

A Main Directional Maximal Difference Analysis for Spotting Micro-expressions

Su-Jing Wang¹, Shuhang Wu², Xiaolan Fu¹

¹CAS Key Laboratory of Behavioral Science, Institute of Psychology, Beijing, 100101, China

wangsujing@psych.ac.cn

²College of Information Science and Engineering, Northeastern University, Shenyang, China.

Abstract. Micro-expressions are facial expressions that have a short duration (generally less than 0.5 seconds), involuntary appearance and low intensity of movement. They are regarded as unique cues revealing the hidden emotions of an individual. Although methods for the spotting and recognition of general facial expressions have been investigated, little progress has been made in the automatic spotting and recognition of micro-expressions. In this paper, we proposed the Main Directional Maximal Difference (MDMD) analysis for micro-expression spotting. MDMD uses the magnitude of maximal difference in the main direction of optical flow as a feature to spot facial movements, including micro-expressions. Based on block-structured facial regions, MDMD obtains more accurate features of the movement of expressions for automatically spotting micro-expressions and macro-expressions from videos. This method obtains both the temporal and spatial locations of facial movements. The evaluation was performed on two spontaneous databases (CAS(ME)² and CASME) containing micro-expressions and macro-expressions.

1 Introduction

The telling of a lie exists everywhere in human social intercourse. Lies are extremely difficult to detect, and although everyone experiences deceiving others or being deceived, even specialists cannot detect them. The polygraph is widely employed in the traditional lie-detection systems, which monitor uncontrolled changes in heart rate and electro-dermal responses when the subject is telling a lie. However, the polygraph makes incursions into the private space of the subject, and the subject can take steps to conceal their genuine emotions [1]. Recently, a type of expression called a micro-expression, which is involuntary, of short duration and low intensity, has aroused the broad concern of affective computational psychologists and researchers. Micro-expressions may appear when individuals are likely to hide their real emotions, and they are very rapid and minute, especially in high-stakes situations [2] [3]. Ekman professed that micro-expressions may be the most reliable clue for detecting lies [3].

Micro-expressions can be detected using a concealed camera during a conversation or interview; therefore, the person will not realize that he is being judged

on whether he is lying. Spotting micro-expressions automatically from videos in a trial may considerably contribute to judicial officials detecting clues of deception by culprits. The automatic spotting of micro-expressions in long videos thus offers great potential.

Regarding micro-expression recognition, a number of papers have been published in recent years. Polikovskiy *et al.* [4] recognized micro-expressions based on the 3D-Gradients orientation histogram descriptor. Pfister *et al.* [5] developed the Temporal Interpolation Model (TIM), handling dynamic features by spatiotemporal local texture descriptors (SLTD), and then used Support Vector Machine (SVM), Multiple Kernel Learning (MKL) and Random Forest (RF) classifiers to recognize spontaneous facial micro-expressions. They [6] also proposed a new spatiotemporal local texture descriptor (CLBP-TOP) to differentiate spontaneous vs. posed (SVP) facial expressions. Wang *et al.* [7] utilized Discriminant Tensor Subspace Analysis (DTSA), which treats a gray facial image as a third-order tensor and Extreme Learning Machine (ELM). However, subtle movements of micro-expressions may be lost in this method. Wang *et al.* [8][9] set up a novel color space model, Tensor Independent Color Space (TICS), because color could provide useful information for expression recognition. Then, they [10] used the sparse part of Robust PCA (RPCA) to extract the subtle motion information of micro-expressions and Local Spatiotemporal Directional Features (LSTD) to extract the local texture features.

The number of papers on micro-expression spotting is smaller than that on micro-expression recognition. Shreve *et al.* [11], [12] primarily used a robust optical flow method [13] for computing the strain from the measured displacement (motion) observed in the video sequence to differentiate macro-expressions and micro-expressions. Polikovskiy *et al.* [4], [14] calculated the durations of the three phases of micro-expressions by the 3D-Gradients orientation histogram descriptor. Moilanen *et al.* [15] proposed a method based on Local Binary Pattern (LBP) histogram features to obtain both the temporal locations and spatial locations for micro-expression spotting. Current research on micro-expression spotting is constrained mostly by micro-expression databases, which employ cropped micro-expression samples or short videos. One exception to this is Shreve *et al.* [12], but the database used in Shreve's study is not yet publicly available.

2 Main Directional Maximal Difference Analysis

2.1 Face alignment, face cropping and block-structure

The inner eye corners were calibrated manually in the first frame of videos to align faces using nonreflective similarity transformation. Nonreflective similarity transformation supports translation, rotation, and isotropic scaling. It has four degrees of freedom and requires two pairs of points, which is similar to affine transformation that needs three pairs of non-collinear points. The inner eye corners are relatively steady [16] and are used in the rest of the video. The original image is shown in Fig. 1 (left), and the aligned image is shown in Fig. 1 (right).



Fig. 1. An example of face alignment.

The method, called Discriminative Response Map Fitting (DRMF) [17], can obtain the outline points of a face and was employed to carry out face cropping. The cropped face image was divided into blocks. We utilized a 6×6 block structure [15] that comprises all the crucial parts of a face and guarantees a relatively low computational complexity. The block structure was based on the horizontal distance between the inner eye corners, the vertical distance between the nasal spine point and the horizontal line connecting the inner eye corners. It is adaptable to faces of different sizes and maintained for each video in this paper because of the measure of face cropping, which is shown in Fig. 2.

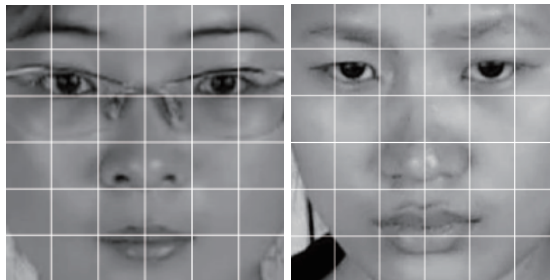


Fig. 2. Examples of the facial 6×6 block structure.

2.2 Main Directional Maximal Difference Analysis

Given a video with n frames, the current frame is denoted as F_i . F_{i-k} is the k -th frame before F_i , and F_{i+k} is the k -th frame after F_i . The optical flow between the F_{i-k} frame (Head Frame) and the F_i frame (Current Frame) after alignment is denoted (u^{HC}, v^{HC}) ¹. Similarly, the optical flow between the F_{i-k} frame (Head Frame), and the F_{i+k} frame (Tail Frame) is denoted (u^{HT}, v^{HT}) . The

¹ For convenience, (u^{HC}, v^{HC}) means the displacement of any point.

(u^{HC}, v^{HC}) and (u^{HT}, v^{HT}) are converted from Euclidean coordinates to polar coordinates (ρ^{HC}, θ^{HC}) and (ρ^{HT}, θ^{HT}) , where ρ and θ represent the magnitude and direction, respectively.

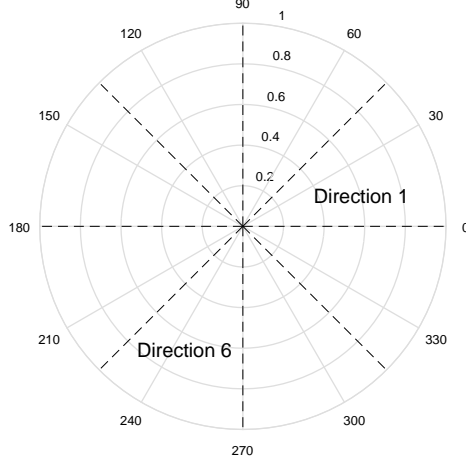


Fig. 3. 8 directions in polar coordinates.

The main direction of the optical flow can well characterize micro-expressions [18]. Based on the directions $\{\theta^{HC}\}$, all the optical flow vectors $\{(\rho^{HC}, \theta^{HC})\}$ are separated into eight directions (see Fig. 3). The *Main Direction* Θ is the direction that has the largest number of optical flow vectors among the eight directions. The main directional optical vector $(\rho_M^{HC}, \theta_M^{HC})$ is the optical flow vector (ρ^{HC}, θ^{HC}) that falls in the *Main Direction* Θ .

$$\{(\rho_M^{HC}, \theta_M^{HC})\} = \{(\rho^{HC}, \theta^{HC}) | \theta^{HC} \in \Theta\} \quad (1)$$

The optical flow vector corresponding to $(\rho_M^{HC}, \theta_M^{HC})$ between the F_{i-k} frame and the F_{i+k} frame is denoted as $(\rho_M^{HT}, \theta_M^{HT})$.

$$\{(\rho_M^{HT}, \theta_M^{HT})\} = \{(\rho^{HT}, \theta^{HT}) | (\rho^{HT}, \theta^{HT}) \text{ and } (\rho_M^{HC}, \theta_M^{HC}) \text{ are two different vectors of the same point in } F_{i-k}\} \quad (2)$$

After sorting the differences $\rho_M^{HC} - \rho_M^{HT}$ in descending order, the maximal difference d^i is the mean difference value of the first $\frac{1}{3}$ of the differences $\rho_M^{HC} - \rho_M^{HT}$ to characterize the frame F_i as the formulation

$$d = \frac{3}{g} \sum \max_{\frac{g}{3}} \{\rho^{HC} - \rho^{HT}\} \quad (3)$$

where $g = |\{(\rho^{HC}, \theta^{HC})\}|$ is the number of elements in the subset $\{(\rho^{HC}, \theta^{HC})\}$, and $\max_m S$ denotes a set that comprises the first m maximal elements in subset S .

In practice, we employed the 6×6 block structure introduced in Section 2.1. We will calculate the maximal difference d_b^i ($b = 1, 2, \dots, 36$) for each block in F_i frame.

It was discovered that picking out approximately one-third of the difference values can obtain a better distinction [15]. For the frame F_i , there are 36 maximal differences d_b^i owing to the 6×6 block structure. Similarly, we arranged the 36 maximal differences d_b^i in descending order. \bar{d}^i is the maximal differences of the first $\frac{1}{3}$, i.e., 12, of the 36 mean values to characterize the frame F_i feature:

$$\bar{d}^i = \frac{1}{12} \sum_{12} \max\{d_b^i\} \quad b = 1, 2, \dots, 36 \quad (4)$$

If a person maintains a neutral expression at F_{i-k} , an emotional expression such as disgust will appear at the onset frame between F_{i-k} and F_i and be repressed at the offset frame between F_i and F_{i+k} , and her facial expression will recover to a neutral expression at F_{i+k} , as is presented in Fig. 4(a). Under this circumstance, the movement between F_i and F_{i-k} is deservedly more intense than that between F_{i+k} and F_{i-k} because the expressions are neutral at both F_{i+k} and F_{i-k} . Therefore, the \bar{d}^i value will be large. In another situation, if a person maintains a neutral expression from F_{i-k} to F_{i+k} , the movement between F_i and F_{i-k} will be similar to that between F_{i+k} and F_{i-k} , so the \bar{d}^i value will be small. In a long video, an emotional expression sometimes appears at the onset frame before F_{i-k} and is repressed at the offset frame after F_{i+k} (see Fig. 4(b)), so the \bar{d}^i value will be small if k is set to a small value. However, k cannot take too large a value, as this could influence the accuracy of computing the optical flow.

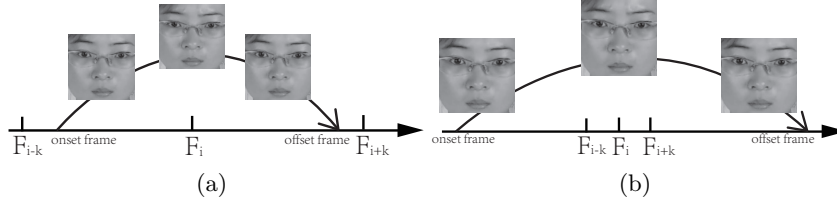


Fig. 4. (a) An emotional expression starts at the onset frame between F_{i-k} and F_i , is repressed at the offset frame between F_i and F_{i+k} and recovers a neutral expression at F_{i+k} ; (b) An emotional expression of a person starts at the onset frame before F_{i-k} and is repressed at the offset frame after F_{i+k} .

2.3 Expression Spotting

We employed the relative difference vector for eliminating background noise which was computed by

$$r^i = \bar{d}^i - \frac{1}{2} (\bar{d}^{i-k+1} + \bar{d}^{i+k-1}) \quad i = k + 1, k + 2, \dots, n - k \quad (5)$$

This is shown in Fig. 5(a), except for the first and the last k frames of a video. The negative difference values indicate that the movement between F_i and F_{i-k} is subtler than the movement between F_{i+k} and F_{i-k} . Accordingly, all negative difference values were set to zero (see Fig. 5(b)).

Thresholding was used to ascertain the frames that have the highest intensity of the facial movements in a video,

$$\begin{aligned}
 & \mathbf{if} \ r_{max} - r_{mean} < a : \\
 & \quad threshold = r_{mean} + b \times (r_{max} - r_{mean}) + c \\
 & \mathbf{else} : \\
 & \quad threshold = r_{mean} + b \times (r_{max} - r_{mean}) \tag{6}
 \end{aligned}$$

where $r_{mean} = \frac{1}{n-2k} \sum_{i=k+1}^{n-k} r^i$ and $r_{max} = \max_{i=k+1}^{n-k} r^i$ are the average and the maximum of all r^i for the entire video. The parameter a is employed due to the condition that the expressions are so subtle that the difference between the emotional expressions and neutral expression in long videos is small, and b is a variable parameter in the range $[0, 1]$. The threshold is more adaptive to improve the robustness of micro-expression detection in long videos, and it is shown in Fig 5 as the red dashed line. The difference values above the red dashed line are the frames that show expressions.

3 Experiments

3.1 Evaluation on CAS(ME)²

The Chinese Academy of Sciences Macro-Expressions and Micro-Expressions (CAS(ME)²) database is the first publicly available database comprising both spontaneous macro-expressions and micro-expressions in long videos (Part A) and separate samples (Part B). Macro-expressions and micro-expressions were collected from the same participants under the same experimental conditions.

In the CAS(ME)² database, Part A has 87 long videos that include spontaneous macro-expressions and micro-expressions collected from 22 participants and Part B contains 300 spontaneous macro-expression samples and 57 micro-expression samples. To our knowledge, there are no publicly available databases that contain macro- and micro-expressions in long videos that can be used for expression detection. The CAS(ME)² database used a Logitech Pro C920 camera with 30 frames per second and a resolution of 640×480 pixels, which satisfied the constraint of brightness constancy. The expression samples were selected from more than 600 elicited facial movements and were coded with the onset, apex, and offset frames, with the AUs marked, emotions labeled, and a self-report for each expression.

In the experiments, we use 59² videos which include 152 macro-expressions and 38 micro-expressions. The maximum duration of the macro-expressions in

² 28 videos were removed because of relatively large movements of the head.

this database is more than 500 ms and less than 4 s, and the maximum duration of the micro-expressions is no greater than 500 ms. The average durations of the macro-expressions and micro-expressions are approximately 1305 ms and 419 ms, respectively. According to the average durations of macro-expressions and micro-expressions, k is set to 12, with $a = 6.6$, $b = 0.15$, $c = 0.15$. The spotting results of a video named 'disgust02' of subject no. 2 are presented in Fig. 5. The green areas denote the durations of expressions or blinking.

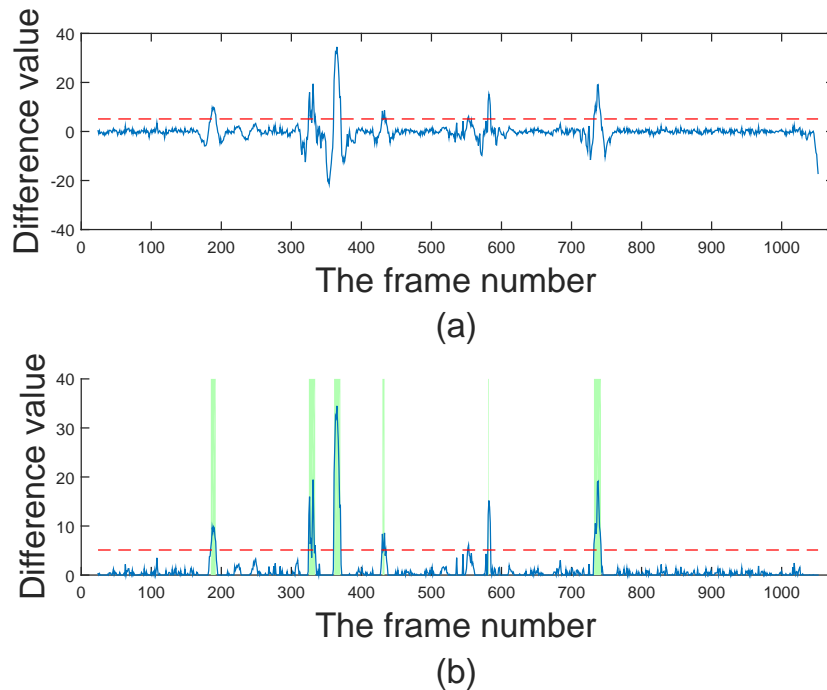


Fig. 5. Spotting results of a video named 16.0102 using the MDMD feature on the CAS(ME)² database.

The LBP feature was used [15] as a comparison by computing the LBP histogram in the (8, 1) neighborhood with $k=12$ and $p=0.25$. The spotting results of the same video are presented in Fig. 6. Apparently, the difference by using the MDMD feature is more notable than that of using the LBP feature on the same video.

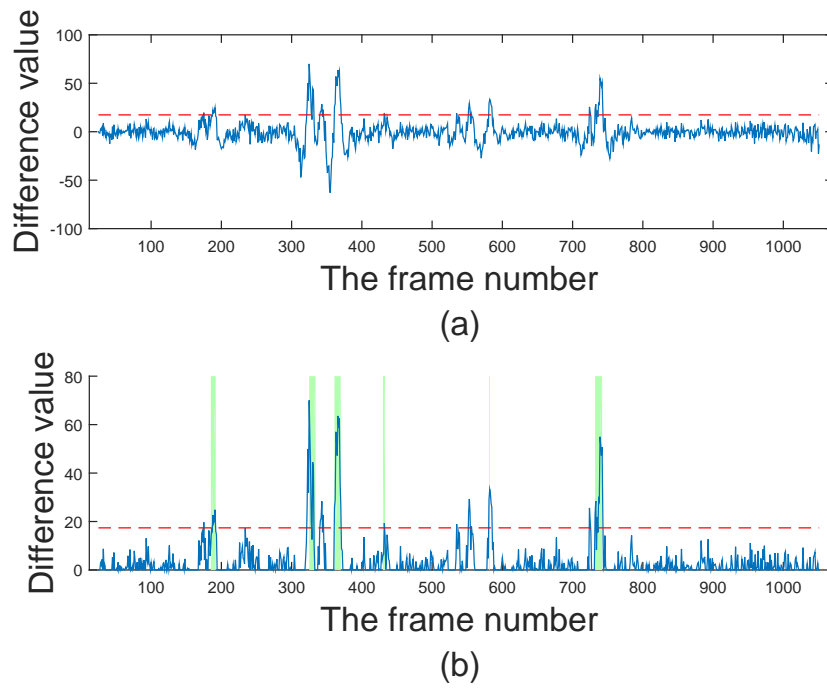


Fig. 6. Spotting results of a video named 16_0102 by using LBP feature on the CAS(ME)² database.

The results of the accuracy of the MDMD feature compared with that of the LBP feature are shown in a table with the following equations:

$$ACC = \frac{\textit{spotted frames which are expressions}}{\textit{all the numbers of spotted frames}} \quad (7)$$

$$TOTAL = \frac{\textit{spotted expressions}}{\textit{all the numbers of expressions}} \quad (8)$$

$$MIC = \frac{\textit{spotted micro - expressions}}{\textit{all the numbers of micro - expressions}} \quad (9)$$

$$MAC = \frac{\textit{spotted macro - expressions}}{\textit{all the numbers of macro - expressions}} \quad (10)$$

We regarded the spotted frames as true results if they fell within the span of $k/2$ before or after the onset or offset of the truth frames provided by the labels of the database. The distance between spotted peaks was not considered, owing to the fact that there are several overlapping micro-expressions in this database. Eye blinks were treated as true results as well because eye blinks can express emotions, e.g., the squinting of the eyes (nervousness or disagreement) or rolling of the eyes (contempt) [3] [19]. They will cause rapid movements in some regions around the eyes [15] and exist in several micro-expressions. An evaluation ac-

Table 1. Evaluation by using MDMD and LBP on CAS(ME)² database

	ACC	TOTAL	MIC	MAC
MDMD	61.3%	74.6%	55.3%	79.5%
LBP	42.7%	70.4%	47.3%	76.2%

ording to the emotional category on the CAS(ME)² database is shown in the table . The MDMD feature performed better than the LBP feature under the conditions of Positive, Negative and Others.

Table 2. Evaluation according to emotional category by using MDMD and LBP on CAS(ME)² database

Category	Number	TOTAL		MIC		MAC	
		LBP	MDMD	LBP	MDMD	LBP	MDMD
Positive	54	70.3%	77.8%	0%	25%	76%	82%
Negative	68	72.1%	75%	20%	30%	81%	82.8%
Surprise	14	71.4%	64.3%	66.7%	66.7%	75%	62.5%
Others	54	69.8%	73.6%	66.7%	72.2%	71.4%	74.3%

3.2 Evaluation on CASME

The Chinese Academy of Sciences Micro-Expression (CASME) database [20] includes 195 spontaneous facial micro-expressions recorded by two 60 fps cameras. The samples were selected from more than 1,500 facial expressions. The CASME database is divided into two classes: Set A and Set B. The samples in Set A were recorded with a BenQ M31 consumer camera with 60fps, with the resolution set to 1280×720 pixels. The participants were recorded in natural light. The samples in Set B were recorded with a Point Grey GRAS-03K2C industrial camera with 60fps, with the resolution set to 640×480 pixels. The participants were recorded in a room with two LED lights. This experiment was performed on Set B because of the fixed light source. The average duration of the micro-expression samples in Set B was approximately 299 ms. All of the samples in Set B were utilized for evaluation in this experiment. According to the average duration of the micro-expressions, k is set to 5 with $a=1.69$, $b=0.6$, and $c=0.3$. The spotting results of a video named EP12.2.5 from subject 8 with 77 frames is presented in Fig. 7, and a micro-expression of disgust is successfully spotted from approximately frame 11 to frame 13.

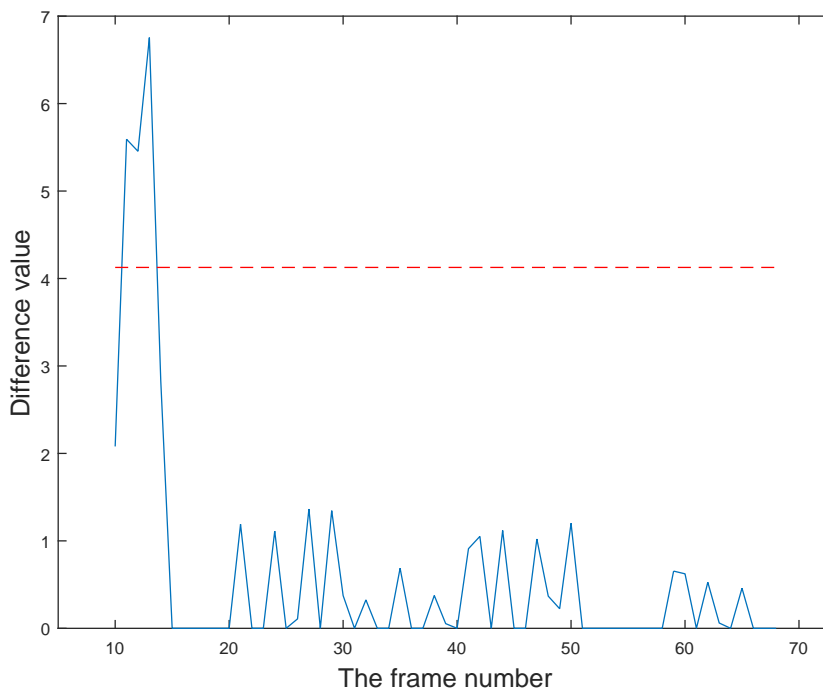


Fig. 7. Spotting results of a video named EP12.2.5 from subject 8 by using MDMD feature on the CASME Set B.

The LBP feature was used [15] as a means of comparison by computing the LBP histogram in the (8, 1) neighborhood with $k=5$ and $p=0.85$. The spotting results of the same video are presented in Fig. 8, and there is no micro-expression at frame 35.

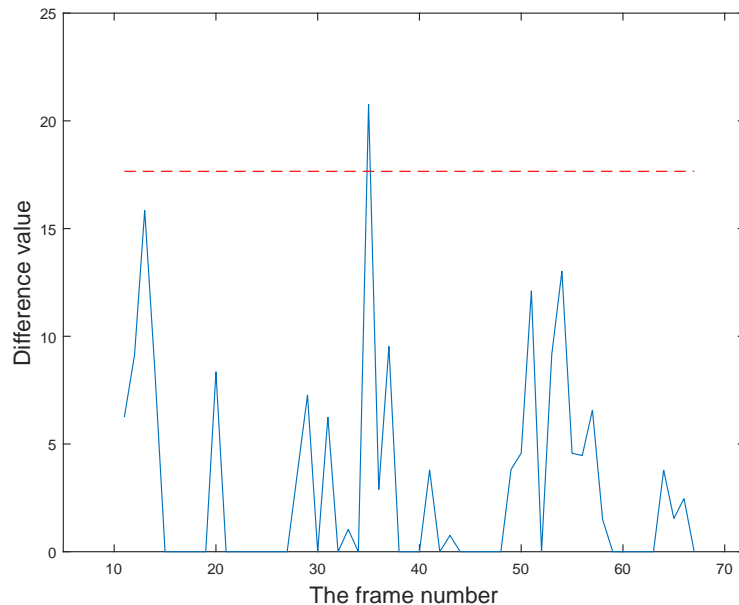


Fig. 8. Spotting results of a video named EP12.2_5 from subject 8 by using LBP feature on the CASME Set B.

The accuracy of the MDMD feature compared with that of the LBP feature is shown in the table . We regarded the spotted frames as true results if they fall between the onset and offset of the truth frames provided by the labels of the database. The constraint in Set B was stricter than that of CAS(ME)² because all the micro-expression in CASME-B are short videos. Eye blinks were considered as true results. The value for k is set to 2 in a video called EP12.4_4 from subject 8 of Set B due to it being a short video with only 13 frames. The flickering light in the video of the CASME-B database influenced the accuracy of the results when using the MDMD feature. The capability of the MDMD feature may be higher when it is employed in long videos. Evaluation according to the emotional category on the Set B database is presented in Table 4. The MDMD feature performed better than the LBP feature under the condition of Surprise, while it performed nearly the same under the conditions of Sadness, Repression, Fear, and Happiness.

Table 3. Evaluation by using MDMD and LBP on CASME Set B

	ACC	MIC
MDND	67.8%	53.5%
LBP	57.2%	58.4%

Table 4. Evaluation according to emotional category by using MDMD and LBP on CASME-B database

Category	Number	MIC	
		LBP	MDMD
Disgust	42	64.3%	59.5%
Sadness	6	50%	50%
Surprise	14	57.1%	71.4%
Tense	23	60.9%	39.1%
Repression	10	50%	50%
Fear	1	100%	100%
Happiness	5	20%	20%

4 Conclusions

In this paper, we proposed the Main Directional Maximal Difference (MDMD) Analysis for micro-expression spotting. We pre-processed databases, including facial alignment, cropping and division, primarily by nonreflective similarity transformation. Based on block-structured facial regions, we calculated a robust local optical flow. We proposed that MDMD will obtain more accurate features of the movement of expressions. The MDMD features were used to spot micro-expressions.

Evaluation was performed on MDMD using two spontaneous databases (CAS(ME)² and CASME) by comparison with LBP [15]. For the CAS(ME)² database, the efficiency and accuracy of the MDMD feature were superior to those of LBP. For the CASME database, the flickering light in the video influenced the accuracy of MDMD. The capability of MDMD feature may be higher in long videos than in short videos.

MDMD achieves not only temporal detection but also the spatial location of facial movements. An example of the 8 largest difference values colored as white, of a person who frowned with a curl of her lip from the CAS(ME)² database is presented in Fig 9. The expression of the frown was labeled AU 4 (brow lower) according to the FACS to convey a negative emotion.

In the future, we can employ MDMD to recognize AU.



Fig. 9. An example of difference values in the space when a person frowned with a curl of her lip.

Acknowledgments

This work was supported by grants from the National Natural Science Foundation of China (61379095, 61375009), and the Beijing Natural Science Foundation (4152055).

References

1. Michael, N., Dilsizian, M., Metaxas, D., Burgoon, J.K.: Motion profiles for deception detection using visual cues. In: *Computer Vision—ECCV 2010*. Springer (2010) 462–475
2. Ekman, P., Friesen, W.V.: Nonverbal leakage and clues to deception. *Psychiatry* **32** (1969) 88–106
3. Ekman, P.: *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage* (Revised Edition). WW Norton & Company (2009)
4. Polikovskiy, S., Kameda, Y., Ohta, Y.: Facial micro-expressions recognition using high speed camera and 3d-gradient descriptor. In: *3rd International Conference on Imaging for Crime Detection and Prevention (ICDP 2009)*. (2009)
5. Pfister, T., Li, X., Zhao, G., Pietikäinen, M.: Recognising spontaneous facial micro-expressions. In: *Computer Vision (ICCV), 2011 IEEE International Conference on*, IEEE (2011) 1449–1456
6. Pfister, T., Li, X., Zhao, G.: Differentiating spontaneous from posed facial expressions within a generic facial expression recognition framework. In: *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, IEEE (2011) 868–875
7. Wang, S.J., Chen, H.L., Yan, W.J., Chen, Y.H., Fu, X.: Face recognition and micro-expression recognition based on discriminant tensor subspace analysis plus extreme learning machine. *Neural processing letters* **39** (2014) 25–43

8. Wang, S.J., Yan, W.J., Li, X., Zhao, G., Fu, X.: Micro-expression recognition using dynamic textures on tensor independent color space. In: Pattern Recognition (ICPR), 2014 22nd International Conference on, IEEE (2014) 4678–4683
9. Wang, S.J., Yan, W.J., Li, X., Zhao, G., Zhou, C.G., Fu, X., Yang, M., Tao, J.: Micro-expression recognition using color spaces. *IEEE Transactions on Image Processing* **24** (2015) 6034–6047
10. Wang, S.J., Yan, W.J., Zhao, G., Fu, X., Zhou, C.G.: Micro-expression recognition using robust principal component analysis and local spatiotemporal directional features. In: Computer Vision-ECCV 2014 Workshops, Springer (2014) 325–338
11. Shreve, M., Godavarthy, S., Manohar, V., Goldgof, D., Sarkar, S.: Towards macro- and micro-expression spotting in video using strain patterns. In: Applications of Computer Vision (WACV), 2009 Workshop on, IEEE (2009) 1–6
12. Shreve, M., Godavarthy, S., Goldgof, D., Sarkar, S.: Macro- and micro-expression spotting in long videos using spatio-temporal strain. In: Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on, IEEE (2011) 51–56
13. Black, M.J., Anandan, P.: The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *Computer vision and image understanding* **63** (1996) 75–104
14. Polikovsky, S., Kameda, Y., Ohta, Y.: Detection and measurement of facial micro-expression characteristics for psychological analysis. *Kameda's Publication* **110** (2010) 57–64
15. Moilanen, A., Zhao, G., Pietikainen, M.: Spotting rapid facial movements from videos using appearance-based feature difference analysis. In: Pattern Recognition (ICPR), 2014 22nd International Conference on, IEEE (2014) 1722–1727
16. Valstar, M.F., Pantic, M.: Fully automatic recognition of the temporal phases of facial actions. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on* **42** (2012) 28–43
17. Asthana, A., Zafeiriou, S., Cheng, S., Pantic, M.: Robust discriminative response map fitting with constrained local models. In: Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, IEEE (2013) 3444–3451
18. Liu, Y.J., Zhang, J.K., Yan, W.J., Wang, S.J., Zhao, G., Fu, X.: A main directional mean optical flow feature for spontaneous micro-expression recognition. *IEEE Transactions on Affective Computing* **PP** (2015) 1–1
19. Ekman, P.: Lie catching and microexpressions. *The philosophy of deception* (2009) 118–133
20. Yan, W.J., Wu, Q., Liu, Y.J., Wang, S.J., Fu, X.: Casme database: a dataset of spontaneous micro-expressions collected from neutralized faces. In: Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on, IEEE (2013) 1–7