# A Delaunay-Based Temporal Coding Model for Micro-expression Recognition

Zhaoyu Lu, Ziqi Luo, Huicheng Zheng, Jikai Chen, Weihong Li

School of Information Science and Technology, Sun Yat-sen University, 510006, Guangzhou, China
zhenghch@mail.sysu.edu.cn, {luzhaoyu, luoziqi, chenjik, liweih3}@ mail2.sysu.edu.cn

**Abstract.** Micro-expression recognition has been a challenging problem in computer vision research due to its briefness and subtlety. Previous psychological study shows that even human being can only recognize micro-expressions with low average recognition rates. In this paper, we propose an effective and efficient method to encode the micro-expressions for recognition. The proposed method, referred to as Delaunay-based temporal coding model (DTCM), encodes texture variations corresponding to muscle activities on face due to dynamical micro-expressions. Image sequences of micro-expressions are normalized not only temporally but also spatially based on Delaunay triangulation, so that the influence of personal appearance irrelevant to micro-expressions can be suppressed. Encoding temporal variations at local subregions and selecting spatial salient subregions in the face area escalates the capacity of our method to locate spatiotemporally important features related to the micro-expressions of interest. Extensive experiments on publicly available datasets, including SMIC, CASME, and CASME II, verified the effectiveness of the proposed model.

## 1 Introduction

Expression recognition has important applications in human computer interaction and psychological study. Facial expressions can be broadly categorized into normal (macro-) and micro-expressions [5]. Usually, normal facial expressions provide rich information about the emotions of human being. But macro-expressions get ineffective when people are hiding their true emotions with faked expressions or do not show macro facial activities at all. On the contrary, micro-expressions often reveal true intents or hidden emotions corresponding to brief and subtle facial activities, which are often hard to be concealed. Ekman noticed the difference between micro-expressions and normal expressions, and started to study micro-expressions in the 1990's [5]. Micro-expressions are rapid, subtle, and involuntary facial expressions which tend to be concealed when people communicate with each other [7]. These micro-expressions reflect the true feelings. But it is hard for humans to notice or recognize such expressions when they take place.

So far, there are few studies on micro-expression recognition and numerous challenges are still to be solved. Firstly, micro-expressions are subtle, which means that the facial activities would be minute and sparse, leading to difficulties of extracting features related to micro-expressions. For example, due to the subtleness, the extracted features are often dramatically influenced by the personal variance of appearance and facial movement irrelevant to expressions. Secondly, because micro-expressions often last for no more than 1/5 of a second [17], there are only a limited number of frames which contain the facial movements corresponding to micro-expressions captured in a video. It is challenging to recognize such facial activities, which are subtle and brief whereas complex and rich in emotional information.

In this paper, we propose the Delaunay-based temporal coding model (DTCM) to address the challenges in micro-expression recognition. In the proposed model, the image sequences containing micro-expressions are normalized not only temporally, but also spatially based on the Delaunay triangulation, to remove the influence of personal appearance on micro-expression recognition. In view of the sensitiveness of locations of facial feature points to noise for micro-expressions, we consider salient coding of texture variations in the subregions generated by triangulation for representation of micro-expressions. Instead of considering fixed areas of the face as in [19], we select subregions related to micro-expressions based on magnitudes of local texture fluctuations, which corresponds to spatial salient coding of local features and handles sparse and subtle facial movements well. For classification, we implement random forest (RF) [2] and support vector machine (SVM) [15]. Extensive experimental results on publicly available SMIC, CASME, and CASME II datasets verified the effectiveness of the proposed method.

The rest of this paper is organized as follows. Section 2 reviews representative related work. In Section 3, the overview of the proposed model is presented. Section 4 presents the coding space of the proposed model, including temporal and spatial normalization of the image sequences corresponding to micro-expressions, and construction of the coding space based on Delaunay triangulation. In Section 5, we explain how we carry out temporal coding at local subregions and select spatially salient subregions. Extensive experimental results are demonstrated in Section 6. This paper is concluded by Section 7.

## 2 Related Work

Micro- and macro-expression recognition are closely related. Research under one case often inspires that under the other case. Relatively more extensive research has been done for macro-expression recognition, which mainly focus on selection of emotion-related facial areas or extraction of features related to expressions.

Hamm et al. proposed a system for analysis of human facial movements, referred to as facial action coding system (FACS), which was successfully used in automatic expression recognition [9]. Micro-expressions are not considered in FACS. Moreover, it is not convenient to apply the system for expression recognition in practice, since action units in the system have to be marked manually. In [16] and [19], regions of interest related to facial expressions are selected from face images for recognition. However, the selected regions are fixed, ignoring distinct distributions among different micro-expressions.

There are mainly two strategies for expression representation, namely static approaches [3, 13] and dynamic approaches [10, 14, 18, 20, 23]. Static approaches only focus on the peak frame of an image sequence and do not exploit the rich information in the dynamical interactions of facial muscles. On the other hand, dynamic approaches extract expression-related features from the full image sequences. For example, in [10], a method based on nonrigid registration using free-form deformations was proposed to represent the texture variation of action units [9] in the spatial-temporal domain. Wang et al. [20] introduced a temporal Bayesian network to capture complex spatiotemporal relations among facial muscles for expression recognition.

For micro-expression recognition, Pfister et al. [14] proposed a temporal interpolation model (TIM) to generate sufficient frames from an image sequence. They utilize the active shape model to detect the facial feature points, and implement local binary patterns extracted from three orthogonal planes (LBP-TOP) [23] to describe the spatiotemporal local textures. In later research, LBP-TOP is combined with tensor independent color space (TICS) and demonstrates better recognition results [19].

In this paper, we apply Delaunay triangulation and standard deviation analysis to locate facial sub-areas related to micro-expressions, which is very effective due to its adaptability to different micro-expressions. We encode the variations of textures instead of the movements of feature points in traditional methods such as [20], since micro-expressions are subtle and feature points are very likely to be static or sensitive to noises. As verified by extensive experiments, the proposed model provides a very effective representation of micro-expressions and leads to promising accuracies for micro-expression recognition.

## 3 Overview of the Proposed Model

Figure 1 illustrates the basic steps in the proposed method. First, image sequences containing micro-expressions are normalized in the time domain by using the temporal interpolation model [14]. Then Delaunay triangulation [1] and mapping [6] is implemented according to the landmarks generated with the active appearance model (AAM) [4] fitting. The face area is normalized spatially based on Delaunay triangulation to avoid the disastrous influence of personal appearance variances. By dividing the overall face area into a collection of local subregions, rich discriminative local descriptors may be extracted for micro-expression representation. More specifically, we extract texture features from the generated Delaunay triangles. We analyze the variations of Delaunay triangles and select those potentially related to micro-expressions according to the magnitudes of local variations. The selected Delaunay triangles are encoded in the time domain to preserve temporal information in micro-expression representation. Finally, the micro-expressions are recognized by using RF or SVM classifiers.

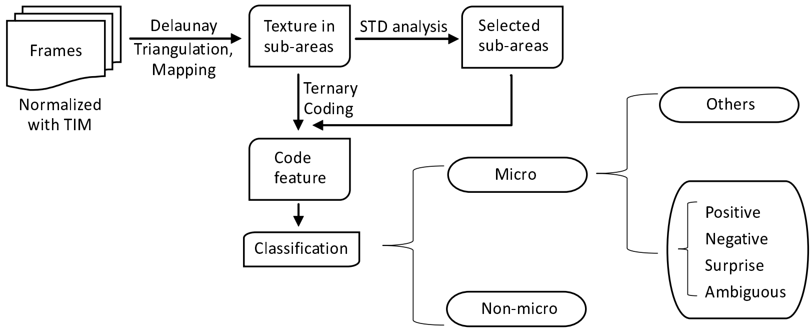



Fig. 1: Overview of the proposed model.

## 4 Coding Space for DTCM

In this section, we construct the spatiotemporal coding space for the proposed DTCM, which will be used later for representation of micro-expressions in image sequences.

### 4.1 Temporal Normalization

To remove temporal fluctuations of micro-expressions, which may introduce noise irrelevant to classification of micro-expressions, we apply TIM [14] to normalize image sequences corresponding to micro-expressions temporally. TIM is a manifold-based interpolation method, which builds a low-dimensional embedding of an image sequence, and then interpolates a curve in the low-dimensional space. The interpolated frames are mapped back to the original high-dimensional space to form the temporally normalized image sequence.

### 4.2 Delaunay Triangulation and Mapping

In order to extract rich description of facial expressions, we define 68 feature points as in [8]. The AAM is implemented to obtain these feature points in a frame with a neutral face from the image sequence. The AAM exploits both texture and shape information to trace facial feature points with a model building phase and a face fitting phase. The fitting process iterates until the Euclidean distance between the model texture vector and the instance texture vector converges. It provides satisfactory feature points for micro-expression recognition based on our method.

Since micro-expressions are revealed by subtle muscle movements in the face, we focus on variations of the image sequence in the facial area. In traditional methods, subregions remain fixed and not adaptive to different faces, which would introduce expression-irrelevant noise. To have a detailed and descriptive representation of the facial expression, we implement Delaunay triangulation based on the point set in the facial area. The Delaunay triangulation is unique

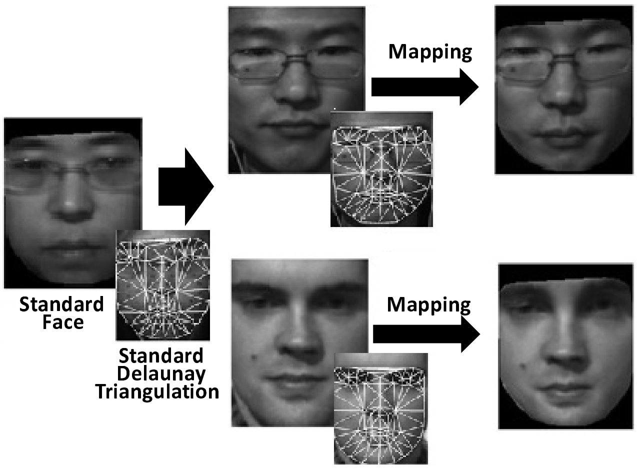



Fig. 2: Delaunay triangulation and mapping. With a standard face, all face images are mapped into the same domain defined by 68 feature points.

for a given set of feature points in general position and there is no other feature point other than the three vertices in the circumcircle of each triangle [12]. The uniqueness with respect to the feature points can be exploited to normalize all faces into a standard space. In traditional methods, subregions remain fixed and not adaptive to different faces, which would introduce expression-irrelevant noise. The triangulation process divides the overall facial area into a number of triangular subregions based on the previously detected feature points. We denote the $n$ sub-areas by $P = \{p_1, p_2, \ldots, p_n\}$. The texture variations in the subregions signal the movements of facial muscles and can be used to characterize micro-expressions.

As micro-expressions are subtle, the landmarks corresponding to feature points rarely change in the image sequence. Based on this observation, we propose to focus on the variations of local textures in the triangles instead of the feature points themselves.

To normalize all faces into a standard space and remove the influence of personal appearances on micro-expression recognition, triangular mapping is applied to all the images in the sequence. As illustrated in Fig. 2, a standard Delaunay triangulation is first determined based on a chosen standard face. With this standard triangulation, other face images with different appearances will be mapped to the same triangulation. In this way, personal appearance difference irrelevant to micro-expressions will be largely removed. After normalization based on the triangular mapping, all face images will have the same amount of pixels in each triangular subregion. The standard face can be a neutral face of any subject in the dataset. In our experiments, we simply choose the neutral face in the first sequence of the first subject for each dataset.

### 4.3 Construction of the Coding Space

Let $F = \{f_1, f_2, \ldots, f_t\}$ be the collection of $t$ frames in the temporally normalized image sequence and $p_{i,j}$, with $j = 1, 2, \ldots, t$, be the subregion $p_i$ in the $j$-th frame. We define the differences between neighboring texture signatures corresponding to $p_{i,j-1}$ and $p_{i,j}$ as the local temporal variations (LTVs) $X = \{x_{i,j} | i = 1, 2, \ldots, n; j = 2, 3, \ldots, t\}$. A spatiotemporal coding space $\Gamma = \{F, P, X\}$ is then constructed. In this paper, we simply choose the grayscale values in each subregion $p_{i,j}$ as the texture feature. For micro-expressions, the facial movements are subtle. Our primary purpose is to describe the weak variations of the expression-related subregions and construct strong descriptors out of the weak ones efficiently. The temporal grayscale variation appears to be a good feature for subtle facial variations according to our experiments. More complex textual features may easily contain a lot of expression-irrelevant information such as the appearance, which overshadows real facial movements.

## 5 Micro-expression Coding Based on DTCM

This section explains how we encode a micro-expression using DTCM based on the previously constructed coding space.

### 5.1 Extraction of Local Temporal Variations

The LTV $x_{i,j}$ reflects the difference between the texture features of neighboring frames in the subregion $p_i$. Let $T_{i,j} \in R^m$ be an $m$-dimensional texture vector extracted from the $j$-th frame in the $i$-th subregion $p_{i,j}$. In this paper, we define the LTV as follows,

$$x_{i,j} = \frac{\sum_{k=1}^{m} (T_{i,j,k} - T_{i,j-1,k})}{m} \tag{1}$$

where $T_{i,j,k}$ denotes the $k$-th element in $T_{i,j}$.

In this paper, $T_{i,j}$ consists of all pixel gray values in the subregion $p_{i,j}$. So $m$ is simply the number of pixels in $p_{i,j}$, and $x_{i,j}$ the difference between mean values of the subregion $p_i$ computed from neighboring frames. The extracted LTVs capture the overall variations in the corresponding subregions between nearby frames. Specifically, a positive $x_{i,j}$ signals the overall lightening in the corresponding subregion and a negative one signals the overall darkening.

### 5.2 Adaptive Threshold and Ternary Coding

The facial movement amplitude, indicated by the value of $x_{i,j}$, is sensitive to noise. To improve the robustness, we only preserve local temporal variations with significant magnitudes, which corresponds to the idea of salient coding. Let $x_{i,j,r}$ be LTVs of the $r$-th sequence, where $r = 1, \ldots, R$. We define $x^{+}_{i,j,r} =$

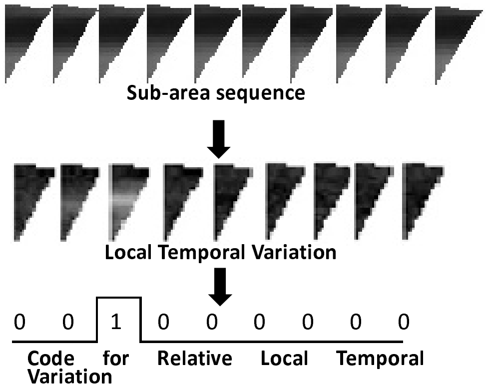



Fig. 3: Coding process for the local temporal variations.

$\{x_{i,j,r}|x_{i,j,r} > 0\}$ and $x^-_{i,j,r} = \{x_{i,j,r}|x_{i,j,r} < 0\}$. Then a positive and a negative thresholds are defined for each subregion $p_i$ as follows,

$$\tau_i^+ = \frac{a}{R}\sum_{r=1}^{R} \text{mean}\{x^+_{i,j,r}\} \tag{2}$$

$$\tau_i^- = \frac{a}{R}\sum_{r=1}^{R} \text{mean}\{x^-_{i,j,r}\} \tag{3}$$

where mean($\cdot$) denotes the mean value of the set, and $a$ is a positive real number to be chosen appropriately. The thresholds defined in (2) and (3) are adaptive for the subregions to cope with spatially-varying noise.

We then propose a ternary coding method to encode the LTVs for representation of micro-expressions by defining

$$c_{i,j} = \begin{cases} 1, & x_{i,j} \geqslant \tau_i^+ \\ 0, & \tau_i^- < x_{i,j} < \tau_i^+ \\ -1, & x_{i,j} \leqslant \tau_i^- \end{cases} \tag{4}$$

The code values stand for three different relative variations in texture, namely, lightening, remaining, and darkening, which can be used to describe significant variations in the subregions. With the salient coding method, noise corresponding to illumination variations and inaccurate alignments can be removed. It is also useful for reducing the influence of personal appearances. The coding process is illustrated in Fig. 3.

### 5.3 Feature Selection

With Delaunay triangulation, abundant subregions will be generated, which leads to a large number of local features. Some of these local features may be

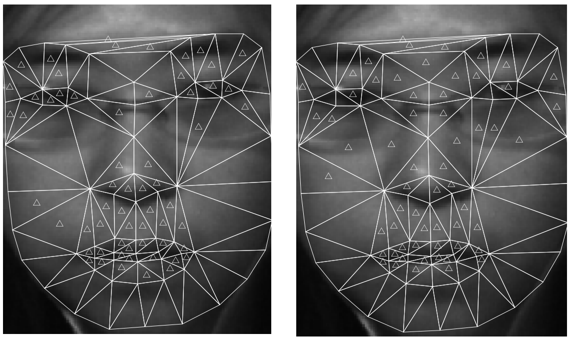


Fig. 4: Top 60 subregions determined as most related to the corresponding micro-expressions based on STD analysis. The small triangles mark the identified subregions. The subregions most related to the corresponding negative and positive micro-expressions are indicated in the left and right pictures, respectively.

noisy and irrelevant to micro-expressions. Therefore, we introduce a feature selection process in our model to determine subregions related to micro-expressions. The approach is based on analysis of the standard deviation (STD) of $x_{i,j}$, which reveals the saliency of a subregion relevant to micro-expressions. Specifically, the proposed STD analysis allows feature selection without a data-dependent pre-training stage, which is prone to overfitting as observed in traditional discriminative methods. Furthermore, the proposed method allows the selected features to be adaptive to specific input instances. The personal STD is computed in each subregion and for each image sequence individually as

$$\mathrm{STD}_i = \sqrt{\frac{1}{t-1}\sum_{j=2}^{t}\left(x_{i,j} - \frac{1}{t-1}\sum_{k=2}^{t} x_{i,k}\right)^2} \tag{5}$$

The magnitude of $\mathrm{STD}_i$ reflects the strength of facial movement in the corresponding facial subregion. We select $N$ subregions with the largest STD values, which are assumed to be mostly related to the micro-expressions. Other subregions with lower STD values are assumed to be irrelevant to micro-expressions and coded as zeros. Figure 4 shows two examples where 60 subregions determined as most related to the corresponding positive or negative micro-expression are identified based on the STD analysis.

We concatenate the code sequences of all subregions in the image sequence into one feature vector as representation of the micro-expression. Then the feature vector is classified by using the RF or SVM classifier, which have demonstrated good performance when dealing with high-dimensional data.

Table 1: The results of separating micro and non-micro expressions

| Method | Accuracy (%) |
|---|---|
| LBP-TOP+TIM10+RF [14] | 74.3 |
| DTCM+RF (Test 1) | 79.80 |
| DTCM+RF (Test 2) | 85.86 |
| DTCM+RF (Test 3) | **88.89** |

## 6 Experimental Results

In this section, we test the proposed DTCM on three publicly available micro-expression datasets, including SMIC, CASME, and CASME II to verify its effectiveness for micro-expression recognition. These datasets consist of spontaneous micro-expressions which appear in real life. For each dataset, we first compare the proposed method to state-of-the-art methods in terms of the micro-expression recognition rates. Then we investigate the influence of the parameters of DTCM on the recognition results. Leave-one-subject-out cross validation is used in all experiments.

### 6.1 Results on SMIC

We first verify the performance of the proposed method on the SMIC dataset. The image sequences taken with a high speed camera are used in the experiments. This dataset consists of 70 sequences of negative micro-expressions, 51 sequences of positive micro-expressions, and 164 sequences without micro-expressions. All the sequences last for no more than 1/2 second, which means that the number of frames are less than 50 for a sequence captured with a camera at 100fps [11].

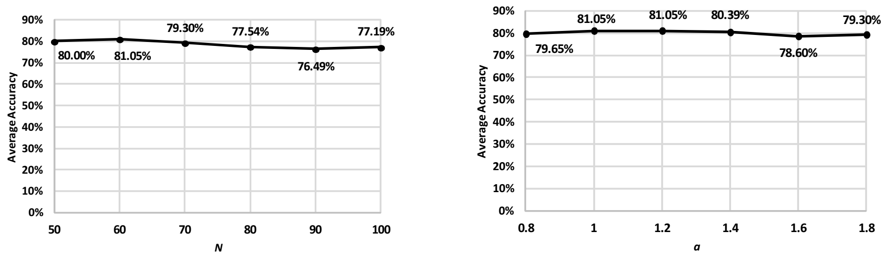
We follow the same experimental protocol as in [14], which combines TIM and LBP-TOP for recognizing micro-expressions on the same dataset. More specifically, we randomly select 77 image sequences without micro-expressions, 18/17 sequences containing negative/positive micro-expressions from the SMIC dataset. We repeat the random selection process several times and record the recognition results.

Tables 1 shows the results of separating micro-expressions from non-micro-expressions, where RF refers to the random forest algorithm. The parameters $a = 1.0$ and $N = 60$ in the proposed method. As we can see, in several random runs of the proposed method, the proposed method obtains better results than that reported in [14]. Table 2 illustrates the results of classifying positive and negative micro-expressions, where SVM (SMO) refers to support vector machine with sequential minimization optimization. Linear SVM is implemented in the experiments. The parameters $a = 1.0$ and $N = 70$ in the proposed method. Again, in several random runs, the proposed method reports higher accuracies than the method based on LBP-TOP+TIM [14].

There are mainly two parameters in the proposed method, namely $a$ for the thresholds $\tau_i^+$ and $\tau_i^-$ defined in (2) and (3), and $N$ for the number of selected

Table 2: Classification of positive and negative micro-expressions

| Method | Accuracy (%) |
|---|---|
| LBP-TOP+TIM10+MKL [14] | 71.4 |
| DTCM+SVM (SMO) (Test 1) | 74.3 |
| DTCM+SVM (SMO) (Test 2) | 80 |
| DTCM+SVM (SMO) (Test 3) | **82.86** |



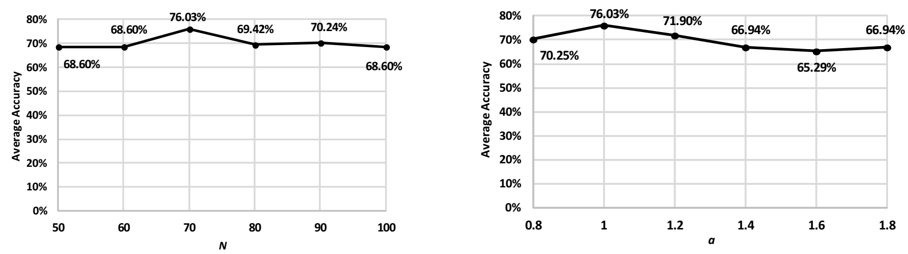

(a) (b)

Fig. 5: Influence of parameters (a) $N$ and (b) $a$ for separating micro-expressions from non-micro-expressions.

subregions related to expressions in Section 5.3. $a$ controls the thresholds for determining if local temporal variations are sufficiently salient. Small values of $a$ lead to higher sensitivity to temporal noises while large values may suppress useful temporal information. The number of selected local regions $N$ controls the sensitivity to noise in the spatial domain. Small values of $N$ can suppress most spatial noise but could also lose useful information. On the other hand, large values of $N$ would be sensitive to noise. To investigate the influence of these parameters in practice, we used all videos including 51 sequences with positive micro-expressions, 70 sequences with negative micro-expressions, and 164 sequences with non-micro-expressions from the SMIC dataset.

Figure 5a shows the influence of $N$ for separating micro-expressions from non-micro-expressions, where $a$ is fixed at 1.0. As we can see, for $N$ from about 50 to 70, which corresponds to about $45 - 64\%$ of the number of local regions in the face area, the proposed model shows reasonably good classification accuracies. Figure 5b illustrates the influence of $a$ in the experiments, where $N$ is fixed at 60. In the tested range of values, the proposed method shows promising results. Generally, a value of $a$ between 0.8 and 1.4 seems to be a good choice.

For the experiments of separating positive and negative micro-expressions, the influence of the parameters $N$ and $a$ are illustrated in Figure. 6a ($a$ is fixed at 1.0) and Figure. 6b ($N$ is fixed at 70), respectively. It seems that the results are sensitive to the choice of $N$ and $a$. This is probably due to the fact that separating two types of micro-expressions is more challenging than sepa-




(a) (b)

Fig. 6: Influence of parameters (a) $N$ and (b) $a$ for separating positive and negative micro-expressions.

rating micro-expressions from non-micro-expressions. Nevertheless, the proposed method obtained reasonable results in the tested range of $N$. A choice of around $N = 70$ gives the best results. For the parameter $a$, the performance is best for $a$ between 0.8 and 1.2, consistent with the conclusion drawn from Figure. 5b.

To verify the effectiveness of Delaunay triangulation, we conduct another experiment for comparison, where Delaunay triangulation is replaced by small-grid-based segmentation as in [14]. Specifically, an image is split into $10 \times 10$ square regions and top 20 regions are selected with STD analysis, which correspond to about the same area as 60 salient Delaunay triangles. The accuracy drops to 63.6% for positive and negative micro-expression classification. This is because the grid-based segmentation is fixed and can not be adapted to different faces. Due to different facial appearances, subregions located at the same coordinate may not correspond to the same semantic region of different people.

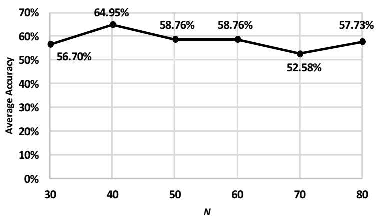
### 6.2 Results on CASME

Experiments in this section are based on sequences in Class B of the CASME dataset, which were recorded by Point Grey GRAS-03K2C camera at 60fps [22]. The dataset contains eight categories of micro-expressions. Since some categories contain very few samples, we follow the experimental settings in [19] and carry out four-category classification on this dataset. The first category is positive, which corresponds to the happiness micro-expression. The second category is negative, which consists of disgust, sadness, and fear micro-expressions. The third category is surprise, and the last category contains ambiguous micro-expressions, including tense and repression. The four categories contain 4, 47, 13, and 33 samples, respectively.

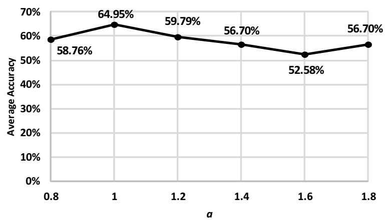
Table 3 shows the results of four-category micro-expression recognition on CASME. The parameters $a = 1.0$ and $N = 40$ in the proposed method. As we can see, under the same experimental protocol, we achieve a higher recognition rate than the method combining LBP-TOP, TIM and TICS [19]. When Delaunay triangulation is replaced by grid-based segmentation ($10 \times 10$ squares), the accuracy drops by 12.37%.

Table 3: Results of four-category micro-expression recognition on CASME

| Method | Accuracy (%) |
|---|---|
| LBP-TOP+TIM70+TICS+SVM [19] | 61.86 |
| DTCM (Grid segmentation) | 52.58 |
| DTCM+RF | **64.95** |




(a)




(b)

Fig. 7: Influence of parameters (a) $N$ and (b) $a$ for four-category micro-expression recognition on CASME.

Figures 7a and 7b show the influence of the parameters $N$ and $a$ in the proposed method, where $a$ and $N$ are fixed at 1.0 and 40, respectively. As we can see, a choice of $N$ around 40 gives the best result, which corresponds to about 36% of the subregions in the face area. For the parameter $a$, a choice between 0.8 and 1.2 would still be recommended.

### 6.3 Results on CASME II

CASME II [21] contains high-quality sequences captured by a Point Grey GRAS-03K2C camera at 200fps, which consists of five categories of micro-expressions, including happiness, surprise, disgust, repression, and others. There are 32, 25, 64, 27, and 99 samples in these categories, respectively.

Wang et al. [19] reported an accuracy of 58.54% in the five-category classification. However, since "others" contains a mixture of different facial activities, in our experiments, this category is singled out first by our method, with a success rate of 72.06%. Our experiments show that the approach is not very sensitive to the number of selected subregions $N$. For the parameter $a$, a choice between about 0.8 and 1.2 is still recommended. Setting $a = 0.8$ and $N = 50$ leads to the overall best accuracy 72.06%. In the experiments of four-category micro-expression recognition, with $a$ fixed at 1.0 and $N$ around 70 (which corresponds to about 63% of the facial subregions), the best recognition accuracy 64.19% is obtained. For the parameter $a$, we observe that when it is around 1.0, a good recognition accuracy can be obtained, consistent with conclusions in the previous experiments.

## 7 Conclusions

In this paper, we propose a method for micro-expression recognition with Delaunay-based temporal coding (DTCM). An arbitrary sequence of micro-expression is normalized not only temporally, but also spatially based on Delaunay triangulation, which aims to remove the influence of personal appearance on recognition. It is especially meaningful for micro-expressions, which are subtle and brief, and can easily be concealed by appearances irrelevant to the expressions of interest. Furthermore, we propose to consider texture information in the subregions generated by the triangulation instead of the fitted feature points, since for micro-expressions, locations of feature points are not discriminative enough and tend to be dominated by noise. The features extracted from local subregions are then coded temporally base on local variations, which captures the saliency distribution in the time domain. Furthermore, spatial selection of local subregions are carried out to extract spatial saliency distribution information. Extensive experimental results on public datasets including SMIC, CASME, and CASME II demonstrates the effectiveness of the proposed method for micro-expression recognition.

**Acknowledgement**. This work is supported by National Natural Science Foundation of China (No. 61172141), Key Projects in the National Science & Technology Pillar Program during the 12th Five-Year Plan Period (No. 2012BAK16B06), Science and Technology Program of Guangzhou, China (2014J4100092), and National Undergraduate Scientific and Technological Innovation Project of China.

## References

1. Barber, C., Dobkin, D., Venue, H.: The quickhull algorithm for convex hulls. ACM Transactions on Mathematical Software **22** (1996) 469–483
2. Breiman, L.: Random forests. Machine Learning **45** (2001) 5–32
3. Chu, W., Torre, F., Cohn, J.: Selective transfer machine for personalized facial action unit detection. In: International Conference on Computer Vision. (2013) 3515–3522
4. Cootes, T., Edwards, G., Taylor, C.: Active appearance models. IEEE Transactions on Pattern Analysis and Machine Intelligence **23** (2001) 681–685
5. Ekman, P.: Facial expressions of emotion: An old controversy and new findings. Philosophical Transactions of The Royal Society of London, Series B: Biological Sciences **335** (1992) 63–69
6. Fournier, A., Montuno, D.: Triangulating simple polygons and equivalent problems. ACM Transactions on Graphics **3** (1984) 153–174
7. Frank, M., Herbasz, M., Sinuk, K., Keller, A., Nolan, C.: I see how you feel: Training laypeople and professionals to recognize fleeting emotions. In: The Annual Meeting of the International Communication Association, New York. (2013) 3515–3522
8. Gross, R., Matthews, I., Cohn, J., Kanade, T., Baker, S.: Multi-PIE. Image and Vision Computing **28** (2010) 807–813
9. Hamm, J., Verma, R., Kohler, C., Gur, R.: Automated facial action coding system for dynamic analysis of facial expressions in neuropsychiatric disorders. Journal of Neuroscience Methods **200** (2011) 237–256

10. 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** (2010) 1940–1954
11. Li, X., Pfister, T., Huang, X., Zhao, G., Pietikainen, M.: A spontaneous micro-expression database: Inducement, collection and baseline. In: IEEE Conference on Automatic Face and Gesture Recognition. (2013) 1–6
12. Mark, B., Cheong, O., Kreveld, M., Overmars, M.: Computational Geometry: Algorithms and Applications. Springer-Verlag, Berlin (2008)
13. Park, S., Kim, D.: Spontaneous facial expression classification with facial motion vectors. In: IEEE Conference on Automatic Face and Gesture Recognition. (2008) 1–6
14. Pfister, T., Li, X., Zhao, G., Pietikainen, M.: Recognising spontaneous facial micro-expressions. In: International Conference on Computer Vision. (2011) 1449–1456
15. Platt, J.: Sequential minimal optimization: A fast algorithm for training support vector machines. Microsoft Research, MSR-TR-98-14 (1998) 1–21
16. Polikovsky, S., Kameda, Y., Ohta, Y.: Facial micro-expressions recognition using high speed camera and 3D-gradient descriptor. In: International Conference on Distributed Platforms. (2009) 1–6
17. Shen, X.B., Wu, Q., Fu, X.L.: Effects of the duration of expressions on the recognition of microexpressions. Journal of Zhejiang University, Science B **13** (2012) 221–230
18. Shreve, M., Godavarthy, S., Goldgof, D., Sarkar, S.: Macro-and micro-expression spotting in long videos using spatio-temporal strain. In: IEEE Conference on Automatic Face and Gesture Recognition. (2011) 51–56
19. Wang, S., Yan, W., Li, X., Zhao, G., Fu, X.: Micro-expression recognition using dynamic textures on tensor independent color space. In: International Conference on Pattern Recognition. (2014)
20. Wang, Z., Wang, S., Ji, Q.: Capturing complex spatio-temporal relations among facial muscles for facial expression recognition. In: Conference on Computer Vision and Pattern Recognition. (2013) 3422–3429
21. Yan, W., Li, X., Wang, S., Zhao, G., Liu, Y., Chen, Y., Fu, X.: CASME II: An improved spontaneous micro-expression database and the baseline evaluation. PLoS ONE **9** (2014) 1–8
22. Yan, W., Wu, Q., Liu, Y., Wang, S., Fu, X.: CASME database: A dataset of spontaneous micro-expressions collected from neutralized faces. In: IEEE Conference on Automatic Face and Gesture Recognition. (2013) 1–7
23. Zhao, G., Pietikainen, M.: Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence **29** (2007) 915–928