Facial Micro-Expression Recognition using Spatiotemporal Local Binary Pattern with Integral Projection

Xiaohua Huang[1], Su-Jing Wang[2], Guoying Zhao[1], Matti Pietikäinen[1]
[1]*Center for Machine Vision Research, Department of Computer Science and Engineering*
University of Oulu, Finland
[2]*State Key Laboratory of Brain and Cognitive Science*
Institute of Psychology, Chinese Academy of Science, Beijing, China
{*xiaohua.huang, gyzhao, mkp*}*@ee.oulu.fi, wangsujing@psych.ac.cn*

Abstract

Recently, there are increasing interests in inferring mirco-expression from facial image sequences. For micro-expression recognition, feature extraction is an important critical issue. In this paper, we proposes a novel framework based on a new spatiotemporal facial representation to analyze micro-expressions with subtle facial movement. Firstly, an integral projection method based on difference images is utilized for obtaining horizontal and vertical projection, which can preserve the shape attributes of facial images and increase the discrimination for micro-expressions. Furthermore, we employ the local binary pattern operators to extract the appearance and motion features on horizontal and vertical projections. Intensive experiments are conducted on three available published micro-expression databases for evaluating the performance of the method. Experimental results demonstrate that the new spatiotemporal descriptor can achieve promising performance in micro-expression recognition.

1. Introduction

Facial expressions of emotion, especially those known as micro-expressions, amongst nonverbal behavior like gestures and voice, have received increasing attention in recent years [26, 14]. A micro-expression is a subtle and involuntary facial expression which usually occurs in high-stakes situations, where people have something to lose or gain. The importance of micro-expression study is apparent in many potential applications, such as in the security field. However, unlike regular facial expressions micro-expressions are very hard for people to detect and recognize. According to psychological research [6], the main reason is that the micro-expressions generally remain less than 0.2 seconds and are very subtle since people are trying to control and repress. In order to improve the ability of people to recognize micro-expressions, the micro-expression training tool was developed by Ekman aiming to train people to better recognize micro-expressions. However, even with its help people achieve only around 40% in recognizing micro-expressions [7]. Therefore there is a great need for a high-quality micro-expression recognition system on assisting people to accurately recognize them.

Previous studies on automatic facial micro-expression analysis primarily focused on spotting facial micro-expressions [21, 22] from macro-expressions. Recently, spontaneous facial micro-expression analysis has been received attention from numerous researchers [11, 17], since spontaneous micro-expressions can reveal genuine emotions which people try to conceal. It is very important to investigate spontaneous facial micro-expressions. The subtle change of micro-expressions causes human difficulties in perceiving micro-expressions, and also brings a serious challenge to computer vision. As a result, it requires a good approach to extract the useful information from micro-expressions with subtle change.

Geometry-based and appearance-based features have been commonly used to analyze facial expressions for facial expression recognition. Specifically, geometric-based features represent face geometry, such as the shapes and locations of facial landmarks. However, geometric features are sensitive to global changes, such as illumination variation. Instead, appearance-based features describe the skin texture of faces. In these years, Local binary pattern from three orthogonal planes (**LBP-TOP**) has demonstrated its efficiency for facial expression recognition [9, 29]. As a result, LBP-TOP has been also widely used in micro-expression analysis [17, 27, 4]. One of the earlier works proposed to use LBP-TOP for spontaneous facial micro-expression recognition method [17]. As well, Yan *et al*. [27] employed LBP-TOP on their proposed CASME 2 database to

achieve micro-expression recognition rate of 63.41%. Some other researcher used LBP-TOP to investigate the difference of micro-facial movement sequences and neutral face sequences [4]. However, there is still a gap to achieve a high-quality micro-expression analysis, although previous research results seem higher than human observation [7].

Recently, many researchers proposed different techniques to improve LBP-TOP for micro-expression recognition. For example, Ruiz-Hernandez and Pietikäinen [19] employed the re-parameterization of second order Gaussian jet on LBP-TOP achieving promising micro-expression recognition result on SMIC database [17]. Wang *et al*. [23] extracted Tensor features from Tensor Independent Colour Space (**TICS**) for micro-expression recognition, achieving promising results on CASME 2 database. Furthermore, Wang *et al*. [24] proposed to use Local Spatiotemporal Directional Features (**LSDF**) on background image obtained by using robust principal component analysis for micro-expressions. In addition, recent work [25] proposed to use six intersection points to reduce redundant information in LBP-TOP, so called **LBP-SIP**, for obtaining better performance than LBP-TOP. Even so, there is still much room for improvement in the recognition performance.

The work in [10] suggests that the shape information is more useful for facial expressions than appearance features. Moreover, the work in [8] demonstrated that local binary pattern (**LBP**) enhanced by shape information can distinguish an image with different shape from those with the same LBP feature distributions. But most of previous methods used in micro-expression recognition just consider texture information to represent face images, but ignore the shape information of face images. Image projection techniques are classical methods for pattern analysis, widely used, e.g., in motion estimation [18] and face tracking [12, 13], as they enhance shape properties and increase discrimination of images. One image projection technique, integral projection, provides simple and efficient computation as well as a very interesting set of properties. It firstly is invariant to a number of image transformations like scale and translation. It is also highly robust to white noise. Then it preserves the principle of locality of pixels and sufficient information in the process of projection. Therefore, we introduce a new spatiotemporal method based on integral projection and LBP [16, 15] to improve the performance of micro-expression recognition, in which integral projection can provide the shape property of facial images.

To explain the concepts of our approach, the paper is organized as follows. In Section 2, we explain our method for exploring the spatiotemporal features for micro-expressions. The results of applying our method for recognizing micro-expressions are provided in Section 3. Finally we summarize our findings in Section 4.

2. Proposed methodology

Recently, interesting approaches combining the integral projection and texture descriptor for bone texture characterization and face recognition were proposed [8, 1]. It shows that the properties of image projection may provide supplementary shape information to enhance the texture descriptor. However, very little research applies the integral projection to temporal domain. In this section, we propose the new spatiotemporal local binary pattern based on integral projection, namely $\underline{S}$patio$\underline{T}$emporal $\underline{L}$ocal $\underline{B}$inary $\underline{P}$attern with $\underline{I}$ntegral $\underline{P}$rojection (**STLBP-IP**), to boost the capability of LBP-TOP for micro-expression recognition.

2.1. Difference-Image based Integral projection

An integral projection produces a one-dimensional pattern, obtained through the sum of a given set of pixels along a given direction. Let $I_t(x, y)$ be the intensity of a pixel at location (x, y) and time t, the random transformation of $f_t(x, y)$ is defined as:

$$\Re[f_t](\theta, s) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I_t(x, y)\delta(x\cos\theta + y\sin\theta - s)dxdy, \tag{1}$$

where δ is a Dirac's delta, θ is a projection angle and s is the threshold value. In this work, we consider the horizontal and vertical directions. In this case, θ on Eq. 1 are 0° and 90° for horizontal and vertical directions, respectively. Thus, Eq. 1 can be re-written as:

$$V_t(x) = \frac{1}{y_2 - y_1}\int_{y_1}^{y_2} I_t(x, y)dy, \tag{2}$$

$$H_t(y) = \frac{1}{x_2 - x_1}\int_{x_1}^{x_2} I_t(x, y)dx, \tag{3}$$

where $H_t(y)$ and $V_t(x)$ represents the horizontal and vertical integral projections, respectively.

The pioneering work by Mateos et al. [12, 13] showed that integral projections can extract the common underlying structure for the same people's face images, which is more relative to face identity. However, we find that this common structure brings the noise to micro-expression recognition, since micro-expression recognition aims to extract the different change between various emotion and less relative to face identity. In order to validate our finding, we use 247 micro-expression images of five emotion classes (happiness, disgust, surprise, repression and others) on apex state from CASME2 database for analyzing whether the integral projection would miss discriminative information of micro-expressions. For 247 facial images, we first use active shape model to obtain the 68 facial landmarks, and apply local weight normalization to transform a face to a canonical frame. The face images are finally cropped and

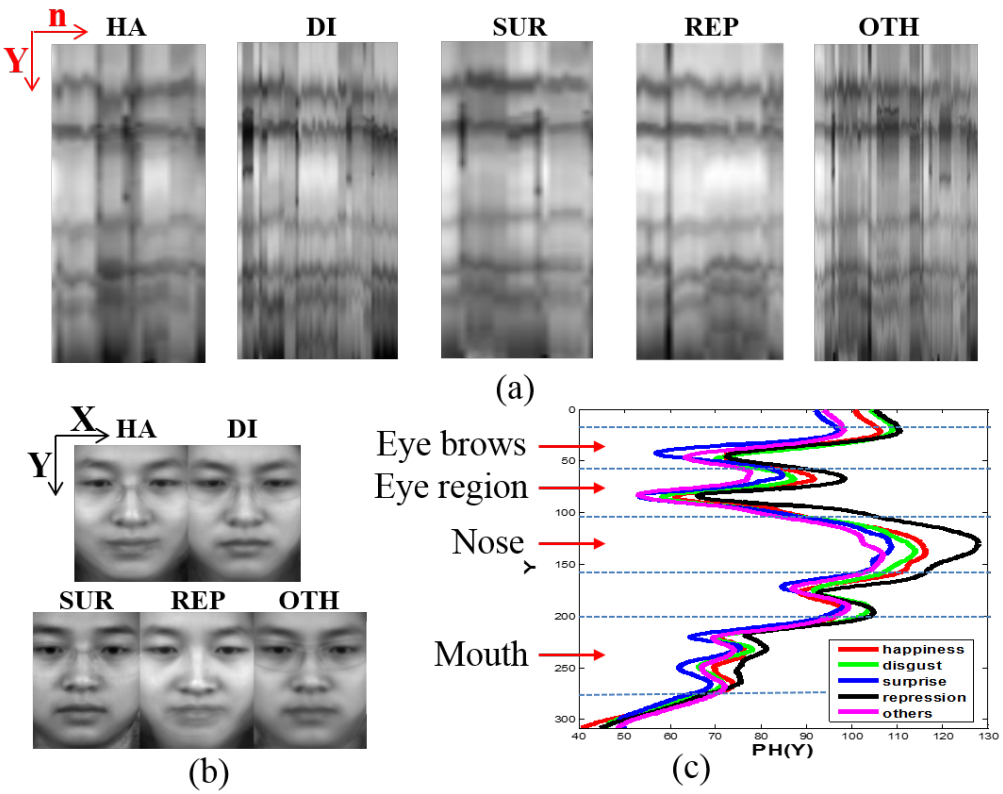


Figure 1: Horizontal integral projection based on micro-expression images. (a) The horizontal integral projections of 247 facial images (happiness (HA), disgust (DI), surprise (SUR), repression (REP) and others (OTH)), where n-axis describes the number of facial images and each projection is represented as a single column, (b) Mean face for each micro-expression, obtained by applying mean operator to facial images of the same class from CASME2 database [27], and (c) The horizontal integral projections of 5 mean faces, where each curve represents the face region, x-axis (PH(Y)) means the value of horizontal integral projection, and y-axis represents the height of an image.

normalized into the same size. Fig. 1 shows the horizontal integration of 247 facial images, five class-mean faces and the graph for horizontal integration of five class-mean faces. As seen from Fig. 1(a), it is found that the integral projection just locates the common regions of all facial images, such as mouth, eyes, but not provides discriminative information for different micro-expressions. Moreover, we present a graph explicitly to show this finding, where we obtain the statistical distribution of various classes by using integral projection on mean face for each micro-expression. As seen from Fig. 1(c), the curves for all classes have the same characteristics. For example, the curves on the mouth region is sharply increased and then decreased. It demonstrates that the original integral projection cannot preserve the discriminative information for micro-expressions. As a result, it is necessary to improve integral projection method to obtain the class information for micro-expressions.

We suppose that a frame $I_n(x, y)$ is neutral in micro-expression video clip. The new facial images are derived by subtracting neutral image from the expressive ones. This method is a simple but efficient to remove face identity. Especially, we find that the difference image method has been used in sparse representation for facial expression recognition [28]. Their work demonstrates that this simple method can reduce the influence of face identity. In addition, we provide the statistical analysis of horizontal integral projections on 247 difference-image based facial images of 5-class micro-expression on CASME2 [27] in Fig. 2. It is found from Fig. 2(a) that the horizontal integral projections from 247 images can capture the various structure of signals for different micro-expressions. In addition, they can obtain the specific structure from such regions of interest of micro-expression as mouth region for happiness expression. Moreover, as seen from Fig. 2(b), comparing with Fig. 1(b), the difference image method can well characterize the specific regions of facial movements for different micro-expressions, for example, disgust expression is mostly appearing in eyebrows and eyes. Another finding in Fig. 2(c) argues that the integral projections based on difference images can serves the discriminative structure of 1D signals for different micro-expressions. From these findings, the integral projection based on difference

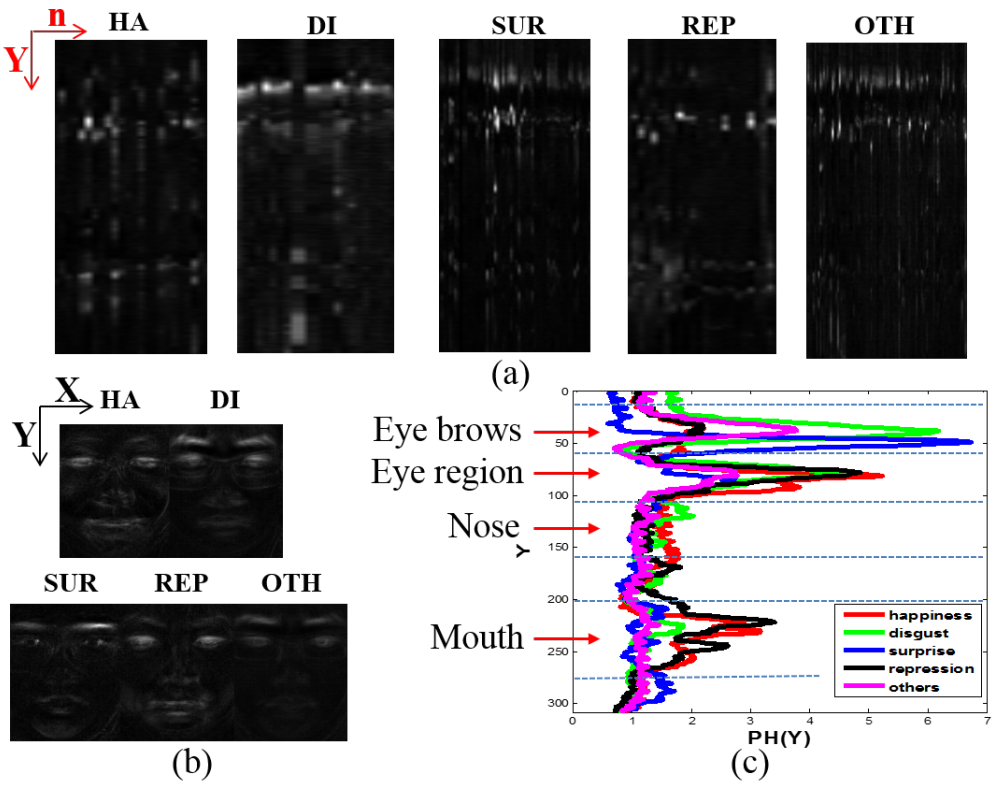


Figure 2: Horizontal integral projection based on difference images. (a) The horizontal integral projections of 247 difference images, where n-axis describes the number of facial images and each projection is represented as a single column. (b) The mean face of difference images for happiness (HA), disgust (DI), surprise (SUR), repression (REP) and others (OTH). (c) The horizontal integral projections of difference images in (b).

images can provide more discriminative information for micro-expressions. Eqs. 2 and 3 can be re-defined as:

$$V_t(x) = \frac{1}{y_2 - y_1} \int_{y_1}^{y_2} (I_t(x,y) - I_n(x,y))dy, \quad (4)$$

$$H_t(y) = \frac{1}{x_2 - x_1} \int_{x_1}^{x_2} (I_t(x,y) - I_n(x,y))dx. \quad (5)$$

2.2. Spatiotemporal Local Binary Pattern based on Integral Projection

Although the integral projection based on difference images preserves the shape of different micro-expressions and has discriminative ability, it is still not robust to describe the appearance and motion of facial images. Recall LBP-TOP [29], which considers micro-expression video clips from three orthogonal planes, representing appearance and motion information, respectively. Next, we borrow the nature of LBP-TOP to get the appearance and motion features from the integral projections.

Firstly, we look at the method to extract the appearance for micro-expressions. Let S_t be the integral projection at time t, where it can be V_t or H_t. The appearance information of an image can be extracted by using one-dimensional local binary pattern (**1DLBP**) [8]. The linear mask of size W, which can be designed as 3, 5, 7 or 9, is established to scan S_t with one element step. The 1DLBP code is calculated by thresholding the neighborhood values against the central element. The neighbors will be assigned the value 1 if they are greater than or equal to the current element and 0 otherwise. Then each binary element of the resulting vector is multiplied by a weight depending on its position. This can be summarized as

$$\text{1DLBP}_{t,W} = \sum_p \delta(S_t(z_p) - S_t(z_c))2^p, \quad (6)$$

where δ is a Dirac's delta, $S_t(z_c)$ is the value at the center of the mask, $z_c \in [y_1, y_2]$ or $[x_1, x_2]$, and z_p is the neighbors of z_c. The distribution of the 1D patterns of each frame is modeled by a histogram which characterizes the frequency of each pattern in the 1D projected signal. This encodes the local and global texture information since the projected signal handles both cues. Fig. 3 shows the procedure to encode the integral projection by using LBP. To describe the appearance of each video clip, the histograms of frames are accumulated. Finally both histogram of horizontal and vertical projections are concatenated into one feature vector,

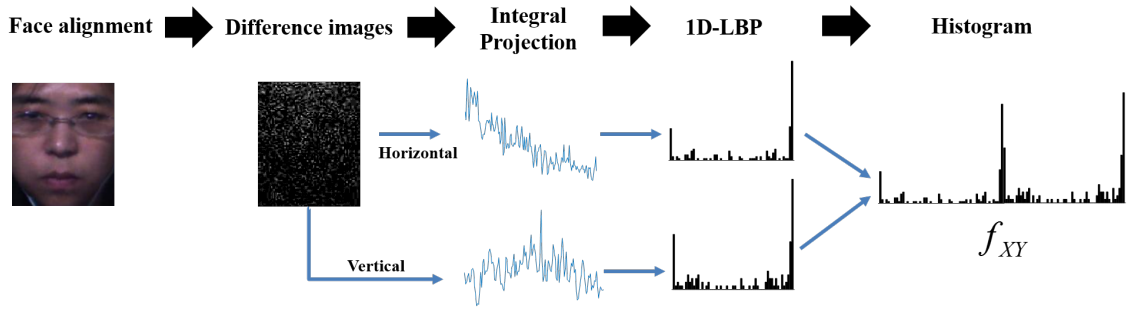


Figure 3: Procedure of encoding difference-image based integral projection on spatial domain.

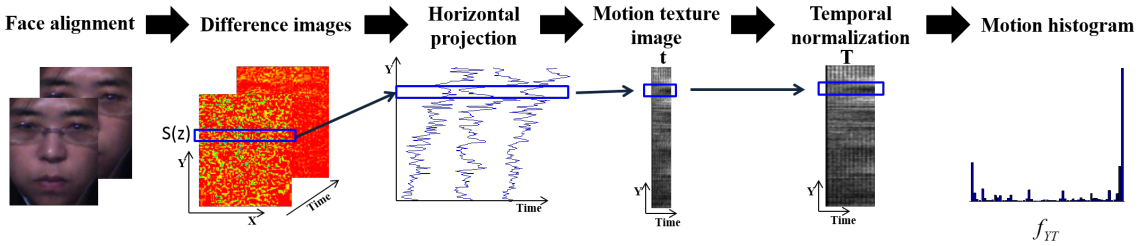


Figure 4: Motion histogram along horizontal direction: The blue rectangle represents the variation of horizontal projection at one position of Y-axis direction, where t and T are the original and temporal normalization time lengths, respectively.

denoted as f_{XY}.

Motion features are extracted from horizontal and vertical direction. Firstly, we consider a simple way to extract the motion histogram along horizontal direction, as shown in Fig. 4. We formulate the horizontal integral projections from all difference images in a video clip as a new texture image. Here, the texture images in horizontal directions can be similar to the YT planes of LBP-TOP, which represents the motion along horizontal direction. As seen from Fig. 4, the change of value $S(z)$ ($z \in [y_1, y_2]$) along the time t definitely shows the motion change of shape of micro-expressions along the horizontal direction. However, the changing rate of micro-expression video clips might be different; it might cause unfair comparison among the motion histograms. Bilinear interpolation is utilized to ensure $S(z)$ along the time t with the same size T. Here we name this procedure *temporal normalization*. Based on the new texture image, a gray-scale invariant texture descriptor, LBP operator [16, 15], which is defined as

$$\text{LBP}_{S,R} = \sum_{s=0}^{S-1} \delta(g_s - g_c)2^s, \tag{7}$$

is exploited to extract the motion histogram, where g_c is the gray value of the center pixel, g_s is the gray value of S equally spaced pixels on a circle of radius R at this center pixel. The same procedure is applied to vertical integral projection.

Empirical experiments tell us that the procedure that normalizes all images into the same size could produce the promising performance. It also allows us to use the same value of R for motion texture images. So far, the motion histograms, which represent motion change along the horizontal (YT) and vertical (XT) directions, are obtained by the process described above. Here, we denote them as f_{YT} and f_{XT}. The final feature vector of a micro-expression video clip can be formulated by $[f_{XY}, f_{XT}, f_{YT}]$, where this feature preserves shape information and discriminative capability.

3. Experiments

In this section, we perform the experiments on two publicly available facial micro-expression databases [27, 11] for evaluating the performance STLBP-IP method. Additionally, we provide the comparison with state-of-the-art approaches.

3.1. Database Description

For evaluating STLBP-IP, we conduct the experiments on CASME2 [27] and SMIC [11] databases for micro-expression recognition.

The CASME2 database includes 247 spontaneous facial micro-expressions recorded by a 200 fps camera and spatial resolution with 640×480 pixel size. In this database, they elicited participants' facial expressions in a well-controlled laboratory environment and proper illumination. These samples are coded with the onset and offset frames, as well as tagged with AUs and emotion. There are 5 classes of the micro-expressions in this database: happiness (32 samples), surprise (25 samples), disgust (64 samples), repression (27 samples) and others (99 samples).

The SMIC database consists of 16 subjects with 164 spontaneous micro-expressions, which were recorded in a

controlled scenario using 100 fps camera with resolution of 640×480. 164 spontaneous facial micro-expressions are categorized into positive (51 samples), negative (70 samples) and surprise (43 samples) classes.

For two databases, we firstly use active shape model (ASM) to obtain the 68 facial landmarks for facial images of video sequence, and align them to a canonical frame. Finally, for CASME2 database, the face images are cropped to 308×257 pixel size, and are divided into 8×9 blocks, while for SMIC database, we crop facial images into 170×139 and divide them into 4×7 blocks. For two database, we employ leave-one-subject-out cross validation protocol in the experiments, where the samples from one subject are used for testing, the rest for training. For the classification, we use SVM with Chi-Square Kernel [3], of which the optimal value of the penalty parameter is determined using the three-fold cross validation.

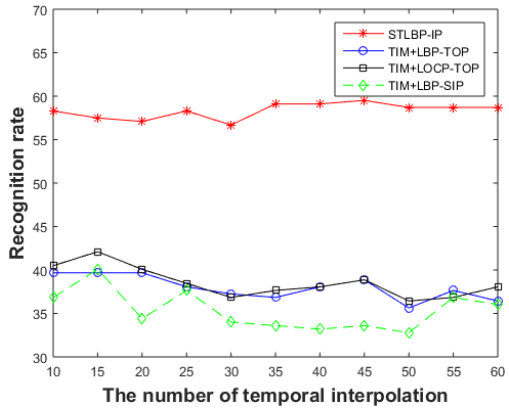
3.2. Parameter Evaluation

Previously mentioned in Section 2.1, we introduce the difference method for integral projection. The mask size W of 1DLBP and the temporal normalization T are two important parameters which determine the computational complexity and classification performance. In this scenario, we aim to evaluate the influence of the difference-image method and two parameters.

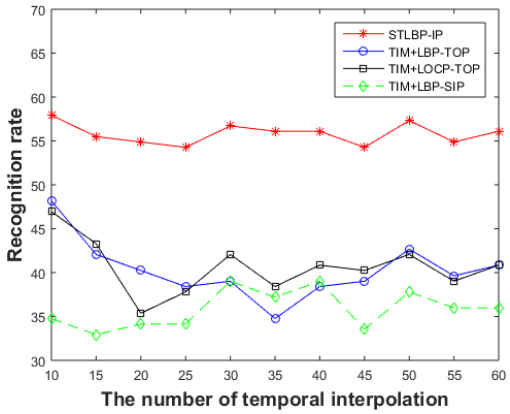
For evaluating the difference-image method, we employ ordinal integral projection instead of difference-image based integral projection in STLBP-IP. We obtained the accuracies of 35.22% and 39.63% for CASME2 and SMIC2 databases, respectively. However, for STLBP-IP, we obtained the accuracies of 59.51% and 54.88% for CASME2 and SMIC2 databases, respectively. Comparing with original integral projection, it is found that the recognition rate is increased by 24.29% and 15.25% for CASME2 and SMIC2 databases, respectively. The performance is substantially improved by using difference-image based integral projection. This may be explained by that difference-image based integral projection can reduce the influence of person identity information.

Furthermore, we evaluate the performance influenced by different W. In this case, we remain T as the number of frame of each video clip. For CASME2, we obtained the accuracies of STLBP-IP as 51.01%, 57.09%, 56.68%, and 58.3% for $W = 3, 5, 7, 9$, respectively. For SMIC2, the accuracies of STLBP-IP are 46.95%, 53.05%, 54.27% and 54.88% for $W = 3, 5, 7, 9$ respectively. From the above result, we can find that the large mask size can help the 1DLBP capture more information in the local neighborhood. From these results, we set W to 9 for next experiments on CASME2 and SMIC2 databases.

Finally, we evaluate the influence of T to STLBP-IP. In this experiment, we test STLBP-IP under various



(a)



(b)

Figure 5: Evaluation of temporal normalization on micro-expression recognition on CASME2 and SMIC2 database.

T on $[5, 60]$. For comparison, we employ the strategy of [11], in which they proposed to use temporal interpolation method (**TIM**) to normalize one video into the same frames length, and then utilize spatiotemporal features on normalized video. Moreover, we use three commonly used spatiotemporal features (LBP-TOP [29], LBP-SIP [25], Local Ordinary Contrast Pattern from Three Orthogonal Planes (**LOCP-TOP**) [2])) for normalized videos.

Fig. 5 shows the influence of T to STLBP-IP and three others features on TIM. As seen from Fig. 5, for STLBP-IP, the larger temporal normalization size is, the better description of the motion change we obtain. However, if T is too large, the motion texture image would provide abundant information for classification, thus the performance will be decreased when T is more than 45 and 10 for CASME2 and SMIC2 databases, respectively. But for LBP-TOP, LOCP-TOP and LBP-SIP, we can see that their performance is much affected by T of TIM, while for STLBP-IP, its performance is just slightly affected by T. From this figure, we can see that STLBP-IP achieves 59.51% and 57.93% for CASME2 and SMIC2 databases, respectively, when T is 45 and 10, respectively.

Table 1: Micro-expression recognition accuracies in CASME2. All methods are based on our implementation, where the values in bracket for LBP-TOP mean the radius for XY, XT and YT planes, for SVM represents kernel function.

Methods	Accuracy (%)
TIM15+LBP-TOP(3,3,3)+SVM(ChiSquare)	39.68
LBP-TOP(3,3,3)+SVM(ChiSquare) [29]	36.03
LBP-TOP(3,3,3)+SVM(linear) [27]	38.87
TIM15+LBP-SIP(3,3,3)+SVM(ChiSquare) [25]	40.08
TIM15+LOCP-TOP(3,3,3)+SVM(ChiSquare) [2]	42.11
STLBP-IP	**59.51**

3.3. Algorithm Comparison

3.3.1 CASME2 Database

We compare the recognition rate of STLBP-IP with the baseline algorithm [27] of CASME2, LBP-TOP [29], LBP-SIP [25], LOCP-TOP [2]. We employ the following setup for each method:

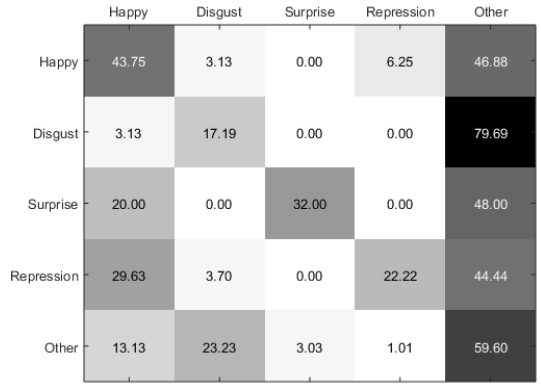
(1) In [27], they used LBP-TOP for 5×5 facial blocks, and radius 3 for LBP operator for three orthogonal planes. In our experiments, following their experimental setup, we re-implement LBP-TOP on 5×5 facial blocks. For classification, we employ linear-kernel based SVM [3]. For convenience, we name this method as LBP-TOP(3,3,3)+SVM(linear).

(2) In [11], they proposed to use temporal interpolation method (TIM) to interpolate each video into the same frames, and then use LBP-TOP for the interpolated video. We re-implement their method as a comparison. In our implementation, we use temporal interpolation method [30] to interpolate all videos into 15 frames, which is denoted as $TIM15$. For spatiotemporal features, we employ LBP-TOP [29], LBP-SIP [25] and LOCP-TOP [2], respectively. The radius and number of neighbors are 3 and 8, respectively. For convenience, we name these methods as TIM15+LBP-TOP, TIM15+LBP-SIP and TIM15+LOCP-TOP, respectively.

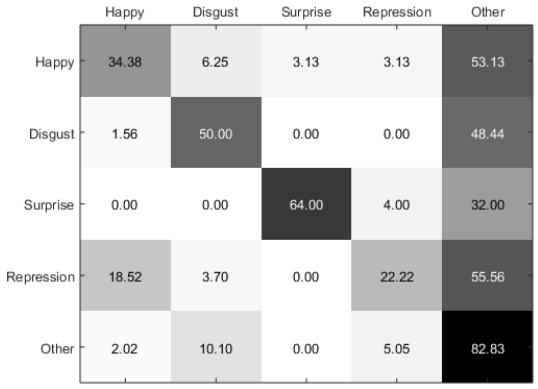
(3) Moreover, we straightly use LBP-TOP [29] on the original videos. For each facial image, we divide it into 8×9 blocks. Here, we denote this method as LBP-TOP.

Results on CASME2 are presented in Table 1. As can be seen, STLBP-IP is shown to outperform the re-implementation of [27]. The accuracy of STLBP-IP is raised by 20.64%. From this table, it is found that temporal interpolation method can boost the performance of LBP-TOP, where LBP-TOP is increased to 39.68%. However, comparing to STLBP-IP, LBP-TOP works worse on micro-expression recognition, and there is a large gap (19.43%) on the performance.

Comparing with LOCP-TOP and LBP-SIP, STLBP-IP increases the performance by 17.4% and 19.43% for micro-expression recognition, respectively. These results demonstrate that STLBP-IP achieves the promising performance rather than LOCP-TOP and LBP-SIP. This is partly explained by STLBP-IP preserves the shape and discriminative ability by using integral projection.



(a)



(b)

Figure 6: The confusion matrix of (a) TIM20+LBP-TOP and (b) STLBP-IP for five micro-expression categorizations on CASME2 database.

The confusion matrix of five micro-expressions is shown in Fig. 6, where we list the results of TIM15+LBP-TOP and STLBP-IP. From this comparison, STLBP-IP performs better on recognizing Disgust, Surprise and Others classes, while works worse on Happy class. We can also find that for LBP-TOP and STLBP-IP, most of micro-expressions, such as happy and repression, are falsely classified into Other class. It may be explained that Other class includes some confused micro-expressions.

3.3.2 SMIC Database

For SMIC database, we compare STLBP-IP with the commonly used spatiotemporal features [29, 2, 25, 5]. In our implementation, we used temporal interpolation method (TIM) to normalize each video into 10 frames.

Table 2: Micro-expression recognition accuracies in SMIC2. All methods are based on our implementation.

Methods	pos/neg/sur (%)
TIM10+LBP-TOP [29]	48.17
TIM10+LOCP-TOP [2]	46.95
TIM10+LBP-SIP [25]	44.51
Periodic+KNN [5]	37.08
LBP+CRF [10]	33.54
Shape+CRF [10]	32.93
Dense optical flow+HMM [20]	20.12
LBP-TOP+SVM(Linear) [11]	51.83
LBP-TOP+SVM(Chi-Square) [11]	46.95
STLBP-IP	**57.93**

(1) We use spatiotemporal feature descriptor (LBP-TOP, LOCP-TOP and LBP-SIP), to 4×7 facial blocks. For convenience, we name them as TIM10+LBP-TOP, TIM10+LOCP-TOP and TIM10+LBP-SIP, respectively.

(2) As well, we re-implement the method of [11], where LBP-TOP is used to 8×8 facial blocks. For classification, we use two SVMs based on linear kernel and Chi-Square kernel functions, respectively. They are denoted as LBP-TOP+SVM(linear) and LBP-TOP+SVM(Chi-Square).

(3) We employ the work of [5] for comparison, where we use periodic feature and k-nearest-neighbor (KNN) classification. The same parameter setup to [5] is exploited.

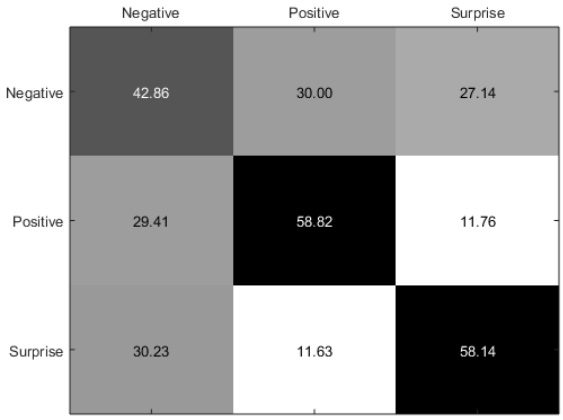
(4) We also employ LBP features with conditional random field(**CRF**) [10], geometric features with CRF [10], dense optical flow with hidden markov model (**HMM**) [20], for comparison.

The comparison results are reported in Table 2. From Table 2, it is noticed that the temporal models with appearance and shape features work poorly in micro-expression recognition. Among the temporal model, LBP+CRF got the best one of 33.54% for micro-expression recognition. STLBP-IP performs much better than the temporal model methods. Comparing with TIM10+LBP-TOP, STLBP-IP increases the accuracies of 9.15% for micro-expression recognition. Comparing with LOCP-TOP, the micro-expression recognition performance is increased by 10.37%. Re-implementation of [11] obtained the best recognition rate of 51.83% among all comparisons, however, it is worse than STLBP-IP. These results demonstrate that STLBP-IP achieves better than geometric features and three spatiotemporal features.

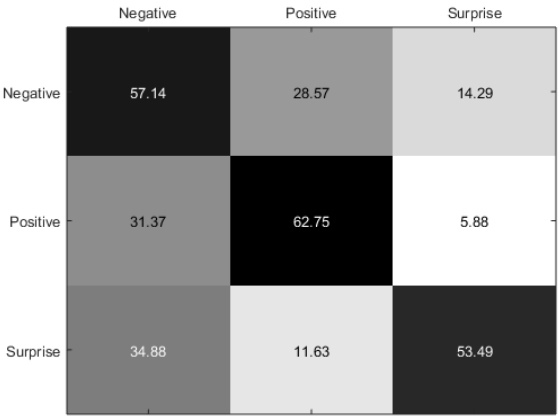
Finally, we present the confusion matrix for LBP-TOP+SVM(linear) and STLBP-IP in Fig. 7. From Fig. 7, it is found that STLBP-IP improves the accuracies of each class. It can better recognize Positive micro-expression.

4. Conclusion

In this paper, we have shown that the spatiotemporal local binary pattern based on integral projection (STLBP-IP) achieves the state-of-the-art performance on two facial micro-expression databases. The use of integral projection based on difference images allows us to preserve the shape property of micro-expressions and then enhance discrimination of micro-expressions. Furthermore, we have presented to use local binary pattern operators to further describe the appearance and motion changes from horizontal and vertical integral projections, well suited for extracting the subtle micro-expressions. As a result, we obtain an effective and efficient system for 'understanding' micro-expressions, which may be applied to widely potential applications.



(a)



(b)

Figure 7: The confusion matrix of (a) LBP-TOP+SVM(linear) [11] and (b) STLBP-IP for three micro-expression categorizations on SMIC2 database.

5. Acknowledgment

This work was supported by the Academy of Finland and Infotech Oulu. This work was supported in part by the National Natural Science Foundation of China (61379095), the Beijing Natural Science Foundation (4152055), the Open Projects Program of National Laboratory of Pattern Recognition (201306295)

References

[1] A. Benzaoui and A. Boukrouche. Face recognition using 1dlbp texture analysis. In *Proc. FCTA*, pages 14–19, 2013.

[2] C. Chan, B. Goswami, J. Kittler, and W. Christmas. Local ordinal contrast pattern histograms for spatiotemporal, lip-based speaker authentication. *IEEE Transactions on Information Forensics and Security*, 7:602–612, 2012.

[3] C. Chang and C. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

[4] A. Davison, M. Yap, N. Costen, K. Tan, C. Lansley, and D. Leightley. Micro-facial movements: an investigation on spatio-temporal descriptors. In *ECCV workshop on Spontaneous Behavior Analysis*, 2014.

[5] P. Dollar, V. Rabaud, G. Cottrell, and S. Belongie. Behavior recognition via sparse spatio-temporal features. In *Proc. VSPETS*, pages 65–72, 2005.

[6] P. Ekman. Lie catching and mico expressions. In E. C. Martin, editor, *The Philosophy of Deception*, pages 118–133. Oxford University Press, 2009.

[7] M. Frank, M. Herbasz, K. Sinuk, A. Keller, and C. Nolan. I see how you feel: Training laypeople and professionals to recognize fleeting emotions. In *International Communication Association*, 2009.

[8] L. Houam, A. Hafiane, A. Boukrouche, E. Lespessailles, and R. Jennane. One dimensional local binary pattern for bone texture characterization. *Pattern Analysis and Applications*, 17:179–193, 2014.

[9] X. Huang, G. Zhao, W. Zheng, and M. Pietikäinen. Towards a dynamic expression recognition system under facial occlusion. *Pattern Recognition Letters*, 33(16):2181–2191, 2012.

[10] S. Jain, C. Hu, and J. Aggarwal. Facial expression recognition with temporal modeling of shapes. In *Proc. ICCV*, pages 1642–1649, 2011.

[11] X. Li, T. Pfister, X. Huang, G. Zhao, and M. Pietikäinen. A spontaneous micro-expression database: Inducement, collection and baseline. In *Proc. AFGR*, 2013.

[12] G. Mateos. Refining face tracking with integral projection. In *Proc. AVBPA*, pages 360–368, 2003.

[13] G. Mateos, A. Ruiz-Garcia, and P. Lopez-de Teruel. Human face processing with 1.5d model. In *Proc. AMFG*, pages 220–234, 2007.

[14] D. Matsumoto, H. Hwang, L. Skinner, and M. Frank. Evaluating truthfulness and detecting deception. *FBI Law Enforcement Bulletin*, 2011. Article available at `http://www.fbi.gov/stats-services/publications/law-enforcement-bulletin/june_2011/school_violence`.

[15] T. Ojala, A. Hadid, and M. Pietikäinen. Face description with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12):2037–2041, 2006.

[16] T. Ojala, M. Pietikäinen, and T. Mäenpää. Multiresolution gray-scale and rotation invariant texutre classification with local binary pattern. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002.

[17] T. Pfister, X. Li, G. Zhao, and M. Pietikäinen. Recognising spontaneous facial micro-expressions. In *Proc. ICCV*, pages 1449–1456, 2011.

[18] D. Robinson and P. Milanfar. Fast local and global projection-based methods for affine motion estimation. *Journal of Mathematical Imaging and Vision*, 18:35–54, 2003.

[19] J. Ruiz-Hernandez and M. Pietikäinen. Encoding local binary patterns using re-parameterization of the second order gaussian jet. In *Proc. AFGR*, 2013.

[20] G. Shin and J. Chun. Spatio-temporal facial expression recognition using optical flow and hmm. *Software Engineering, Artificial Intelligence, Network, and Parallel/Distributed Computing*, pages 27–38, 2008.

[21] M. Shreve, S. Godavarthy, D. Goldgof, and S. Sarkar. Macro- and micro-exprssion spotting in long videos using spatio-temporal strain. In *Proc. AFGR*, pages 51–56, 2011.

[22] M. Shreve, S. Godavarthy, V. Manohar, D. Goldgof, and S. Sarkar. Towards macro-and micro-expression spotting in video using strain patterns. In *Proc. WACV*, pages 1–6, 2009.

[23] S. Wang, W. Yan, X. Li, G. Zhao, and X. Fu. Micro-expression recognition using dynamic textures on tensor independent color space. In *Proc. ICPR*, 2014.

[24] S. Wang, W. Yan, G. Zhao, and X. Fu. Micro-expression recognition using robust principal component analysis and local spatiotemporal directional features. In *ECCV workshop on Spontaneous Behavior Analysis*, 2014.

[25] Y. Wang, J. See, R. Phan, and Y. Oh. LBP with six intersection points: Reducing redundant information in LBP-TOP for micro-expression recognition. In *Asian Conference on Computer Vision*, 2014.

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

[27] W. Yan, X. Li, S. Wang, G. Zhao, Y. Liu, Y. Chen, and X. Fu. CASME II: An improved spontaneous micro-expression database and the baseline evaluation. *PLOS ONE*, 9(1):1–8, 2014.

[28] S. Zafeiriou and M. Petrou. Sparse representations for facial expressions recognition via L1 optimization. In *Proc. CVPRW*, pages 32–39, 2010.

[29] G. Zhao and M. Pietikäinen. Dynamic texture recognition using local binary pattern with an application to facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6):915–928, 2009.

[30] Z. Zhou, G. Zhao, Y. Guo, and M. Pietikäinen. An image-based visual speech animation system. *IEEE Transactions on Circuits and Systems for Video Technology*, 22(10):1420–1432, 2012.