A Main Directional Maximal Difference Analysis for Spotting Facial Movements from Long-term Videos

Su-Jing Wang[a,*], Shuhang Wu[b], Xingsheng Qian[a], Jingxiu Li[c], Xiaolan Fu[d,e]

[a] *CAS Key Laboratory of Behavioral Science, Institute of Psychology, Beijing, 100101, China*

[b] *College of Information Science and Engineering, Northeastern University, Shenyang, China*

[c] *Department of Cardiology, the Fourth Affiliated Hospital of Harbin Medical University, Harbin, 150001, China*

[d] *State Key Laboratory of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, 100101, China*

[e] *University of the Chinese Academy of Sciences, Beijing, 100049, China*

Abstract

There is an increasing interests in micro-expression researches. Spotting micro-expressions in long-term videos is very important, not only for providing clues for lie detection, but also for reducing the labor required to collect micro-expression data. However, little progress has been made in spotting micro-expressions. In this paper, we propose a Main Directional Maximal Difference (MDMD) Analysis for micro-expression spotting. MDMD uses the magnitude maximal difference in the main direction of optical flow features to spot facial movements, including micro-expressions. Using block structured facial regions, MDMD obtains more accurate features of movement of expressions for automatically spotting micro-expressions and macro-expressions from videos. This method involves both the temporal and spatial locations of face movements. Evaluations using the CAS(ME)^2 database containing micro-expressions and macro-expressions show that MDMD is more robust than some state-of-the-art algorithms.

Keywords: Micro-expression Recognition, Macro-expression, Micro-expression Spotting, Optical Flow

*Corresponding author

Email address: wangsujing@psych.ac.cn (Su-Jing Wang)

1. INTRODUCTION

Telling a lie is very common in human social intercourse. Lies are extremely difficult to detect although everyone has deceived others or has been deceived even specialists cannot detect them. The polygraph is employed in the traditional lie-detection system to monitor uncontrolled transformations in heart rate and electro-dermal responses when the subject is telling a lie. However, the polygraph makes incursion into the private space of the subject, and the subject can take steps to conceal their true emotions [1]. Recently, psychologists have found that micro-expressions provide important clues for detecting lies. Micro-expressions may appear when individuals are hiding their real emotions. Micro-expressions are very rapid and tiny, especially in high-stake situations [2] [3]. A lie detection system based on micro-expressions captures facial movements using a concealed camera during a conversation or interview, and therefore, the person will not realize that he is being observed when he is lying. Spotting micro-expressions from long-term videos of facial movements is a key technology that can be used in a lie detection system based on micro-expressions.

Research on facial expressions originated with Darwin *et al.* [4]. A previous study conducted by Mehrabian *et al.* [5] revealed that 55% of messages regarding feelings and attitudes are conveyed by facial expressions. Micro-expressions were first discovered by Haggard *et al.* [6], and were called rapid expressions that showed repressed emotions. Ekman *et al.* [2] found this type of expression by observing a psychotic inpatient, who wanted to commit suicide but concealed the negative expression within 1/12 of a second by smiling and named it a micro-expression. A Facial Action Coding System (FACS) [7] and a Micro Expression Train Tool (METT) were developed later. Micro-expressions can reveal authentic emotions and are considered one of the most important non-verbal methods for determining clues to judge whether someone is lying or being honest [8] [9]in important areas, such as clinical medicine [10] [11] [12] and political psychology [13]). There is extensive research concerning facial expressions, however, knowledge regarding micro-expressions needs to be further studied.

Spotting and automatically recognizing micro-expressions are indispensable and a field of frontier research related research is rare. In micro-expression recognition, several studies have been published. Polikovsky *et al.* [14] recognized micro-expressions based on a 3D-Gradient orientation histogram descriptor. Pfister *et al.* [15] developed the Temporal Interpola-

tion Model (TIM), which handles dynamic features by spatiotemporal local texture descriptors (SLTD) and then uses a Support Vector Machine (SVM), Multiple Kernel Learning (MKL) and Random Forest (RF) classifiers to recognize spontaneous facial micro-expressions. Meanwhile, Pfister *et al.* [16] proposed a new spatiotemporal local texture descriptor (CLBP-TOP) to differentiate spontaneous vs. posed (SVP) facial expressions. Wang *et al.* [17] utilized a Discriminant Tensor Subspace Analysis (DTSA), which treated a gray facial image as a third order tensor, and an Extreme Learning Machine (ELM). However,the subtle movements of micro-expressions may be lost using this method. Wang *et al.* [18][19] established a novel color space model, Tensor Independent Color Space (TICS), because color could provide useful information for expression recognition. Then, they [20] used the sparse part of Robust PCA (RPCA) to extract the subtle motion information of the micro-expressions and Local Spatiotemporal Directional Features (LSTD) to extract local texture features.

The amount of research regarding micro-expression spotting is less than that for micro-expression recognition. Shreve *et al.* [21], [22] used a robust optical flow method [23] to compute strain from the measured displacement (motion) observed in a video sequence to differentiate macro-expressions from micro-expressions. Polikovsky *et al.* [14], [24] calculated the duration of the three phases of micro-expressions using a 3D-Gradient orientation histogram descriptor. Moilanen *et al.* [25] proposed a method based on Local Binary Patterns (LBP) histogram features to obtain both temporal locations and spatial locations for micro-expression spotting. Xia *et al.* [26] utilized Adaboost to estimate the initial probability for each frame and used random walk to model the correlation between the frames. In this paper, this method is called RW-Adaboost. Davison *et al.* [27] used a HOG feature to replace the LBP feature in [25] and proposed an individualized baseline threshold for spotting micro-expression. Current research on micro-expression spotting is constrained by the micro-expression databases using cropped micro-expression samples or short videos, except for Shreve et al. [22]; unfortunately, the database used in Shreve's study is still not publicly available.

The number of micro-expression databases is far smaller than that of macro-expression databases. There are only the following seven published micro-expression databases: (1) USF-HD [22]; (2) the Polikovsky's database [14]; (3) SMIC (The Spontaneous Micro-expression Corpus) [15]; (4) SMIC2 [28]; (5) CASME (The Chinese Academy of Sciences Micro-expression) [29];

(6) CASME II [30]; and (7) CAS(ME)^2 (it has been accepted and will be published soon). It is difficult to elicit micro-expressions and encode them. This is the main reason why there are few micro-expression databases. The work to manually code frame by frame is tedious and very time consuming. Although the accuracy of spotting micro-expressions has not achieved a satisfactory level currently, it can greatly reduce the work of manual coding. Spotting micro-expressions is not only an essential step for micro-expression recognition, but is also significant for reducing manual codes and increasing the efficiency of obtaining databases.

This paper is an extended version of our Asian Conference on Computer Vision (ACCV) paper [31] in which we proposed using a Main Directional Maximal Difference (MDMD) analysis to spot micro-expressions in long-term videos. In this paper, we analyze the influence of the parameters on performance and compare MDMD to some state-of-the-art algorithms for spotting micro-expressions. The remainder of this paper is organized as follows. In Section 2, we will introduce the pre-process, propose the Main Directional Maximal Difference (MDMD) analysis, and analyze the influence of the parameters k on the performance. In Section 3, a database based on long videos will be introduced. MDMD, LBP, HOG, and RW-Adaboost are conducted using the database, and the parameters of MDMD are thoroughly discussed. Finally, in Section 4, conclusions are drawn, and several issues for future work are described.

2. Main Directional Maximal Difference Analysis

To improve the micro-expression spotting performance on long-term videos, we proposed a Main Directional Maximal Difference (MDMD) Analysis. The main steps of our method are as follows: (1) Pre-processing, including facial alignment, cropping and dividing facial images into a block-structure; (2) applying a robust local optical flow (RLOF) on the block-structure facial regions; (3) measuring features frame by frame by calculating the maximal difference values in the main direction of the optical flows (MDMD feature); and (4) proposing a threshold and spotting micro-expressions based on the MDMD feature.

2.1. Face alignment, face cropping and block-structure

The inner eye corners were calibrated manually in the first frame of the video to align the faces using a non-reflective similarity transformation. A

non-reflective similarity transformation supports translation, rotation, and isotropic scaling. It has four degrees of freedom and requires two pairs of points, which is similar to the affine transformation, which requires three pairs of non-collinear points. The inner eye corners are relatively steady [32]. The same transformation is used in the remaining frames of the video. The original image is shown in Fig. 1 (left) and the aligned image is shown in Fig. 1 (right).

Figure 1: An example of face alignment.

Discriminative Response Map Fitting (DRMF) [33] can obtain the outline points of a face, and we use it to crop the face (see Fig. 2). The cropped face image is divided into $b \times b$ blocks. A 6×6 block structure is shown in Fig. 3. The structure comprises all the crucial parts of the face and guarantees a relatively low computational complexity. The block structure is based on the horizontal distance between the inner eye corners, the vertical distance between the nasal spine point, and the 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.

2.2. Robust Local Optical Flow

We employed a robust local optical flow (RLOF) [34] on the block-structural facial regions to estimate facial motion. RLOF not only adapts different region sizes and moderates changing illuminations well but also possesses higher effectiveness with a slight increase of computational complexity than standard KLT, especially when the assumptions made by Lucas/Kanade [35] are violated.

The optical flow computes the motion of objects or scenes by detecting the changing intensity of the pixels between two image frames over time. A pixel

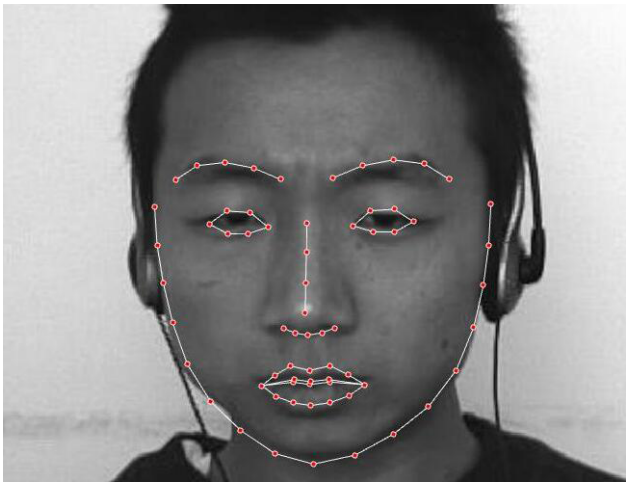

Figure 2: An example of face cropping using DRMF.

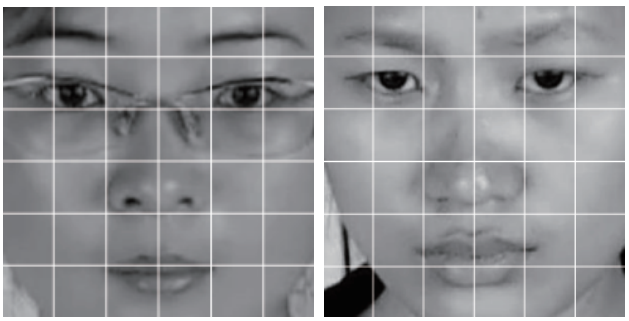

Figure 3: Examples of facial 6×6 block structure.

at location $(x,\ y,\ t)$ with intensity $I(x,\ y,\ t)$ will move to the location $(x+u, y+v, t+\triangle t)$ with intensity $I(x+u,\ y+v,\ t+\triangle t)$ between two frames. Because of the constraints of temporal coherence, spatial coherence, and brightness constancy, we obtain the equation of the intensity constancy within a small identical region in two consecutive images:

$$I(x,\ y,\ t) = I(x+u,\ y+v,\ t+\triangle t) \tag{1}$$

$\mathbf{d} = (u, v)^T$ denotes the displacement of the point at location $(x,\ y,\ t)$, and $\triangle t$ is a small temporal interval.

The RLOF method reduces the number of the Hampel estimator[36] parameters by shrinking the high and low flat segment to obtain the influence function φ:

$$\varphi(\epsilon_i,\ \sigma) = \begin{cases} 2\epsilon_i & |\epsilon_i| \leq \sigma_1 \\ 0 & |\epsilon_i| \geq \sigma_2 \\ \frac{\sigma_1(\epsilon_i - sign(\epsilon_i)\cdot\sigma_2)}{\frac{1}{2}(\sigma_1-\sigma_2)} & \text{otherwise.} \end{cases} \tag{2}$$

ϵ_i is the i-th of all observations in Ω, Ω denotes an image region in which we are interested, and σ_1 and σ_2 are scale parameters. In addition, ϵ is computed as:

$$\epsilon = \nabla\mathbf{I}(\mathbf{x})^T \cdot \mathbf{d} + I_t(\mathbf{x}) \tag{3}$$

with

$$\nabla\mathbf{I}(\mathbf{x}) = (I_x(\mathbf{x}, t),\ I_y(\mathbf{x}, t))^T \tag{4}$$

$$\mathbf{x} = (x,\ y) \tag{5}$$

$$I(\mathbf{x},\ t) = I(\mathbf{x},\ t-1) + I_t(\mathbf{x}) \tag{6}$$

$I_x(\mathbf{x}, t),\ I_y(\mathbf{x}, t))$ and $I_t(\mathbf{x})$ are, respectively, the x directional, y directional and temporal derivative of $I(x,\ y,\ t)$. The inverse compositional RLOF residual error represents the integrals derived from $\varphi(\epsilon_i,\ \sigma)$ as:

$$E_{RLOF} = \sum_{\Omega 1 \subset \Omega} \epsilon^2 + \sum_{\Omega 3 \subset \Omega} \sigma_1\sigma_2 + \sum_{\Omega 2 \subset \Omega} \left(\frac{\sigma_1}{\sigma_1 - \sigma_2}(|\epsilon| - \sigma_2)^2 + \sigma_1\sigma_2\right) \tag{7}$$

To gain the displacement $\mathbf{d}$, the residual error E_{RLOF} is minimized,and $\mathbf{x} \in \Omega$, $\Omega 1$, $\Omega 2$, $\Omega 3$ are the subset of data in Ω fulfilling, respectively, $|\epsilon_i| \leq \sigma_1$, $\sigma_1 < |\epsilon_i| < \sigma_2$ and $|\epsilon_i| \geq \sigma_2$.The estimation of $\mathbf{d}$ can be solved to the

displacement $\Delta\mathbf{d}$ by an iterative solution in a Newton-Raphson fashion [37]:

$$\Delta\mathbf{d}^k = \mathbf{G}_{RLOF}^{-1} \cdot \Bigg[\sum_{\Omega 1 \subset \Omega} \nabla\mathbf{I}(\mathbf{x}) \cdot I_t^{k-1}(\mathbf{x}) + \sum_{\Omega 2 \subset \Omega} \frac{\sigma_1}{\sigma_1 - \sigma_2} \cdot \nabla\mathbf{I}(\mathbf{x}) \cdot \left(I_t^{k-1}(\mathbf{x}) - sign(I_t^{k-1}(\mathbf{x})) \cdot \sigma_2 \right) \Bigg] \tag{8}$$

$\mathbf{G}$ is the modified Hessian matrix:

$$\mathbf{G}_{RLOF} = \sum_{\Omega 1 \subset \Omega} \nabla\mathbf{I}(\mathbf{x}) \cdot \nabla\mathbf{I}(\mathbf{x})^T + \sum_{\Omega 2 \subset \Omega} \frac{\sigma_1}{\sigma_1 - \sigma_2} \nabla\mathbf{I}(\mathbf{x}) \cdot \nabla\mathbf{I}(\mathbf{x})^T \tag{9}$$

The $\mathbf{d}^k$ is updated as:

$$\mathbf{d}^{k-1} + \Delta\mathbf{d}^k \rightarrow \mathbf{d}^k \tag{10}$$

thus, the intensity of the next frame after updating at each iteration k is:

$$I(\mathbf{x},\, t) = I(\mathbf{x} + \mathbf{d}^{k-1},\, t-1) + I_t^{k-1}(\mathbf{x}) \tag{11}$$

The iterative solution is initialized with $\mathbf{d} = (0,\, 0)^T$, $\mathbf{L}^2$ norm is utilized as a monotone φ for the first iteration, and the non-monotone φ corresponding to the non-convex shrunk Hampel norm is added to cycles for the remainder of the iterations.

2.3. 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 the F_i, and F_{i+k} is the k-th frame after the F_i. The optical flow between the F_{i-k} frame (Head Frame) and the F_i frame (Current Frame) after alignment is denoted by (u^{HC}, v^{HC}). For convenience, (u^{HC}, v^{HC}) means the displacement of any point. Similarly, the optical flow between the F_{i-k} frame (Head Frame) and the F_{i+k} frame (Tail Frame) is denoted by (u^{HT}, v^{HT}). Then, (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, respectively, the magnitude and direction.

The main direction of the optical flow can well characterize micro-expressions [38]. Based on the directions $\{\theta^{HC}\}$, all the optical flow vectors $\{(\rho^{HC}, \theta^{HC})\}$ are divided into a directions (see Fig. 4). The *Main Direction* Θ is the direction

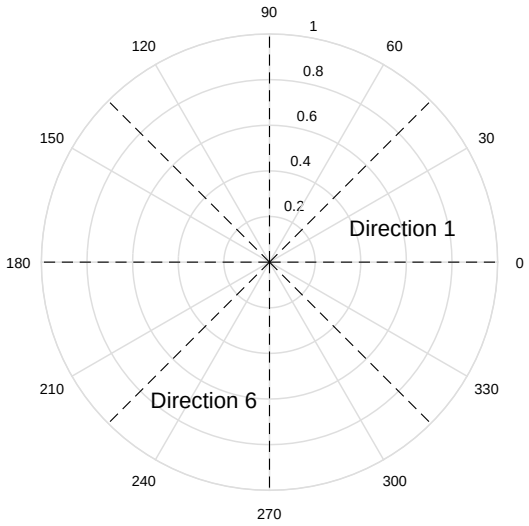

Figure 4: Eight directions in the polar coordinates.

.

that has the largest number of optical flow vectors among the a 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.

$$\{(\rho_M^{HC}, \theta_M^{HC})\} = \{(\rho^{HC}, \theta^{HC}) | \theta^{HC} \in \Theta\} \tag{12}$$

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

$$\begin{aligned} \{(\rho_M^{HT}, \theta_M^{HT})\} = \{(\rho^{HT}, \theta^{HT}) | (\rho^{HT}, \theta^{HT}) \text{ and } (\rho_M^{HC}, \theta_M^{HC}) \\ \text{are two different vectors of the same point in } F_{i-k}\} \end{aligned} \tag{13}$$

After the differences $\rho_M^{HC} - \rho_M^{HT}$ is sorted in a 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 in the formula:

$$d = \frac{3}{g} \sum \max_{\frac{g}{3}} \{\rho^{HC} - \rho^{HT}\} \tag{14}$$

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 comprised of the first m maximal elements in the subset S.

In practice, we employ the $b \times b$ block-structure that was introduced in Section 2.1. We will calculate the maximal difference d_j^i $(j = 1, 2, \ldots, b^2)$ for each block in the F_i frame. For frame F_i, there are b^2 maximal differences d_j^i due to the $b \times b$ block structure. We arrange the b^2 maximal differences d_j^i in a descending order where $\bar{d}^i$ is the first s maximal difference and characterizes the frame F_i feature:

$$\bar{d}^i = \frac{1}{s} \sum \max_s \{d_j^i\} \quad j = 1, 2, \ldots, b^2 \tag{15}$$

If a person maintains a neutral expression at F_{i-k}, his emotional expression, such as disgust, 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 then his facial expression recovers a neutral expression at F_{i+k}, which is presented in Fig. 5. In this circumstance, the movement between F_i and F_{i-k} is more intense than the movement between F_{i+k} and F_{i-k} because the expression is neutral at both F_{i+k} and F_{i-k}. Therefore, the $\bar{d}^i$ value will be large. Another situation is a person maintaining a neutral expression from F_{i-k} to F_{i+k}. The movement between F_i and F_{i-k} is similar to the movement between F_{i+k} and F_{i-k}; thus, the $\bar{d}^i$ value will be small. In a long video, sometimes an emotional expression starts at the onset frame before F_{i-k} and is repressed at the offset frame after F_{i+k} (see Fig. 6). In this case, the $\bar{d}^i$ value will also be small if k is set to be a small value. However, k cannot be set as a large value because this would influence the accuracy of the computing optical flow.

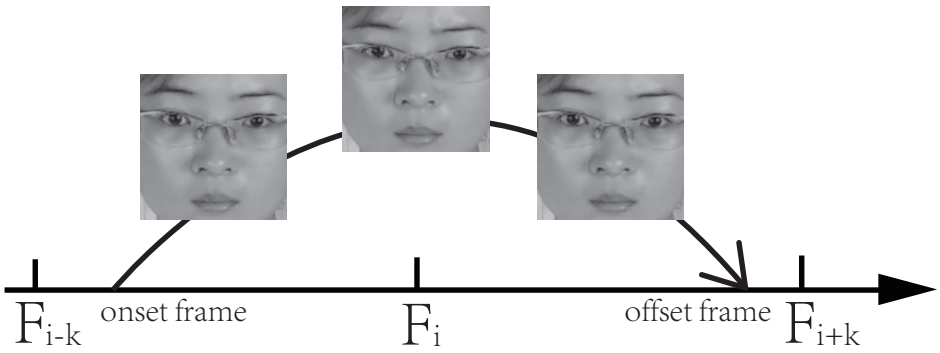

Figure 5: An emotional expression starting at the onset frame between F_{i-k} and F_i is repressed at the offset frame between F_i and F_{i+k} and recovers a neutral expression at F_{i+k}

.

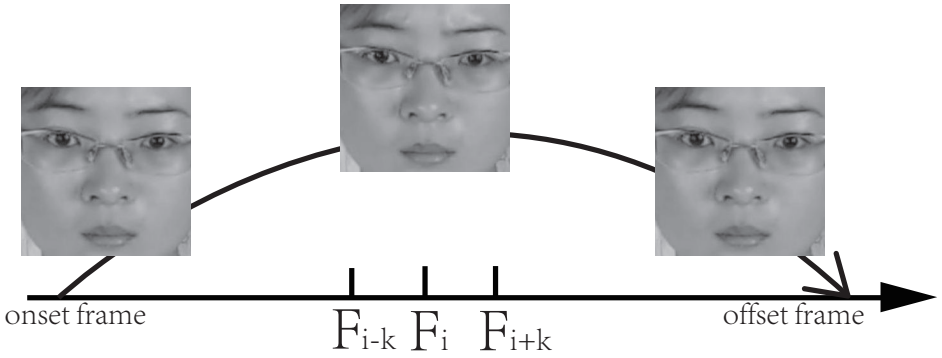

Figure 6: An emotional expression starting at the onset frame before F_{i-k} is repressed at the offset frame after F_{i+k}

.

2.4. Expression Spotting

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

$$r^i = \bar{d}^i - \frac{1}{2}\left(\bar{d}^{i-k+1} + \bar{d}^{i+k-1}\right) \quad i = k+1, k+2, \ldots, n-k \tag{16}$$

As shown in Fig. 7(a), we excluded the first and the last k frames of the video because the negative difference values illustrated that the movement between F_i and F_{i-k} is more subtle than the movement between F_{i+k} and F_{i-k}. Accordingly, all negative difference values were set to zero (see Fig. 7(b)).

A threshold was used to obtain the frames that had peaks representing the facial movements in a video

$$threshold = r_{mean} + p \times (r_{max} - r_{mean}) \tag{17}$$

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 whole video. The parameter p is a variable parameter in the range $[0, 1]$. The threshold is more adaptive to improve the robustness of micro-expression detection in long videosIt is denoted as the red dashed line in Fig 7. The frames with difference values above the red dashed line are the frames where expressions appear. The green areas denote the durations of the expressions or blinks. Spotting results of LBP for the same video is presented in Fig. 8. For the same video, the differences obtained from MDMD features are more notable than those from LBP features.

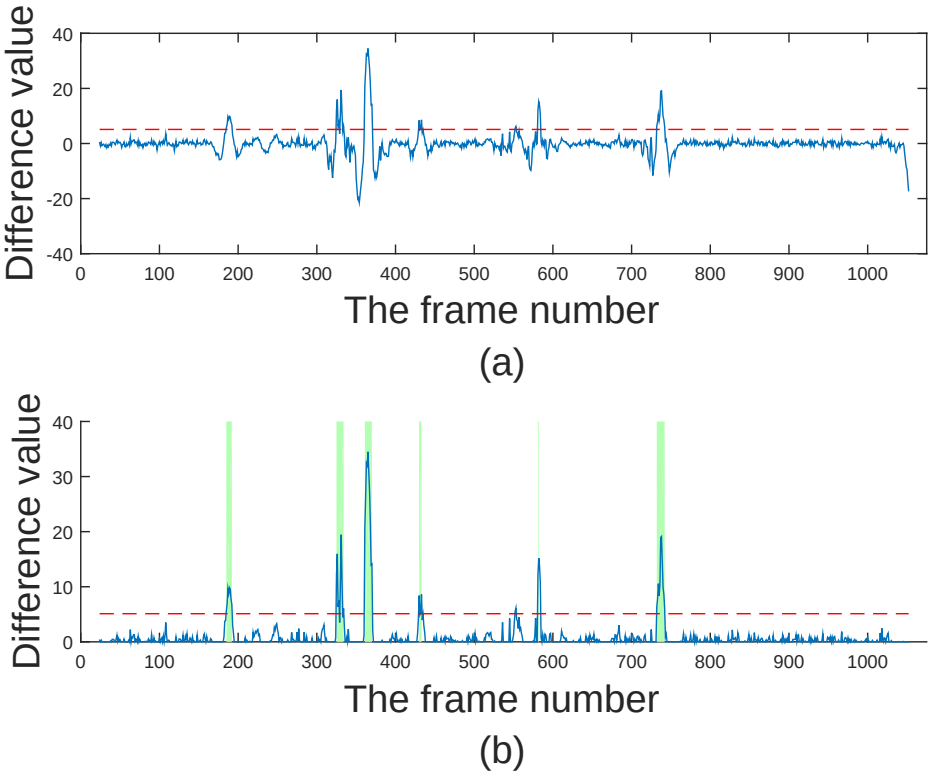

Figure 7: Spotting results of a video named 16_0102 using the MDMD feature on the CAS(ME)^2 database.

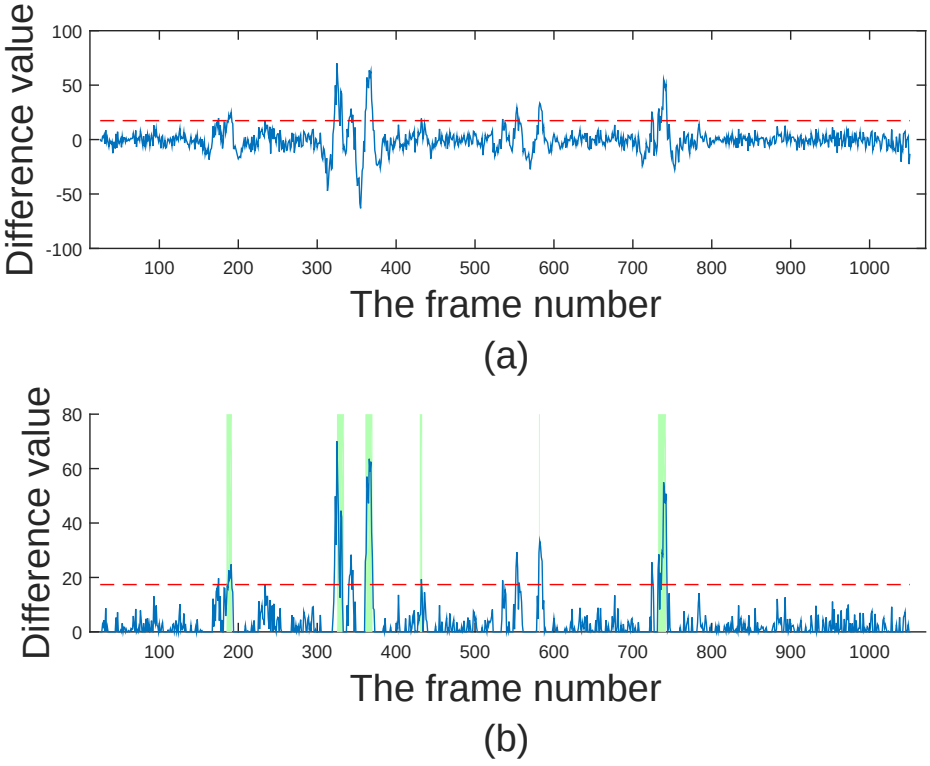

Figure 8: Spotting results of a video named 16_0102 using the LBP feature on the CAS(ME)^2 database.

3. EXPERIMENTS

3.1. CAS(ME)^2 database

To our knowledge, there are no publicly available databases that contain macro-expressions and micro-expressions in long videos that can be used for spotting expressions. The Chinese Academy of Sciences Macro-Expressions and Micro-Expressions (CAS(ME)^2) database will be 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 includes 87 long videos of spontaneous macro-expressions and micro-expressions collected from 22 participants and Part B contains 300 spontaneous macro-expression samples and 57 micro-expression samples. 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 steady consistent brightness. The expression samples were selected from more than 600 elicited facial movements and were coded with the onset, apex, and offset frames, with AUs marked, emotions labeled, and a self-report for each expression.

3.2. Experimental Evaluation

In Part A of CAS(ME)^2, there are 87 long videos. Among these videos, 28 videos were removed because there are relatively large movements of the head. Thus, we use 59 videos that include 152 macro-expressions and 38 micro-expressions. The maximum duration of the macro-expressions is more than 500 ms and less than 4 s, and the maximum duration of micro-expressions is no more 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. The numbers of blocks are 5×5, 6×6, 7×7 and 8×8. The number of directions are 4, 6, 8 and 10. Fig. 9 shows the ROC curves for the 16 different combinations. The 16 ROC curves are almost consistent, showing that the performance of MDMD is stable for the block size and the number of directions. Among the 16 combinations, AUC (Area Under Curve) has the highest value of 0.5862 for the combination where the number of blocks is 6×6 and the number of directions is 4. If the number of blocks is fixed, AUC obtains the highest value when the number of blocks is 4.

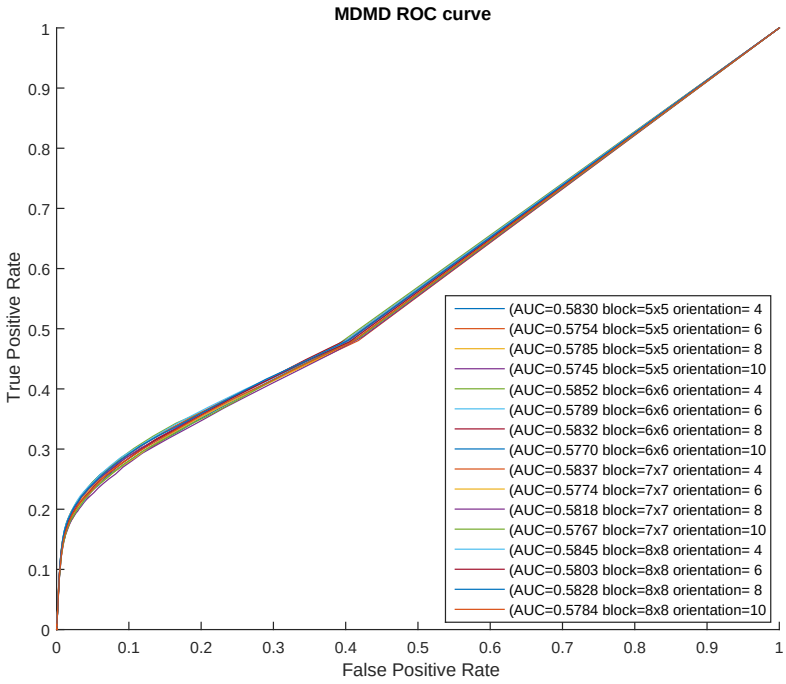

Figure 9: ROC curves for 16 different combinations.

To evaluate MDMD, we applied the LBP feature [25], HOG feature [27], and RW-Adaboost [26] to the same dataset. For the LBP feature, we used 6×6 blocks and set k as 10. For the HOG feature, Spatial HOG features are extracted using Piotr Dollar's Matlab Toolbox. The number of direction bins is set as 8, and the signed gradient direction binning is set as 2π. The number of blocks is 6×6, and k is set as 15. For RW-Adaboost, all parameters are set as described in [26]. The random walk framework (with $\alpha = 0.4$ and 20 update times) was used to generate the final probability after the procedures such as procrustes analysis, and a geometric deformation (with $\beta = 0.7$ and $L+1$ frames is a temporal window where $L = 6$) and Gentle AdaBoost using GML AdaBoost Toolbox (with 30% data for training 40 iterations) were performed. The ROC curves of the four methods are plotted in Fig. 10.

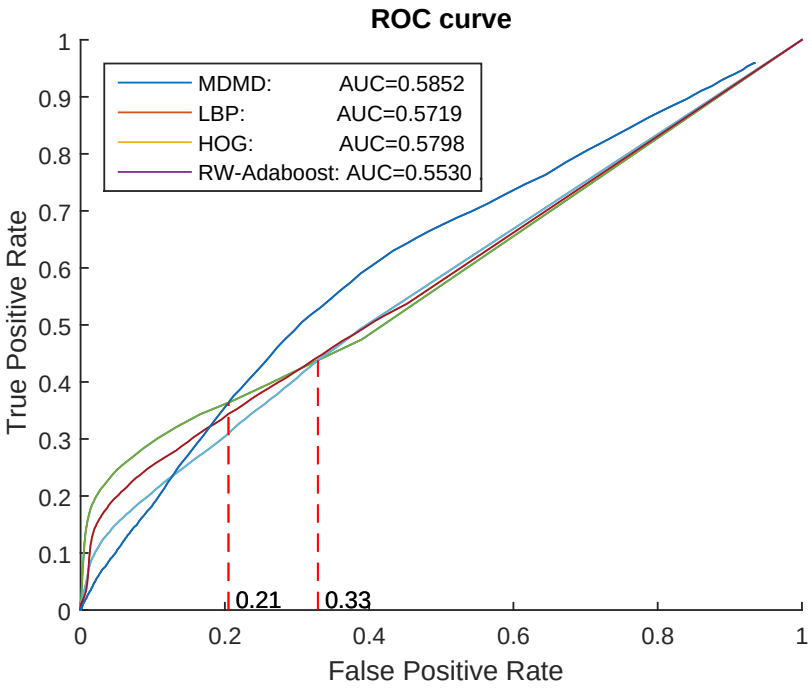

Figure 10: ROC curves of MDMD, LBP, HOG and RW-Adaboost.

Among the four methods, the AUC of MDMD obtains the best value of 0.5852. When the False Positive Rate (FPR) is less than 0.21, the True Positive Rate (TPR) of MDMD is better than those of the other three methods. MDMD performs better than the three other methods when the FPR

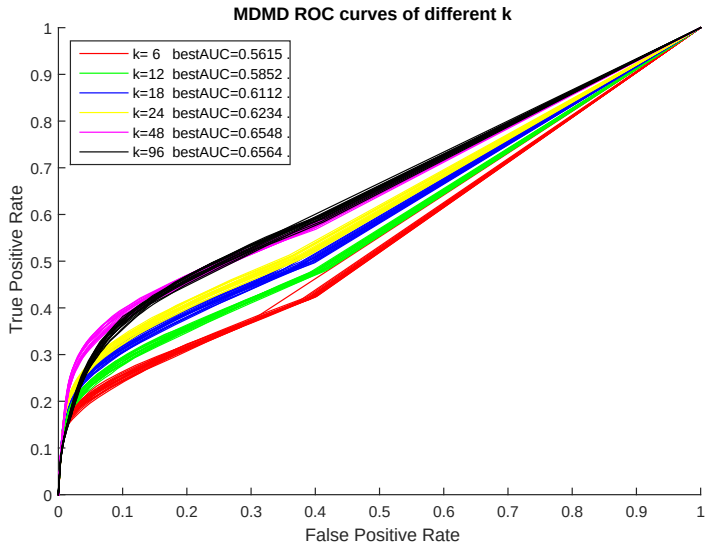

Figure 11: MDMD ROC curves with various k

decreases. When the FPR is greater than 0.33, the ROC curves of MDMD, LBP and HOG are almost consistent and parallel to the line whose angle with the abscissa is 45 degrees. It shows that when the FPR is greater than 0.33, the predications of the three methods are random. For spotting facial movements, the random probability of predication is 0.5 (movement or not). When the FPR is greater than 0.21, RW-Adaboost is slightly better than the other three methods. The reason is that the Adaboost classifier is slightly better than the random level.

We select various k ($k = 6, 12, 18, 24, 48, 96$) to repeat the same experiment. Given k, there are 16 ROC curves on various numbers of blocks and directions. The 16 ROC curves have the same color in Fig. 11. The 16 ROC curves are almost consistent for each k (expect for $k = 6$). This again demonstrates that the performance of MDMD is stable for the numbers of blocks and directions. With an increasing k value, the AUC is larger. Although AUC of $k = 96$ is slightly larger than that of $k = 48$, the performance of $k = 48$ is better than that of $k = 96$ because the slope of the curve of $k = 48$ is larger than that of $k = 96$ in the case that the FPR is small. This is coincident with the analysis of k in Section 2.3.

We select various k ($k = 6, 12, 18, 24, 48, 96$) and 5×5, 6×6, 7×7, 8×8 blocks to repeat the LBP experiment. For each k, the ROC curve with

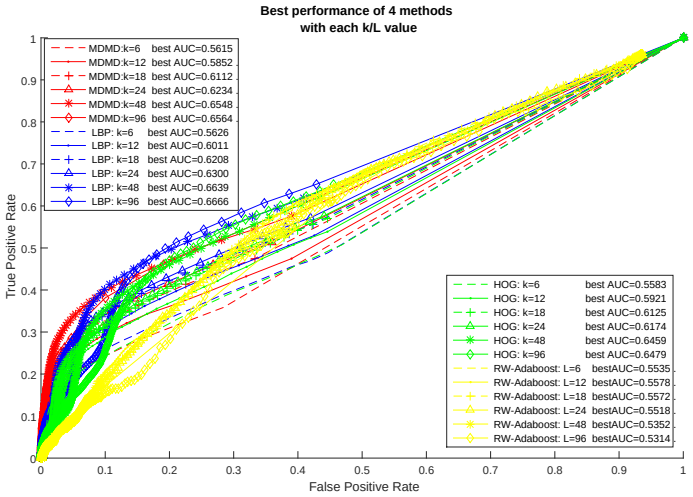

Figure 12: The ROC curves with various k of MDMD, LBP, HOG and RW-Adaboost

the largest AUC is plotted in Fig. 12. For HOG and RW-Adaboost, the same experiments are also conducted. With an increased k value, the AUCs of LBP and HOG are larger. Although the AUC of LBP is slightly larger than that of MDMD, the performance of MDMD is better than that of LBP, because the slope of the curve of MDMD is larger than that of LBP in the case where the FPR is small.

Recall, *Precision*, and F_1 *score* are also used to measure experimental results. There are two classes (movement frame and non-movement frame). The number of correct positive results of the cth class is cp_c, ap_c is the number of all positive results of the cth class and rp_c is the number of positive results that should have been returned of the c class [39]. *Precision* and *Recall* are defined as follows:

$$Precision = \frac{1}{2}\sum_{c=1}^{2}\frac{cp_c}{ap_c} \tag{18}$$

and

$$Recall = \frac{1}{2}\sum_{c=1}^{2}\frac{cp_c}{rp_c} \tag{19}$$

However, the numbers of movement frames and non-movement frames are not balanced. We use the F_1 score to address this. The F_1 score is defined

as follows:

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{20}$$

For MDMD, we select the related best performance in the case with 8×8 blocks, 4 directions, $p = 7$, and $k = 48$. For LBP, we select the related best performance in the case with 8×8 blocks, $p = 11$, and $k = 48$. For HOG, we choose the parameters 6×6 blocks and $k = 48$. For RW-Adaboost, k is set as 6. These performances are listed in Table 1. On *Recall*, HOG obtains the best performance, and MDMD obtains the second-best performance. HOG is more suited for assisting in encoding micro-expressions. MDMD obtains the best performance on *Precision* and F_1 *score*.

Table 1: *Recall*, *Precision*, and F_1 *score* of MDMD, LBP, HOG and RW-Adaboost

	Recall	Precision	F_1 score
MDMD ($p = 7, k = 48$)	31.90%	35.21%	33.48%
LBP ($p = 11, k = 48$)	27.17%	32.26%	29.50%
HOG ($k = 96$)	46.89%	18.15%	26.17%
RW-Adaboost ($k = 6$)	9.30%	26.13%	13.72%

We also investigate the influence of s in Eq. 15 on the performance of MDMD. All parameters are the same as those in the previous experiments. The parameter s varies from 1 to 64. The AUCs of MDMD are plotted in Fig. 13. The curve is parallel to the abscissa. It shows little influence on the performance of MDMD.

4. CONCLUSION

In this paper, we proposed a Main Directional Maximal Difference (MDMD) Analysis for micro-expression spotting. We pre-processed databases that included facial alignment, cropping and division primarily by non-reflective similarity transformation. Based on block structured facial regions, we calculate robust local optical flows. We propose MDMD to obtain more accurate features of the movement of expressions; the MDMD feature was used to spot micro-expressions. The results were evaluated on CAS(ME)^2 databases (and CASME) using four methods, MDMD, LBP, HOG, and RW-Adaboost. On the CAS(ME)^2 database, MDMD performs well in spotting micro-expressions from long videos.

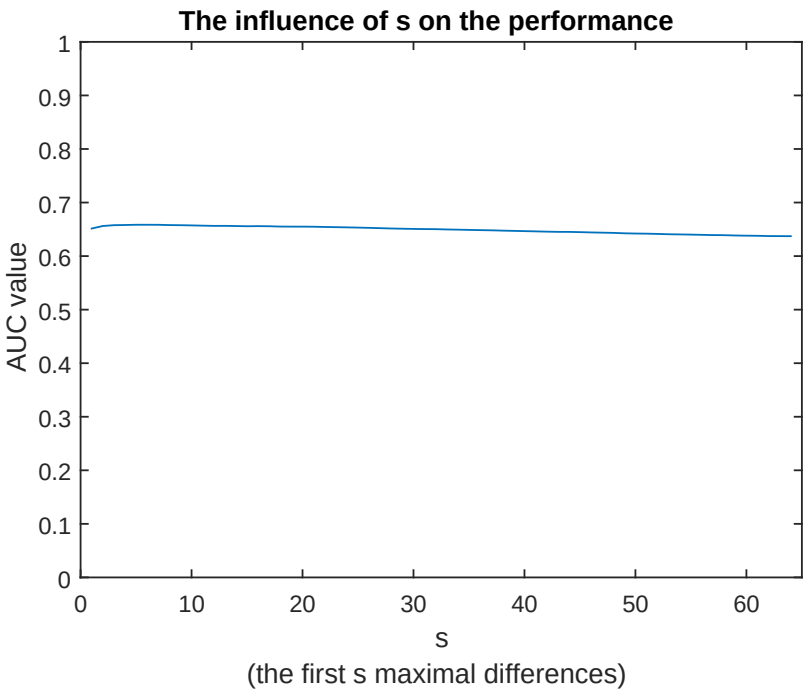

Figure 13: The influence of s in Eq. 15 on the performance of MDMD.

Acknowledgements

This paper is supported in part by grants from the National Natural Science Foundation of China (61379095, 61375009) and the Beijing Natural Science Foundation (4152055).

[1] N. Michael, M. Dilsizian, D. Metaxas, J. K. Burgoon, Motion profiles for deception detection using visual cues, in: Computer Vision–ECCV 2010, Springer, 2010, pp. 462–475.

[2] P. Ekman, W. V. Friesen, Nonverbal leakage and clues to deception, Psychiatry 32 (1) (1969) 88–106.

[3] P. Ekman, Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage (Revised Edition), WW Norton & Company, 2009.

[4] C. Darwin, The expression of the emotions in man and animals, Vol. 526, University of Chicago press, 1965.

[5] A. Mehrabian, Communication without words, Psychological today 2 (1968) 53–55.

[6] E. A. Haggard, K. S. Isaacs, Micromomentary facial expressions as indicators of ego mechanisms in psychotherapy, Springer, 1966.

[7] P. Ekman, W. Friesen, Facial action coding system.

[8] P. Ekman, Lie catching and microexpressions, The philosophy of deception (2009) 118–133.

[9] G. Warren, E. Schertler, P. Bull, Detecting deception from emotional and unemotional cues, Journal of Nonverbal Behavior 33 (1) (2009) 59–69.

[10] T. A. Russell, E. Chu, M. L. Phillips, A pilot study to investigate the effectiveness of emotion recognition remediation in schizophrenia using the micro-expression training tool, British Journal of Clinical Psychology 45 (4) (2006) 579–583.

[11] T. A. Russell, M. J. Green, I. Simpson, M. Coltheart, Remediation of facial emotion perception in schizophrenia: concomitant changes in visual attention, Schizophrenia research 103 (1) (2008) 248–256.

[12] M. Swart, R. Kortekaas, A. Aleman, Dealing with feelings: characterization of trait alexithymia on emotion regulation strategies and cognitive-emotional processing, PLoS One 4 (6) (2009) e5751.

[13] P. A. Stewart, B. M. Waller, J. N. Schubert, Presidential speechmaking style: Emotional response to micro-expressions of facial affect, Motivation and Emotion 33 (2) (2009) 125–135.

[14] S. Polikovsky, Y. Kameda, Y. Ohta, 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.

[15] T. Pfister, X. Li, G. Zhao, M. Pietikäinen, Recognising spontaneous facial micro-expressions, in: Computer Vision (ICCV), 2011 IEEE International Conference on, IEEE, 2011, pp. 1449–1456.

[16] T. Pfister, X. Li, G. Zhao, 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, pp. 868–875.

[17] S.-J. Wang, H.-L. Chen, W.-J. Yan, Y.-H. Chen, X. Fu, Face recognition and micro-expression recognition based on discriminant tensor subspace analysis plus extreme learning machine, Neural processing letters 39 (1) (2014) 25–43.

[18] S.-J. Wang, W.-J. Yan, X. Li, G. Zhao, X. Fu, Micro-expression recognition using dynamic textures on tensor independent color space, in: Pattern Recognition (ICPR), 2014 22nd International Conference on, IEEE, 2014, pp. 4678–4683.

[19] S.-J. Wang, W.-J. Yan, X. Li, G. Zhao, C.-G. Zhou, X. Fu, M. Yang, J. Tao, Micro-expression recognition using color spaces, IEEE Transactions on Image Processing 24 (12) (2015) 6034–6047. doi:10.1109/TIP.2015.2496314.

[20] S.-J. Wang, W.-J. Yan, G. Zhao, X. Fu, C.-G. Zhou, Micro-expression recognition using robust principal component analysis and local spatiotemporal directional features, in: Computer Vision-ECCV 2014 Workshops, Springer, 2014, pp. 325–338.

[21] M. Shreve, S. Godavarthy, V. Manohar, D. Goldgof, S. Sarkar, Towards macro-and micro-expression spotting in video using strain patterns, in: Applications of Computer Vision (WACV), 2009 Workshop on, IEEE, 2009, pp. 1–6.

[22] M. Shreve, S. Godavarthy, D. Goldgof, S. Sarkar, 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, pp. 51–56.

[23] M. J. Black, P. Anandan, The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields, Computer vision and image understanding 63 (1) (1996) 75–104.

[24] S. Polikovsky, Y. Kameda, Y. Ohta, Detection and measurement of facial micro-expression characteristics for psychological analysis, Kameda's Publication 110 (2010) 57–64.

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

[26] Z. Xia, X. Feng, J. Peng, X. Peng, G. Zhao, Spontaneous micro-expression spotting via geometric deformation modeling, Computer Vision and Image Understanding 147 (2016) 87–94.

[27] A. K. Davison, M. H. Yap, C. Lansley, Micro-facial movement detection using individualised baselines and histogram-based descriptors, in: Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on, IEEE, 2015, pp. 1864–1869.

[28] X. Li, T. Pfister, X. Huang, G. Zhao, M. Pietikainen, A spontaneous micro-expression database: Inducement, collection and baseline, in: Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on, IEEE, 2013, pp. 1–6.

[29] W.-J. Yan, Q. Wu, Y.-J. Liu, S.-J. Wang, X. Fu, 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, pp. 1–7.

[30] W.-J. Yan, X. Li, S.-J. Wang, G. Zhao, Y.-J. Liu, Y.-H. Chen, X. Fu, Casme ii: An improved spontaneous micro-expression database and the baseline evaluation, PloS one 9 (1).

[31] S.-J. Wang, S. Wu, X. Fu, A main directional maximal difference analysis for spotting micro-expressions, in: Workshop on The 13th Asian Conference on Computer Vision, 2016.

[32] M. F. Valstar, M. Pantic, Fully automatic recognition of the temporal phases of facial actions, Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 42 (1) (2012) 28–43.

[33] A. Asthana, S. Zafeiriou, S. Cheng, M. Pantic, Robust discriminative response map fitting with constrained local models, in: Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, IEEE, 2013, pp. 3444–3451.

[34] T. Senst, V. Eiselein, T. Sikora, Robust local optical flow for feature tracking, Circuits and Systems for Video Technology, IEEE Transactions on 22 (9) (2012) 1377–1387.

[35] B. D. Lucas, T. Kanade, et al., An iterative image registration technique with an application to stereo vision., in: IJCAI, Vol. 81, 1981, pp. 674–679.

[36] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, W. A. Stahel, Robust statistics: the approach based on influence functions, Vol. 114, John Wiley & Sons, 2011.

[37] J.-Y. Bouguet, Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm, Intel Corporation 5 (2001) 1–10.

[38] Y.-J. Liu, J.-K. Zhang, W.-J. Yan, S.-J. Wang, G. Zhao, X. Fu, A main directional mean optical flow feature for spontaneous micro-expression recognition, IEEE Transactions on Affective Computing PP (99) (2015) 1–1. doi:10.1109/TAFFC.2015.2485205.

[39] S.-J. Wang, W.-J. Yan, T. Sun, G. Zhao, X. Fu, Sparse tensor canonical correlation analysis for micro-expression recognition, Neurocomputing 214 (2016) 218 – 232. doi:http://dx.doi.org/10.1016/j.neucom.2016.05.083.
URL http://www.sciencedirect.com/science/article/pii/S0925231216305501