Micro-expression Recognition Using Color Spaces

Su-Jing Wang, *Member, IEEE,* Wen-Jing Yan, Xiaobai Li, Guoying Zhao *Senior Member, IEEE,* Chun-Guang Zhou, Xiaolan Fu *Member, IEEE,* Minghao Yang, Jianhua Tao *Member, IEEE*

***Abstract*—Micro-expressions are brief involuntary facial expressions that reveal genuine emotions and, thus help detect lies. Because of their many promising applications, they have attracted the attention of researchers from various fields. Recent research reveals that two perceptual color spaces (CIELab and CIELuv) provide useful information for expression recognition. This paper is an extended version of our International Conference on Pattern Recognition (ICPR) paper [1], in which we propose a novel color space model, Tensor Independent Color Space (TICS), to help recognize micro-expressions. In this paper, we further show that CIELab and CIELuv are also helpful in recognizing micro-expressions, and we indicate why these three color spaces achieve better performance. A micro-expression color video clip is treated as a fourth-order tensor, i.e., a four-dimension array. The first two dimensions are the spatial information, the third is the temporal information, and the fourth is the color information. We transform the fourth dimension from RGB into TICS, in which the color components are as independent as possible. The combination of dynamic texture and independent color components achieves a higher accuracy than does that of RGB. In addition, we define a set of Regions of Interest (ROIs) based on the Facial Action Coding System (FACS) and calculated the dynamic texture histograms for each ROI. Experiments are conducted on two micro-expression databases, CASME and CASME 2, and the results show that the performances for TICS, CIELab and CIELuv are better than those for RGB or gray.**

***Index Terms*—Micro-expression recognition, Color Spaces, Tensor Analysis, Local Binary Patterns, Facial Action Coding System.**

This work was supported by grants from the National Natural Science Foundation of China (61379095, 61375009, 61472138, 31500875, 61332017), the Beijing Natural Science Foundation (4152055), the Open Projects Program of National Laboratory of Pattern Recognition (201306295), the open project program of Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, and the Academy of Finland, and Infotech Oulu.

S.J Wang is with the Key Laboratory of Behavior Sciences, Institute of Psychology, Chinese Academy of Sciences, Beijing, 100101, China, and also with the Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun, 130012, P.R. China (e-mail:wangsujing@psych.ac.cn).

W.J Yan is with the College of Teacher Education, Wenzhou University, Wenzhou, 325035, China. (e-mail: yanwj@wzu.edu.cn).

X.B Li and G.Y Zhao are with the Department of Computer Science and Engineering, University of Oulu, P. O. Box 4500, FI-90014, Finland. (e-mail: gyzhao@ee.oulu.fi).

C.G Zhou is with the College of Computer Science and Technology, Jilin University, Changchun 130012, China, and also with Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun, 130012, P.R. China (e-mail:cgzhou@jlu.edu.cn).

X Fu is with the State Key Laboratory of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, 100101, China. (e-mail:fuxl@psych.ac.cn).

M.H Yang and J.H Tao are with the National Laboratory of Pattern Recognition (NLPR) Institute of Automation, Chinese Academy of Sciences (CASIA) No.95 ZhongGuanCun East Street, HaiDian District, Beijing, 100190, China. (e-mail:{mhyan, jhtao}@nlpr.ia.ac.cn).

I. Introduction

Micro-expressions are brief facial expressions that reveal the emotions that a person tries to conceal, especially in high-stakes situations [2][3]. Compared with normal facial expressions, micro-expressions have three significant characteristics. They are of short duration, and low intensity and are generally fragments of prototypical facial expressions. Micro-expressions are generally known for their potential use in many fields, such as clinical diagnosis [4], national security [5] and interrogations [6], because they may reveal genuine emotions and thus help detect lies. The polygraph is invasive because it must be connected to the individual's body throughout the session [7]. Thus, individuals are aware that they are being monitored and may develop countermeasures. In comparison, lie detection based on micro-expressions is unobtrusive, and the individuals being observed are less likely to develop countermeasures.

More than 30 years ago, psychologists began to show interest in micro-expressions. Haggard and Isaacs first discovered micro-expressions (micro-momentary expressions) and viewed them as cues for repressed emotions [8][9]. In 1969, Ekman analyzed a video of an interview with a patient, who was suffering from depression and had tried to commit suicide, and observed micro-expressions. Since that time, Ekman's group has conducted many studies on micro-expressionds [10]. According to Ekman, micro-expressions are the most promising approach for detecting deception [3].

Although micro-expressions have potential application in a variety of fields, humans have difficulty in detecting and recognizing them. This difficulty may be the result of their short duration, and low intensity in addition to fragmental action units [2][11]. Although there is debate regarding their duration, the generally accepted limit is 0.5 seconds [11][12]. Micro-expressions are usually very subtle because individuals try to control and repress them [11]. In addition, micro-expressions usually exhibit only parts of the action units of fully-stretched facial expressions. For example, only the upper face or lower face may show action units, not both at the same time as in macro-expressions [13]. To improve human performance in recognizing micro-expressions, Ekman [14] developed the Micro-Expression Training Tool (METT), which trains people to better recognize seven categories[1] of micro-expressions. Due to the increasing power of computers and the overwhelming quantity of expressions to monitor, researchers have turned to automatic micro-expression recognition.

However, there are few papers addressing micro-expression recognition. Polikovsky *et al.* [15] used a 3D-gradient descriptor for micro-expressions recognition. Wang *et al.* [16]

[1]Contempt was added besides the basic six emotions

treated a micro-expression gray-scale video clip as a 3rd-order tensor and used Discriminant Tensor Subspace Analysis (DTSA) and Extreme Learning Machine (ELM) to recognize micro-expression. However, the subtle movements of micro-expressions may be lost in the process of solving DTSA. Pfister *et al.* [17] used a Temporal Interpolation Model (TIM) based on Laplacian matrix to normalize the frame numbers of micro-expression video clips. Then, the LBP-TOP [18] was used to extract the motion and appearance features of micro-expressions and multiple kernel learning was used for classification.

The LBP operator has been widely used in ordinary texture analysis. It efficiently describes the local structures of images, and in recent years, has aroused increasing interests in many areas of image processing and computer vision, exhibiting its effectiveness in a number of applications. Recently, research on LBP has flourished. Tan and Triggs [19] developed a generalization of the local texture descriptor named as Local Ternary Pattern (LTP) for face recognition, which is more discriminative and less sensitive to noise in uniform regions. Zhu *et al.* [20] divided P neighbors into $[P/4]$ groups and calculated an LBP histogram for each group including at most 4 points. The lines connecting two neighboring points to the central point are orthogonal. They named the method as the orthogonal combination of local binary patterns (OC-LBP). The objective of OC-LBP is to reduce the dimensionality of the original LBP operator while keeping its discriminative power and computational efficiency. The authors also proposed six new local descriptors based on OC-LBP enhanced with color information for image region description. The main idea is to independently calculate the original OC-LBP descriptor over different channels in a given color space, and then concatenate them to obtain the final color OC-LBP descriptor [20]. Lee et al. [21] presents a novel expression recognition method that exploits the effectiveness of color information and sparse representation.

Color is a fundamental aspect of human perception, and its effects on cognition and behavior have intrigued generations of researchers [22]. Recent research efforts [23][24][25][26][27][28] revealed that color may provide useful information for face recognition. In [23], it is also revealed that the face recognition system for various color spaces (such as RGB, PCS and YIQ) is better than for gray images. Liu [24] applied PCA, ICA and LDA to obtain the uncorrelated color space (UCS), the independent color space (ICS), and the discriminating color space (DCS) for face recognition, respectively. In [25], the authors took advantage of the ICS to improve performance of the Face Recognition Grand Challenge (FRGC) [29]. Yang and Liu proposed [26] the Color Image Discriminant (CID) model borrowing the idea of LDA to not only obtain the discriminative color space but also to obtain the optimal spatial transformation matrix. Wang *et al.* [27] presented a Tensor Discriminant Color Space (TDCS) model that uses a 3rd-order tensor to represent a color facial image. To make the model more robust to noise, they [28] also used an elastic net to propose a Sparse Tensor Discriminant Color Space (STDCS). Lajevardi and Wu [30] also treated a color facial expression image as a 3rd-order tensor and showed that the perceptual color spaces (CIELab and CIELuv) are better overall than other color spaces for facial expression recognition.

When the emotional and physiological states of humans change, the facial skin color hue subtly changes, due to variations in the levels of hemoglobin and oxygenation under the skin. Ramirez *et al.* [31] showed that facial skin color changes can be used to infer the emotional state of a person in the valence dimension with an accuracy of 77.08%. We infer that when different micro-expressions occur, the facial skin color hues are also different. Given such a consideration, facial color information could help improve micro-expression recognition.

This paper is an extended version of our International Conference on Pattern Recognition (ICPR) paper [1], in which we propose a novel color space model, Tensor Independent Color Space (TICS), to help recognize micro-expressions. In this paper, we further show that CIELab and CIELuv are also helpful in recognizing micro-expressions. In these color spaces, LBP-TOP is used to extract the dynamic texture features of micro-expression clips from three color components. Then, the histograms of the LBP-TOP codes are concatenated as a long feature vector, which is treated as the input of SVM to recognize micro-expressions. The results in TICS are slightly better than those in CIELab and CIELuv. A key difference with CIELab and CIELuv is that TICS is modeled by learning from samples. The three color components in TICS are as independent from each other as possible. We use the mutual information to explain why TICS, CIELab and CIELuv achieve better performance than the RGB color space. In addition, we use tensors to generalize LBP-TOP to a higher-dimensional space, and propose Tensor Orthogonal LBP (TO-LBP). We also show LBP-TOP is a special case of TO-LBP in 3D space.

The rest of this paper is organized as follows: in Section II, we briefly review the fundamentals of tensors and the perceptual color spaces; in Section III, we present the Tensor Independent Color Space model (TICS); in Section IV, we introduce LBP-TOP, which is used to extract the dynamic texture features of micro-expression clips from three components in TICS; in Section V, we design 16 Region of Interests (ROIs) based on Action Units; in Section VI, we describe the micro-expression recognition framework based on TICS and LBP-TOP; in Section VII, the experiments are conducted on two micro-expression databases CASME and CASME 2, the results showing the efficiency and performance of TICS; finally in Section VIII, conclusions are drawn and several issues for future work are discussed.

II. Background

A. Tensor Fundamentals

A tensor is a multidimensional array. It is the higher-order generalization of a scalar (zero-order tensor), vector (1st-order tensor), and matrix (2nd-order tensor). In this paper, lowercase italic letters (*a, b, ...*) denote scalars, bold lowercase letters (**a, b, ...**) denote vectors, bold uppercase letters (**A, B, ...**) denote matrices, and calligraphic uppercase letters ($\mathcal{A}$, $\mathcal{B}$, ...) denote tensors. The formal definition is given below[32]:

Definition 1. *The* order *of a tensor* $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ *is* N. *An element of* $\mathcal{A}$ *is denoted by* $\mathcal{A}_{i_1 i_2 \ldots i_N}$ *or* $a_{i_1 i_2 \ldots i_N}$, *where* $1 \leq i_n \leq I_n$, $n = 1, 2, \ldots, N$.

Definition 2. *The* n-mode vectors *of* $\mathcal{A}$ *are the* I_n*-dimensional vectors obtained from* $\mathcal{A}$ *by fixing every index but index* i_n.

Definition 3. *The* n-mode unfolding matrix *of* $\mathcal{A}$, *denoted by* $(\mathcal{A})_{(n)} \in \mathbb{R}^{I_n \times (I_1 \ldots I_{n-1} I_{n+1} \ldots \times I_N)}$, *contains the element* $a_{i_1 \ldots i_N}$ *at the* i_n*th row and* j*th column, where*

$$j = 1 + \sum_{k=1, k \neq n}^{N} (i_k - 1) J_k, \qquad \textit{with} \qquad J_k = \prod_{m=1, m \neq n}^{k-1} I_m. \tag{1}$$

We can generalize the product of two matrices to the product of a tensor and a matrix.

Definition 4. *The* n-mode product *of a tensor* $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ *by a matrix* $\mathbf{U} \in \mathbb{R}^{J_n \times I_n}$, *denoted by* $\mathcal{A} \times_n \mathbf{U}$, *is an* $(I_1 \times I_2 \times \ldots \times I_{n-1} \times J_n \times I_{n+1} \times \ldots \times I_N)$*-tensor for which the entries are given by:*

$$(\mathcal{A} \times_n \mathbf{U})_{i_1 i_2 \ldots i_{n-1} j_n i_{n+1} \ldots i_N} \stackrel{def}{=} \sum_{i_n} a_{i_1 i_2 \ldots i_{n-1} i_n i_{n+1} \ldots i_N} u_{j_n i_n}. \tag{2}$$

Definition 5. *The* scalar product *of two tensors* $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$, *denoted by* $\langle \mathcal{A}, \mathcal{B} \rangle$, *is defined in a straightforward way as* $\langle \mathcal{A}, \mathcal{B} \rangle \stackrel{def}{=} \sum_{i_1} \sum_{i_2} \ldots \sum_{i_N} a_{i_1 i_2 \ldots i_N} b_{i_1 i_2 \ldots i_N}$. *The* Frobenius norm *of a tensor* $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ *is then defined as* $\|\mathcal{A}\|_F \stackrel{def}{=} \sqrt{\langle \mathcal{A}, \mathcal{A} \rangle}$

B. Perceptual Color Spaces

In this section, we introduce two perceptual color spaces: CIELab, and CIELuv, which can enhance the performance of expression recognition [30]. In computer vision, the most widely used color space is RGB color space, which is the basis for other color spaces (such as YCbCr, CIELab, and CIELuv) that are usually defined by its linear or nonlinear transformations. To reduce the lighting effect, the RGB color space is usually normalized, and denoted as $(R_{norm}, G_{norm}, B_{norm})$.

To convert from RGB to perceptual color spaces (CIELab or CIELuv), the RGB color space is first converted to the CIE XYZ color space, which is the basis for conversion to perceptual color spaces. The component L is the same for both the CIELab and CIELuv color spaces. The component L indicates lightness and is independent of the other two components. The conversion procedure is as follows [30]:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.431 & 0.342 & 0.178 \\ 0.222 & 0.707 & 0.071 \\ 0.020 & 0.130 & 0.939 \end{bmatrix} \begin{bmatrix} R_{norm} \\ G_{norm} \\ B_{norm} \end{bmatrix} \tag{3}$$

$$L = \begin{cases} 116 \times (\frac{Y}{Y_n})^{\frac{1}{3}} - 16, & \frac{Y}{Y_n} > 0.008856 \\ 903 \times (\frac{Y}{Y_n}), & \frac{Y}{Y_n} \leq 0.008856 \end{cases} \tag{4}$$

$$a = 500 \times \left(f(\frac{X}{X_n}) - f(\frac{Y}{Y_n}) \right) \tag{5}$$

$$b = 200 \times \left(f(\frac{Y}{Y_n}) - f(\frac{Z}{Z_n}) \right) \tag{6}$$

where X_n, Y_n, and Z_n are reference white tristimulus values, which are defined in the CIE chromaticity diagram [33] and

$$f(t) = \begin{cases} t^{\frac{1}{3}}, & t > 0.008856 \\ 7.787 \times t + \frac{16}{116}, & t \leq 0.008856 \end{cases} \tag{7}$$

For the u and v color components, the conversion is defined by

$$u = 13 \times L \times (u' - u'_n) \qquad \text{and} \qquad v = 13 \times L \times (v' - v'_n). \tag{8}$$

The equations for u' and v' are given below

$$u' = \frac{4X}{X + 15Y + 3Z} \qquad \text{and} \qquad v' = \frac{9Y}{X + 15Y + 3Z} \tag{9}$$

The quantities u'_n and v'_n are the (u', v') chromaticity coordinates of a specified white object and are defined by

$$u'_n = \frac{4X_n}{X_n + 15Y_+ 3Z_n} \qquad \text{and} \qquad v'_n = \frac{9Y_n}{X_n + 15Y_n + 3Z_n} \tag{10}$$

In Section VII, we will show that the mutual information among each component in CIELab (or CIELuv) is small, and this will explain why perceptual color spaces are better than RGB for micro-expression recognition.

III. Tensor Independent Color Space (TICS)

Unlike CIELab and CIELuv, Tensor Independent Color Space (TICS) is not a fixed linear or nonlinear transformation of RGB. Its transformation matrix is obtained by learning from the provided samples. A color micro-expression video clip is naturally represented by a 4th-order tensor, where mode-1 and mode-2 of a tensor are facial spatial information, mode-3 is the temporal information and mode-4 is the color space information. For instance, a color micro-expression video clip with a resolution of $I_1 \times I_2$ is represented as a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3 \times I_4}$, where I_3 is the number of frames and $I_4 = 3$ has 3 components corresponding to $\mathbf{R}$, $\mathbf{G}$ and $\mathbf{B}$ in RGB space. However, the $\mathbf{R}$, $\mathbf{G}$ and $\mathbf{B}$ components are correlated. If we can transform the three correlated components into a series of uncorrelated components $\mathbf{T}_1$, $\mathbf{T}_2$ and $\mathbf{T}_3$, and extract the dynamic texture features from each uncorrelated component, we can obtain better results.

Given the assumption that M is the number of color micro-expression video clips, $\mathcal{X}_i$ is the ith color micro-expression video clip. We want to seek a color space transformation matrix $\mathbf{U}_4 \in \mathbb{R}^{I_4 \times L_4}$ (usually $L_4 = I_4$) for transformation

$$\begin{aligned} &\mathcal{Y}_i = \mathcal{X}_i \times_4 \mathbf{U}_4^T, \\ &i = 1, 2, \ldots, M. \end{aligned} \tag{11}$$

such that the components of mode-4 of $\mathcal{Y}_i$ are as independent as possible. To obtain $\mathbf{U}_4$, we use ICA [2] to decorrelate the RGB color space. M 4th-order tensor $\mathcal{X}_i$ are concatenated to a 5th-order tensor $\mathcal{F} \in \mathbb{R}^{I_1 \times I_2 \times I_3 \times I_4 \times M}$. The mode-4 unfolding matrix $\mathbf{F}_{(4)}$ is a $3 \times K$ matrix, where $K = I_1 \times I_2 \times I_3 \times M$ and the three rows of $\mathbf{F}_{(4)}$ correspond to the three components in RGB space.

[2]For ICA operations, we used Hyvarinen's fixed-point algorithmhttp://www.cis.hut.fi/projects/ica/fastica/.

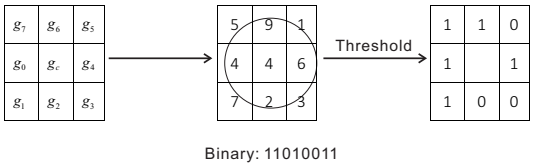


Fig. 1. Example of a basic LBP operator.

The color space transformation matrix $\mathbf{U}_4$ may be derived using ICA on $\mathbf{F}_{(4)}$. The ICA of $\mathbf{F}_{(4)}$ factorizes the covariance matrix Σ_F into the following form:

$$\Sigma_F = \mathbf{U}_4^{-1} \nabla \mathbf{U}_4^{-T} \tag{12}$$

where $\nabla \in \mathbb{R}^{3\times3}$ is diagonal real positive and $\mathbf{U}_4$ transforms RGB color space to a new color space whose three components are independent or the most possible independent. $\mathbf{U}_4$ in Eq. (12) may be derived using Comon's ICA algorithm by calculating mutual information and higher-order statistics [34].

IV. LBP DESCRIPTION FROM THREE ORTHOGONAL PLANES

Local Binary Patterns (LBPs) [35] are used on gray images to extract texture features. Given a pixel c in the gray image, its LBP code is computed by comparing it with its P neighbors p. The neighbors lie on a circle with center c and a radius equal to R.

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p \tag{13}$$

where g_c is the gray value of the given pixel c, g_p is the value of its neighbor p, and $s(u)$ is 1 if $u \geq 0$ and 0 otherwise. If the coordinates of c are (x_c, y_c), the coordinates of p are $(x_c + Rcos(2\pi p/P), y_c - Rsin(2\pi p/P))$. The coordinates of the neighbors that do not fall exactly on pixels are approximated by bilinear interpolation. The LBP encoding process is illustrated in Fig. 1. It is possible to use only a subset of 2^p binary patterns to describe the texture of the images. Ojala et al. named these patterns *uniform patterns*. An LBP is called uniform, if it contains at most two bitwise transitions from 0 to 1 or vice versa when the corresponding bit string is considered circular.

After the LBP of each pixel is coded, a histogram is calculated to represent the texture

$$H(k) = \sum_{i=1}^{I}\sum_{j=1}^{J} f(LBP_{P,R}, k), k \in [0, K) \tag{14}$$

where K is the number of LBP pattern values, and I and J are the height and width of the image, respectively. $f(x, y)$ is 1 if $x = y$ and 0 otherwise.

The LBP is defined on a gray image, which is treated as a 2D object. To extract the dynamic texture of a 3D object (such as a gray micro-expression video clip), a dynamic LBP description from three orthogonal planes (LBP-TOP) was formed.

Fig. 2 shows the spatiotemporal volume of a video. It also illustrates the XY plane and the resulting XT and YT planes

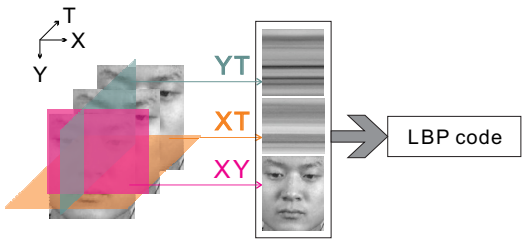


Fig. 2. Illustration of a spatiotemporal volume of a video, the XY plane (original frames) and the resulting temporal planes for LBP feature extraction.

from a single row and column of the volume. The LBP-TOP description is formed by calculating the LBP features from the planes and concatenating the histograms. Intuitively it can be understood that XT and YT planes encode the vertical and horizontal motion patterns respectively.

The original LBP operator was based on a circular sampling pattern; however, different radii and neighborhoods can also be used. An elliptic sampling is used for the XT and YT planes:

$$LBP(x_c, y_c, t_c) = \sum_{p=0}^{P_{plane}-1} s(g_p - g_c)2^p \tag{15}$$

where g_c is the gray value of the center pixel (x_c, y_c, t_c) and g_p are the gray values at the P_{plane} sampling points. $s(u)$ is 1 if $u \geq 0$ and 0 otherwise. P_{plane} can be different on each plane. The gray values g_p are taken from the sampling point: $(x_c - R_x \sin(2\pi p/P_{xt}), y_c, t_c - R_t \cos(2\pi p/P_{xt}))$ on the XT plane and similarly $(x_c, y_c - R_y \sin(2\pi p/P_{yt}), t_c - R_t \cos(2\pi p/P_{yt}))$ on the YT plane. R_d is the radius of the ellipse in the direction of axis d (x, y or t). As the XY plane encodes only the appearance, i.e., both axes have the same meaning, circular sampling is suitable. Values g_p for points that do not fall on pixels are estimated using bilinear interpolation. The length of the feature histogram for LBP-TOP is $2^{P_{xy}} + 2^{P_{xt}} + 2^{P_{yt}}$ when all three planes are considered.

For 4D or higher dimensional objects, the LBP can be extended to higher-dimensional space from the tensor viewpoint. From the conceptual formal, given a pixel c in a D dimensional object, its D dimensional LBP is computed by comparing it with its P neighbors p. The neighbors lie on a D-dimensional hyper-sphere with center c and a radius equal to R. However, the conceptual formal is infeasible. In the higher-dimensional space, a large enough number of neighbors P ensures that the maximum local information of c is coded. This means that the length of the LBP code is very long, beyond the capacity of a PC. An additional problem is how to choose the P neighbors on a D-dimensional hyper-sphere such that the P neighbors are evenly distributed.

To address these problems, we introduce Tensor Orthogonal LBP (TO-LBP). In D-dimensional space, the number of $D-1$-dimensional hyper-planes is D. These $D-1$-dimensional hyper-planes are orthogonal to each other. In each $D-1$-dimensional hyper-plane, we can find a $D-1$-dimensional hyper-sphere, with center c and a radius equal to R. Similarly, in each $D-1$-dimensional space, the number of $D-2$-

dimensional hyper-planes is $D-1$. These $D-2$-dimensional hyper-planes are orthogonal to each other. In each $D-2$-dimensional hyper-plane, we can find a $D-2$-dimensional hyper-sphere, with center c and a radius equal to R. This is a recursive procedure, until the 2-dimensional plane, on which there is a circle with center c and a radius equal to R. Hence, we have $D\times(D-1)\times\ldots\times 3$ circles, each of which represents a special direction in D-dimensional space. When $D=3$, TOTOP degenerates into LBP-TOP.

Although a color micro-expression video clip may be represented as a fourth-order tensor, its mode-4 has only 3 elements. We can therefor use LBP-TOP to extract the dynamic textures from each color component, and then concatenate them as a long feature vector to represent the micro-expression sample. Similar to LBP, LBP-TOP also need to divide the sample into several patches, then code for each patch. In following section, we will design a set of Regions of Interest (ROIs) for coding LBP-TOP.

V. Action Units and Regions of Interest

The Facial Action Coding System (FACS) [36] is an objective method for quantifying facial movement based on a combination of 38 elementary components. These elementary components comprising 32 action units (AUs) and 6 action descriptors (ADs), can be seen as the *phonemes* of facial expressions. As words are temporal combinations of phonemes, micro-expression are spatial combinations of AUs. Each AU depicts a local facial movement. We selected a frontal neutral facial image as the template face and divided it into 16 Regions of Interest (ROIs) based on these AUs. Each ROI corresponds to one or more AUs. Fig. 3 shows the template face, the 16 ROIs and the corresponding AUs. Table I also lists the AU number, FACS name and the corresponding ROI according to [36].

In [36], the AUs are divided into 6 groups. Group 6 *Miscellaneous Actions and Supplementary Codes* is not specific to any muscular basis and has no corresponding ROI, the muscular anatomy and muscular action of the other groups are illustrated in Fig. 4 and Fig. 5, which are taken from [36]. The ROIs are drawn according to the muscular action. In Group 5, AUs were seldom found in micro-expressions, perhaps because they last longer than 500 milliseconds. Thus the ROIs do not take these AUs into account.

Many of the 16ROIs correspond to multiple AUs with different directions of movement. For example, AU16 and AU20 are contained in ROI R_{14} (or R_{13}). The direction of movement of AU16 is vertical (up/down) and that of AU20 is horizontal. So the texture descriptor (LBP or LBP-TOP) on R_{14} therefore has more discriminant power to discriminate between AU16 and AU20.

VI. LBP-TOP on TICS for Micro-expression Recognition

LBP-TOP is a dynamic texture operator and can represent not only appearance information but also motion information. It has already successfully been used for expression recognition [18] and micro-expression recognition [17]. However,

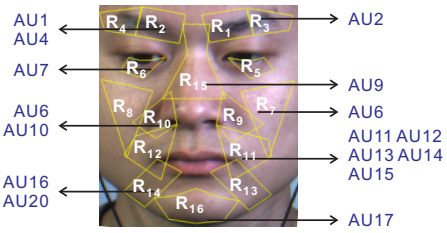


Fig. 3. Template face and 16 ROIs [1].

TABLE I
AU Number, FACS Name and corresponding ROI

AU Number	FACS Name	Region of Interest
Group 1: *Upper Face Action Units*		
AU1	Inner Brow Raiser	R_1, R_2
AU2	Outer Brow Raiser	R_3, R_4
AU4	Brow Lowerer	R_1, R_2
AU5	Upper Lid Raiser	No ROI
AU7	Lid Tightener	R_5, R_6
AU6	Cheek Raiser and Lid Compressor	R_7, R_8, R_9, R_{10}
AU43	Eye Closure - Optional	No ROI
AU45	Blink - Optional	No ROI
AU46	Wink - Optional	No ROI
Group 2: *Lower Face Action Units - Up/Down Actions*		
AU9	Nose Wrinkler	R_{15}
AU10	Upper Lip Raiser	R_9, R_{10}
AU17	Chin Raiser	R_{16}
AU15	Lip Corner Depressor	R_{11}, R_{12}
AU25	Lips Part	No ROI
AU26	Jaw Drop	No ROI
AU27	Mouth Stretch	No ROI
AU16	Lower Lip Depressor	R_{13}, R_{14}
Group 3: *Lower Face Action Units - Horizontal Actions*		
AU20	Lip Stretcher	R_{13}, R_{14}
AU14	Dimpler	R_{11}, R_{12}
Group 4: *Lower Face Action Units - Oblique Actions*		
AU11	Nasolabial Furrow Deepener	R_{11}, R_{12}
AU12	Lip Corner Puller	R_{11}, R_{12}
AU13	Sharp Lip Puller	R_{11}, R_{12}
Group 5: *Lower Face Action Units - Orbital Actions*		
AU18	Lip Pucker	No ROI
AU22	Lip Funneler	No ROI
AU23	Lip Tightener	No ROI
AU24	Lip Presser	No ROI
AU28	Lips Suck	No ROI
Group 6: *Miscellaneous Actions and Supplementary Codes*		
AU8+25	Lips Toward Each Other	No ROI
AU21	Neck Tightener	No ROI
AU31	Jaw Clencher	No ROI
AU38	Nostril Dilator	No ROI
AU39	Nostril Compressor	No ROI

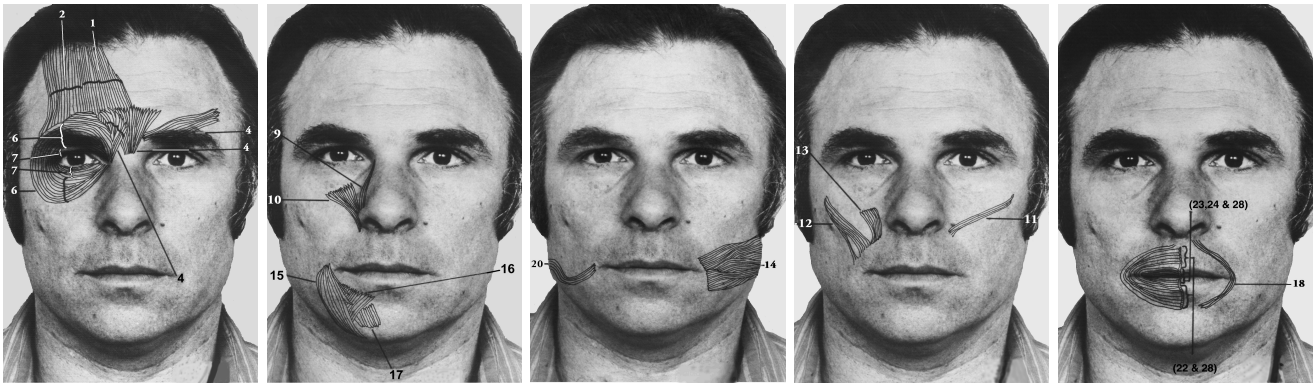


Fig. 4. Muscular Anatomy [36]. The numbers in the figures are the AU numbers. (a) Upper Face Action Units; (b) Lower Face Action Units - Up/Down Actions; (c) Lower Face Action Units - Horizontal Actions; (d) Lower Face Action Units - Oblique Actions; (e) Lower Face Action Units - Orbital Actions.

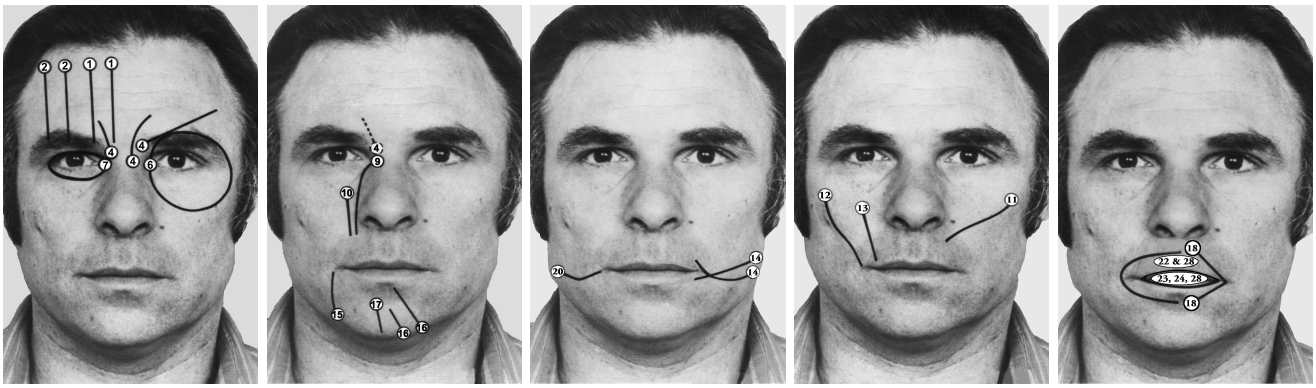


Fig. 5. Muscular Action [36]. The numbers in the figures are the AU numbers. (a) Upper Face Action Units; (b) Lower Face Action Units - Up/Down Actions; (c) Lower Face Action Units - Horizontal Actions; (d) Lower Face Action Units - Oblique Actions; (e) Lower Face Action Units - Orbital Actions.

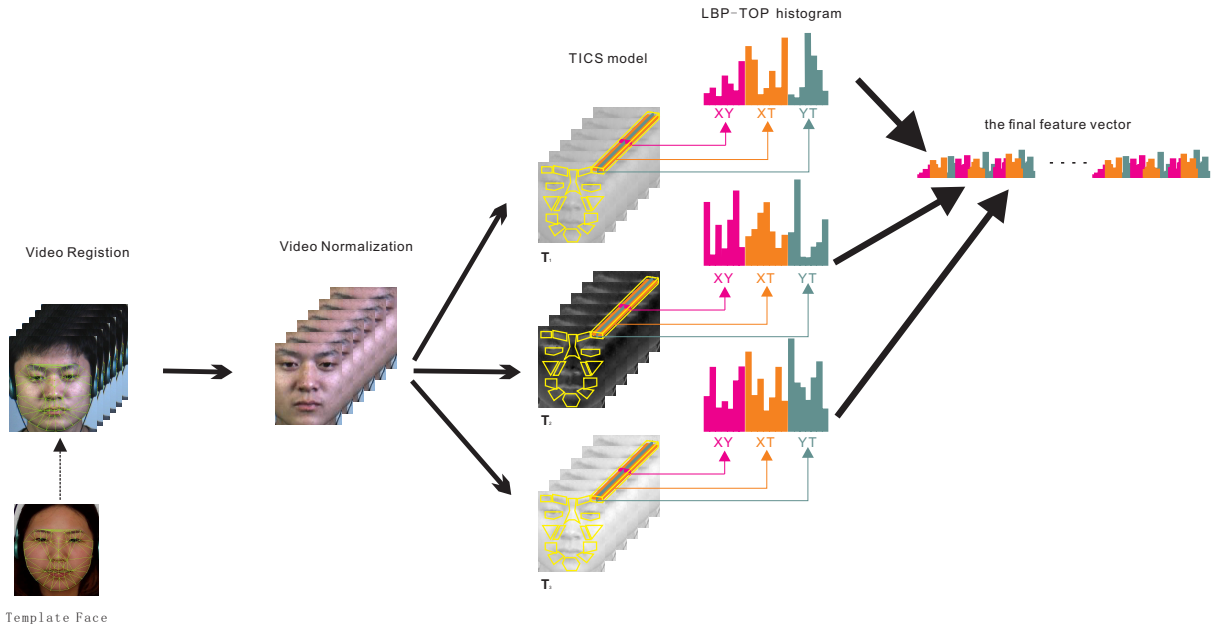


Fig. 6. Level diagram of our method.

only gray video clips were used in these studies. Recent research has shown that the use of color information may significantly improve the discriminative power of LBP [20], while other research has shown that expression recognition achieves better performance in perceptual color spaces [30]. Motivated by these studies, we propose a novel idea to use LBP-TOP on Tensor Independent Color Space (TICS) for micro-expression recognition. Fig. 6 shows the level diagram of our method.

First, we register the micro-expression video to address the large variations in the spatial appearances of faces. A face in the first frame of each video clip was normalized to a template face by registering 68 facial landmark points detected using the Active Shape Model (ASM) [37]. The registration transformation is denoted as T. Others frames are transformed by the same T.

Then, the registered video was normalized. The facial region of each frame was cropped and normalized to 163×134 pixels. The frame numbers of each video clip were normalized to I_3 by using linear interpolation. Hence, each video was normalized to a fourth-order tensor $\mathcal{X}^{163\times134\times I_3\times3}$. Its 4-mode includes 3 color components (R, G and B) in RGB color space.

TICS is performed to transform the 4-mode from RGB into TICS. TICS has three color components $\mathbf{T}_1$, $\mathbf{T}_2$ and $\mathbf{T}_3$. Each color component video was divided in 16 ROIs, and an LBP-TOP histogram was calculated from each ROI. The 48 histograms (3 color components, 16 ROIs) were concatenated to form the final vector.

Fig. 7 illustrates the color components in RGB color space and TICS color space. LBP-TOP is used to extract dynamic texture features from the $\mathbf{T}_1$, $\mathbf{T}_2$ and $\mathbf{T}_3$ components. Fig. 7 also shows the LBP codes on the XT plane in the color components. The LBP codes of the $\mathbf{R}$, $\mathbf{G}$ and $\mathbf{B}$ color components are all 01110000, while the LBP codes of the $\mathbf{T}_1$, $\mathbf{T}_2$ and $\mathbf{T}_3$ color components are 11111000, 00001111 and 1111000, respectively.

Why are the LBP codes of $\mathbf{R}$, $\mathbf{G}$ and $\mathbf{B}$ color components usually are the same? The explanation is given as follows. Given a pixel c in the RGB image, its value in the $\mathbf{R}$ color components is denoted as g_c^r, and the values of its neighbors in the $\mathbf{R}$ color components are denoted as g_p^r $(p = 0, 1, \ldots, 7)$. In the $\mathbf{G}$ color components, the values of the given pixel c and its neighbors are similarly denoted as g_c^g and g_p^g (see Fig. 8). In an extreme case, we assume that $\mathbf{R}$ and $\mathbf{G}$ have a linear correlation.

$$\frac{g_p^r}{g_c^r} = \frac{g_p^g}{g_c^g} = k_p \tag{16}$$

According to Eq. (13), the LBP codes of the given pixel c in color components $\mathbf{G}$ and $\mathbf{B}$ are

$$LBP_r = \sum_{p=0}^{7} s(g_p^r - g_c^r)2^p = \sum_{p=0}^{7} s((k_p - 1)g_c^r)2^p \tag{17}$$

and

$$LBP_g = \sum_{p=0}^{7} s(g_p^g - g_c^g)2^p = \sum_{p=0}^{7} s((k_p - 1)g_c^g)2^p. \tag{18}$$

Because $g_c^r \geq 0$ and $g_c^r \geq 0$, we have

$$s((k_p - 1)g_c^r) = s(k_p - 1) \tag{19}$$

and

$$s((k_p - 1)g_c^g) = s(k_p - 1). \tag{20}$$

Hence, $LBP_r = LBP_g$, which means that the LBP codes of the given pixel c in color components $\mathbf{G}$ and $\mathbf{B}$ are the same. However, Eq. (16) does not always hold. We assume that

$$\frac{g_p^r}{g_c^r} = k_p^r \qquad \text{and} \qquad \frac{g_p^g}{g_c^g} = k_p^g. \tag{21}$$

Hence, we have

$$LBP_r = \sum_{p=0}^{7} s(k_p^r - 1)2^p \tag{22}$$

and

$$LBP_g = \sum_{p=0}^{7} s(k_p^g - 1)2^p. \tag{23}$$

If the conditions $k_p^r \geq 1$ and $k_p^g \geq 1$ or $k_p^r < 1$ and $k_p^g < 1$ are met at same time, we also obtain $LBP_r = LBP_g$.

What is the probability that these conditions are met? An investigation was performed using the CASME and CASME 2 databases. In CASME, the probability that the conditions are met in color components $\mathbf{R}$ and $\mathbf{G}$ is 89.01%, and in CASME 2, the probability is 88.98%.

We use similar method to investigate the probabilities that the conditions are met in each color component pairs in TICS, CIELab, CIELuv and RGB. Table II shows the probabilities that the conditions are met in each color component pairs. Based on the table, we can see that the probability that the LBP codes of the three color components are the same is over 85% in RGB color space. Hence, the combination of the LBP of $\mathbf{R}$, $\mathbf{G}$ and $\mathbf{B}$ color components can not significantly improve the final performance. In TICS, there is at least one color component pair for which the probability of the conditions being met at the same time is very low, which means the probability of the LBP codes in the two color components being the same is very low. Hence, the combination of the LBP of the color components in TICS may significantly improve upon the final performance.

TABLE II
PROBABILITY (%) THAT THE CONDITIONS ARE MET IN COLOR COMPONENT PAIRS.

Database	Component Pairs	TICS	CIELuv	CIELab	RGB
CASME	(1, 2)	56.61	59.14	52.95	89.01
	(1, 3)	42.48	59.30	51.63	87.28
	(2, 3)	**8.83**	65.25	56.62	90.00
CASME 2	(1, 2)	**13.25**	55.99	50.01	88.98
	(1, 3)	21.51	64.81	52.37	86.03
	(2, 3)	71.20	62.68	57.00	89.37

The mutual information of two random variables is a measure of their mutual dependence. Formally, the mutual

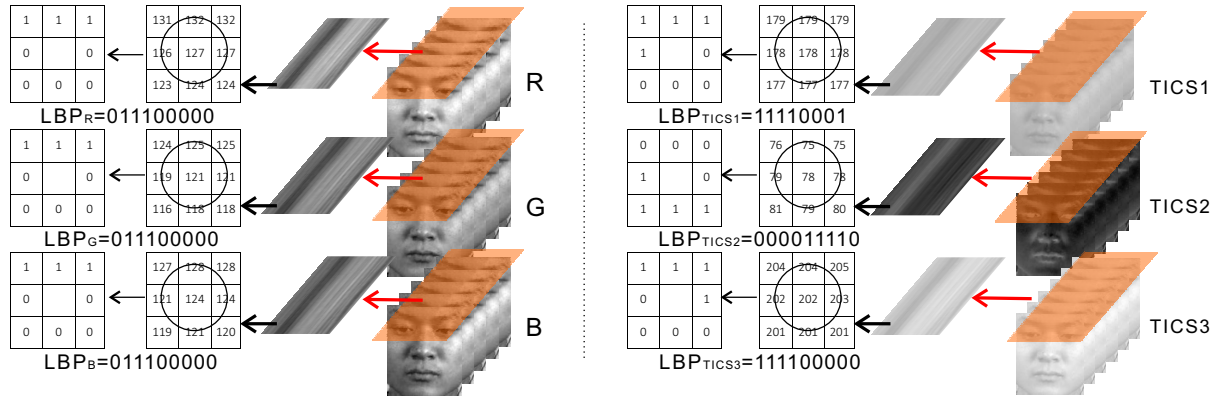


Fig. 7. Illustration of R, G, and B color components, the various components generated by TICS and the corresponding LBP-TOP codes on the XT plane [1].

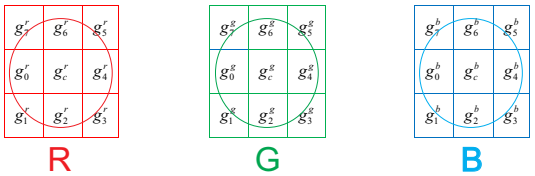


Fig. 8. Example of the color LBP operator.

information of two discrete random variables X and Y can be defined as:

$$I(X;Y) = \sum_{y\in Y}\sum_{x\in X} p(x,y)\log(\frac{p(x,y)}{p(x)p(y)}) \qquad (24)$$

where $p(x,y)$ is the joint probability distribution function of X and Y, and $p(x)$ and $p(y)$ are the marginal probability distribution functions of X and Y respectively. The larger $I(X;Y)$ is, the more sharing information between X and Y there will be. This means that the combination of X and Y provides less additional information.

A face image is randomly picked from CASME 2, and the mutual information is calculated from each pair of color components. $I(LBP_r;LBP_g) = I(LBP_g;LBP_b) = I(LBP_b;LBP_r) = 0.42$, $I(LBP_{\mathbf{T}_1};LBP_{\mathbf{T}_2}) = I(LBP_{\mathbf{T}_2};LBP_{\mathbf{T}_3}) = 0$ and $I(LBP_{\mathbf{T}_1};LBP_{\mathbf{T}_3}) = 0.38$. Hence, the LBP features in TICS are more independent than RGB space such that the performance in TICS superior.

VII. Experiments

A. CASME

The Chinese Academy of Sciences Micro-Expression (CASME) database [38] includes 195 spontaneous facial micro-expressions recorded by two 60 fps cameras. The samples were selected from more than 1500 facial expressions. The selected micro-expressions had either a total duration of less than 500 ms or an onset duration (time from onset frame to apex frame[3]) of less than 250 ms. These samples are coded with the onset, apex and offset frames, and tagged with action units (AUs) [39]. In this database, micro-expressions are labeled into seven categories (happiness, surprise, disgust, fear, sadness, repression and tense). Fig. 9 is an example.

The CASME database is divided into two classes: Set A and Set B. The samples in Set A were recorded with a BenQ M31 consumer camera with 60fps, with the resolution set to 1280×720 pixels. The participants were recorded in natural light. The samples in Set B were recorded with a Point Grey GRAS-03K2C industrial camera with 60fps, with the resolution set to 640×480 pixels. The participants were recorded in a room with two LED lights. Industrial cameras may capture the subtler movements of micro-expressions with higher frame rate, but the color depth is no more than 16-bit, which is far lower than that of consumer cameras. Therefore, Set B is used in the following experiments.

In the experiments, we merged the seven categories into four classes. Such a classification may be more easily applied in practice. *Positive* (4 samples) contains happy micro-expression, which indicates "good" emotions in the individual. *Negative* (47 samples) contains disgust, sadness and fear, which are usually considered "bad" emotions. *Surprise* (13 samples) usually occurs when there is a difference between expectations and reality and can be neutral/moderate, pleasant, unpleasant, positive, or negative. Tense and repression indicate the ambiguous feelings of an individual and require further inference, so they were categorized in the *Other* class (33 samples). We selected 97 samples[4] from Set B. Nonetheless, there is a tremendous difference in the sample numbers in each class. Therefore, in the experiments, we use the Leave-One-Sample-Out (LOSO) cross-validation; i.e., in each fold, one sample is used as the test set, and the others are used as the training set. After 97 folds, each sample has been

[3]The onset is the first frame that changes from the baseline (usually a neutral facial expression). The apex is the frame that reaches the highest intensity of the facial expression. The offset is the last frame of the expression (before turning back to a neutral facial expression). Occasionally, a facial expression faded very slowly, and the changes between frames are very difficult to detect by eyes. For such offset frames, the coders only coded the last clear frame as the offset frame while ignoring the nearly imperceptible changing frames.

[4]There is a sample with 122 frames. It was removed.

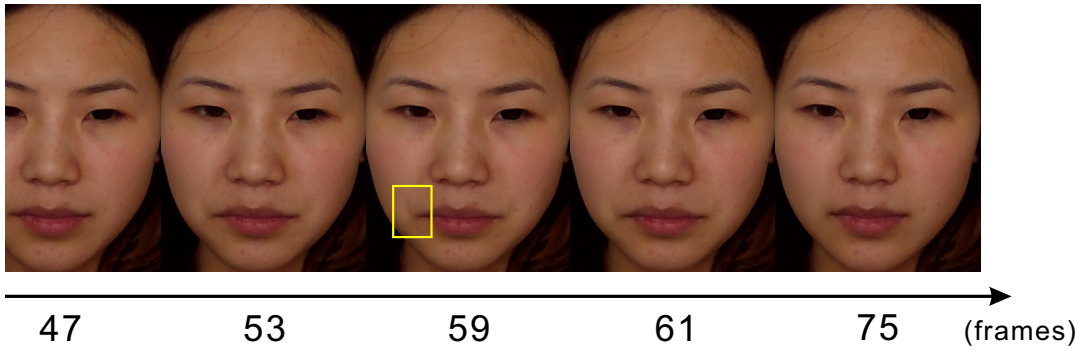


Fig. 9. A demonstration of the frame sequence of a micro-expression in CASME. The AU number for this micro-expression is 15, which indicates *Lip Corner Depressor*.

used as the test set once and the final recognition accuracy is calculated based on all of the results. Of the 97 samples, the frame number of the shortest sample is 10 and that of the longest sample is 68. The frame numbers of all samples are normalized to 70 using linear interpolation. Hence, each sample was normalized to a fourth-order tensor with a size of $163 \times 134 \times 70 \times 3$.

In the experiments, we compared the micro-expression recognition rates in TICS, RGB, gray, and two perceptual color spaces [30]: CIELuv and CIELab. For the gray color space, we extracted LBP-TOP to represent the dynamic texture features for each ROI and built histograms. Then, the histograms were concatenated into a vector as an input for the classifier. A support vector machine (SVM) classifier was selected and used the linear kernel as the kernel function. For the TICS, CIELuv, CIELab, and RGB color spaces, we used the same method to build the LBP-TOP histograms and concatenate them into a vector for each color component. The vectors were concatenated to a long vector as an input for SVM. For LBP-TOP, the radii in axes X and Y (be marked as R_x and R_y) were set as 1 and the radii in axies T (marked as R_t) was assigned various values from 2 to 4. The number of neighboring points (marked as P) in the XY, XT and YT planes were all set to 4 and 8. The uniform pattern and the basic LBP were used in LBP coding. The results are shown in Table III.

From the table, we can see that the performances in the TICS color space is the best in most cases. When $R_t = 2$, CIELab achieves the best performances in the first three cases, but the performances of TICS is better than those of RGB and gray. We can see that the performance of the uniform pattern is the same as that of the basic LBP in most cases, although its code length is far shorter. In addition, the accuracies with $P = 8$ are not better than the accuracies with $P = 4$ in many cases. Therefore, we used the uniform pattern and set P as 4 in the following experiments.

Fig. 10 shows five confusion matrices of TICS, CIELuv, CIELab, RGB, and GRAY in Set B of CASME. No *Positive* sample is misrecognized as *Negative*, and no *Negative* sample is misrecognized as *Positive*. According to the field of psychology, *Positive* and *Negative* expressions are opposites. *Positive* facial expressions usually have distinct differences in their appearance from *Negative* facial expressions. For happiness, AU6 and AU12 are linked to upward movements of the mouth corner, while for *Negative* facial expressions such as disgust, fear and sadness, the mouth corners move downward (such as AU16) and/or there are movements of the eye-brows (such as AU 4) or chin (such as AU 17). *Surprise*, however, can be positive, negative, or neutral, depending on different situations, which makes it more likely to be misrecognized as other categories.

The recognition accuracy of each class of micro-expression (except for *Positive* in TICS) is higher in the TICS, CIELuv, and CIELab color spaces than in the RGB color space. For *Others*, TICS has the highest recognition accuracy 51.52%. For *Negative*, TICS and CIELuv achieved better performances than CIELab. For *Positive*, however, TICS achieved a worse performance than CIELuv and CIELab. From the figure, we can see that the recognition accuracy of *Positive* is lower in RGB than in gray because the number of *Positive* samples was too small (only 4 samples).

B. CASME2

The CASME2 [40] database includes 255 spontaneous facial micro-expressions recorded by two 200 fps cameras. These samples were selected from more than 2,500 facial expressions. Compared with CASME, this spontaneous micro-expression database is improved in its increased sample size, fixed illumination, and higher resolution (both temporal and spatial). This database selected micro-expressions either with a total duration of less than 500 ms or an onset duration (time from the onset frame to apex frame) of less than 250 ms. These samples are coded with the onset and offset frames ans tagged with action units (AUs) and emotions. Fig. 11 is an example. In this database, micro-expressions are labeled into seven categories (happiness, surprise, disgust, fear, sadness, repression and tense).

We also merged the seven categories into four classes: *Positive* (32 samples), *Negative* (66 samples), *Surprise* (25 samples), and *Others* (129 samples). The LOSO cross-validation mentioned in the previous experiment is also used in this experiment. To address the large variations in the

TABLE III
MICRO-EXPRESSION RECOGNITION ACCURACIES (%) IN GRAY, RGB AND TICS COLOR SPACES IN SET B OF CASME.

		TICS	CIELuv	CIELab	RGB	GRAY
$R_t = 2$	$P = 4$, uniform pattern	57.73	58.76	**61.86**	52.58	51.55
	$P = 4$, basic LBP	57.73	58.76	**61.86**	52.58	51.55
	$P = 8$, uniform pattern	59.79	59.79	**60.82**	54.64	52.58
	$P = 8$, basic LBP	**59.79**	**59.79**	58.76	53.61	51.55
$R_t = 3$	$P = 4$, uniform pattern	**61.86**	**61.86**	58.76	55.67	53.61
	$P = 4$, basic LBP	**61.86**	**61.86**	58.76	55.67	53.61
	$P = 8$, uniform pattern	**61.86**	55.67	59.79	54.64	54.64
	$P = 8$, basic LBP	**60.82**	59.79	56.70	54.64	54.64
$R_t = 4$	$P = 4$, uniform pattern	**60.82**	**60.82**	58.76	54.64	54.64
	$P = 4$, basic LBP	**60.82**	**60.82**	58.76	54.64	54.64
	$P = 8$, uniform pattern	57.73	54.64	**59.79**	57.73	54.64
	$P = 8$, basic LBP	**60.82**	59.79	**60.82**	55.67	54.64

		Predicted											
		TICS				**CIELuv**				**CIELab**			
		Positive	Negative	Surprise	Others	Positive	Negative	Surprise	Others	Positive	Negative	Surprise	Others
Ground Truth	Positive	**25.00**	0	0	75.00	**50.00**	0	0	50.00	**50.00**	0	0	50.00
	Negative	0	**80.85**	0	19.15	0	**80.85**	0	19.15	0	**78.72**	2.13	19.15
	Surprise	15.38	23.08	**30.77**	30.77	7.69	23.08	**30.77**	38.46	15.38	23.08	**30.77**	30.77
	Others	9.09	24.24	15.15	**51.52**	9.09	33.33	9.09	**48.48**	9.09	33.33	9.09	**48.48**
		RGB				**GRAY**							
		Positive	Negative	Surprise	Others	Positive	Negative	Surprise	Others				
Ground Truth	Positive	**25.00**	0	0	75.00	**50.00**	0	0	50.00				
	Negative	0	**78.72**	2.13	19.15	0	**74.47**	4.26	21.28				
	Surprise	7.69	38.46	**23.08**	30.77	7.69	30.77	**23.08**	38.46				
	Others	12.12	45.45	9.09	**33.33**	12.12	42.42	12.12	**33.33**				

Fig. 10. Five confusion matrices of TICS, CIELuv, CIELab, RGB, and GRAY in Set B of CASME.

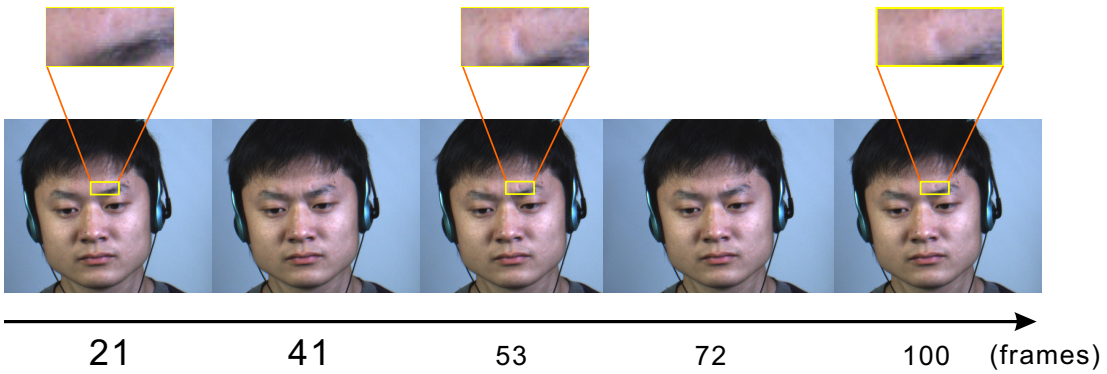


Fig. 11. A demonstration of the frame sequence of a micro-expression in CASME 2. The AU number for this micro-expression is 4, which indicates *Brow Lowerer*. The three rectangles above the images show the right inner brow (AU4) in zoom in mode.

spatial appearances of micro-expressions, we used the same transformation method as in the previous experiments. In the samples, the frame number of the shortest sample is 24, and that of the longest sample is 146. The frame numbers of all samples are normalized to 150 by linear interpolation. The size of each frame is normalized to 163×134 pixels. Therefore, each sample was normalized to a fourth-order tensor with a size of $163 \times 134 \times 150 \times 3$.

To estimate the performance of micro-expression recognition in TICS, CIELuv, and CIELab color spaces, we compared them with RGB and gray color spaces. The radii in axes X and Y were assigned various values from 1 to 4. To avoid too many combinations of parameters, we made $R_x = R_y$. The radius in axis T was assigned various values from 2 to 4. The number of neighboring points in the XY, XT and YT planes was set as 4. A uniform pattern was used in LBP coding. The other settings are the same as in previous experiments. The results are listed in Table IV.

From the table, the performances in the TICS, CIELuv, and CIELab color spaces are better than those of RGB and GRAY in most cases. Both TICS and CIELuv reach the highest recognition accuracy 62.30%. The amount of information in the RGB color space is three times as much as that in gray. However, the accuracy in the RGB color space is sometimes (for example, in $R_x = 2, R_y = 2, R_t = 2$ cases) worse than that in gray. This is due to the large amount of redundant information in the RGB color space in general, which is an obstruction of the further improvement in accuracy in the RGB color space. As the redundancy is removed from the TICS color space, the accuracy is better in general.

Fig. 12 shows the five confusion matrices of TICS, CIELuv, CIELab, RGB, and GRAY in CASME 2. Compared with RGB and GRAY, the recognition accuracies of *Positive*, *Negative* and *Others* in TICS, CIELuv, and CIELab color spaces are improved. The recognition accuracy of *Others* in TICS achieves a highest value of 72.09%, the recognition accuracy of *Positive* in CIELab achieves a highest value of 53.13%, and the recognition accuracy of *Negative* in CIELuv achieves a highest value of 57.58%.

The classical descriptors based on LBP are only applied on gray images. Color information, however, may significantly improve the discriminative power of descriptors. There exist two main methods to combine color and texture cues to improve the discriminative power [41][42].

Early Fusion: Early fusion involves combining color and texture cues at the pixel level. A common way is to compute the texture descriptors on the color channels and then to concatenate them.

$$T_E = [T_R, T_G, T_B] \tag{25}$$

where T can be any texture descriptor.

Late Fusion: Late fusion involves combining color and texture cues at the image level. The color and texture cues are processed independently. The two histograms are then concatenated into a single representation.

$$T_L = [H_T, H_C] \tag{26}$$

where H_T and H_C are explicit texture and color histograms.

The proposed method belongs to early fusion. Following, we also use later fusion in CASME 2. In the experiments, H_T is the histogram of LBP-TOP on gray video. For RGB color space, we use the RGB histogram designated as H_C. The RGB histogram [43] is a combination of three 1D histograms based on the $\mathbf{R}$, $\mathbf{G}$ the and $\mathbf{B}$ color components of the RGB color space. Each histogram is normalized to $[0, 1]$. For TICS color space, the values of TICS are normalized to $[0, 255]$. Three histograms are calculated from the $\mathbf{T}_1$, $\mathbf{T}_2$ and $\mathbf{T}_3$ color components and are then concatenated into H_C. Table V lists the recognition accuracies of early fusion and late fusion, in which the recognition accuracies of early fusion are from Table IV. From the table, we can see that in most cases the recognition accuracies of early fusion are higher than those of late fusion.

TABLE V
MICRO-EXPRESSION RECOGNITION ACCURACIES (%) OF EARLY FUSION AND LATE FUSION IN TICS AND RGB.

	Early Fusion		Later Fusion	
	TICS	RGB	TICS	RGB
$R_x = 1, R_y = 1, R_t = 2$	56.75	**58.33**	56.35	55.56
$R_x = 1, R_y = 1, R_t = 3$	**58.73**	56.35	56.75	55.16
$R_x = 1, R_y = 1, R_t = 4$	**61.90**	58.73	56.75	55.56
$R_x = 2, R_y = 2, R_t = 2$	**59.92**	55.95	55.95	55.16
$R_x = 2, R_y = 2, R_t = 3$	**61.11**	57.54	56.35	55.56
$R_x = 2, R_y = 2, R_t = 4$	**62.30**	59.52	56.35	55.16
$R_x = 3, R_y = 3, R_t = 2$	53.97	54.76	**55.95**	55.16
$R_x = 3, R_y = 3, R_t = 3$	55.16	**56.35**	**56.35**	55.56
$R_x = 3, R_y = 3, R_t = 4$	56.75	**59.52**	56.75	55.56
$R_x = 4, R_y = 4, R_t = 2$	**58.33**	56.35	55.95	55.16
$R_x = 4, R_y = 4, R_t = 3$	**58.33**	53.57	56.35	55.95
$R_x = 4, R_y = 4, R_t = 4$	**57.54**	55.16	56.75	55.95

To investigate the mutual information among the three components in these color spaces, we calculate the LBP-TOP codes for each component of the micro-expression video clips. Then, the mutual information based on these LBP-TOP codes is calculated. Each sample then has three mutual information values: components 1 and 2, components 2 and 3, components 3 and 1. We plot these values in Fig. 13. From the figure, the mutual information values in TICS, CIELuv, and CIELab are smaller than those in RGB color space. This explains why the performances of TICS, CIELuv, and CIELab are better than that of RGB.

We also use the template face as the target image, and produce scatter plots of more than 5,000 randomly chosen data points in the four color spaces. Fig. 14 depicts these scatter plots, which show three pairs of axes plotted against each other. The data points are decorrelated if the data are axis-aligned. The RGB color space shows an almost complete correlation between all pairs of axes because of the data cluster around a line with a 45-degree slope. The TICS color space shows that the correlation between $\mathbf{T}_2$ and $\mathbf{T}_3$ is small. Based on this observation, it is deduced that the combination of $\mathbf{T}_2$ and $\mathbf{T}_3$ maybe achieve better performance.

To verify this assumption, the same experiment was conducted on all pairs of components in TICS. Table VI lists the experimental results. In most cases, the combination of

TABLE IV
MICRO-EXPRESSION RECOGNITION ACCURACIES (%) IN TICS RGB AND GRAY COLOR SPACES IN CASME 2.

Parameters	TICS	CIELuv	CIELab	RGB	GRAY
$R_x = 1, R_y = 1, R_t = 2$	56.75	57.54	**59.52**	58.33	54.37
$R_x = 1, R_y = 1, R_t = 3$	58.73	58.73	**59.13**	56.35	55.16
$R_x = 1, R_y = 1, R_t = 4$	**61.90**	59.13	59.92	58.73	55.95
$R_x = 2, R_y = 2, R_t = 2$	**59.92**	59.92	59.52	55.95	56.35
$R_x = 2, R_y = 2, R_t = 3$	**61.11**	59.92	59.92	57.54	55.16
$R_x = 2, R_y = 2, R_t = 4$	**62.30**	58.33	61.11	59.52	56.75
$R_x = 3, R_y = 3, R_t = 2$	53.97	**59.92**	57.94	54.76	51.59
$R_x = 3, R_y = 3, R_t = 3$	55.16	**60.32**	59.13	56.35	55.14
$R_x = 3, R_y = 3, R_t = 4$	56.75	**61.90**	59.13	59.52	56.75
$R_x = 4, R_y = 4, R_t = 2$	**58.33**	**58.33**	55.56	56.35	51.59
$R_x = 4, R_y = 4, R_t = 3$	58.33	**60.71**	57.53	53.57	50.79
$R_x = 4, R_y = 4, R_t = 4$	57.54	**62.30**	58.33	55.16	55.16

		Predicted											
		TICS				**CIELuv**				**CIELab**			
		Positive	Negative	Surprise	Others	Positive	Negative	Surprise	Others	Positive	Negative	Surprise	Others
Ground Truth	Positive	**50.00**	9.38	6.25	34.38	**43.75**	6.25	6.25	43.75	**53.13**	12.50	6.25	28.13
	Negative	7.58	**54.55**	6.06	31.82	6.06	**57.58**	4.55	31.82	6.06	**53.03**	6.06	34.85
	Surprise	20.00	16.00	**48.00**	16.00	8.00	12.00	**48.00**	32.00	12.00	16.00	**48.00**	24.00
	Others	7.75	17.83	2.33	**72.09**	11.63	20.93	3.10	**64.34**	8.53	18.60	3.10	**69.77**
		RGB				**GRAY**							
		Positive	Negative	Surprise	Others	Positive	Negative	Surprise	Others				
Ground Truth	Positive	**40.63**	9.38	3.13	46.88	**40.63**	12.50	6.25	40.63				
	Negative	6.06	**54.55**	3.03	36.36	6.06	**50.00**	4.55	39.39				
	Surprise	8.00	24.00	**60.00**	8.00	8.00	16.00	**60.00**	16.00				
	Others	10.85	18.60	3.88	**66.67**	13.18	20.93	2.33	**63.57**				

Fig. 12. Five confusion matrices of TICS, CIELuv, CIELab, RGB, and GRAY in CASME 2.

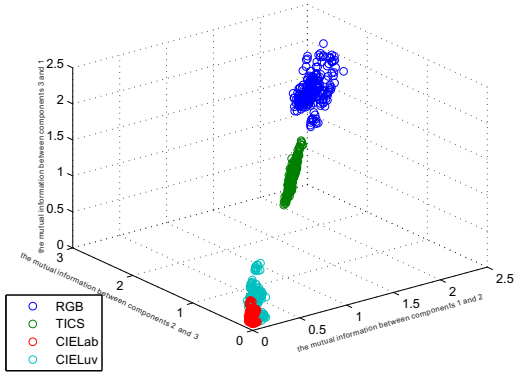
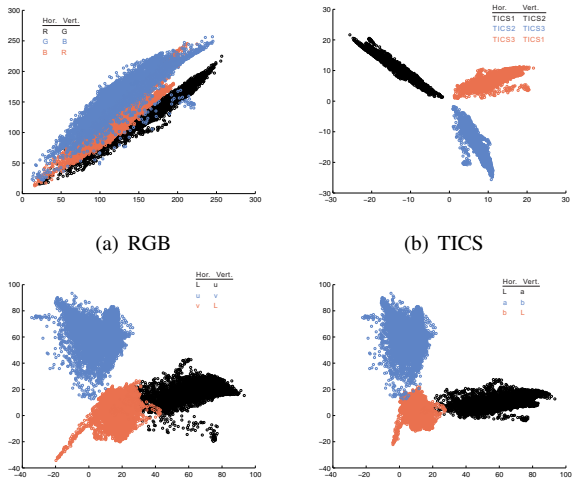


Fig. 13. Mutual information between the three color components of TICS, CIELuv, CIELab, and RGB.



(a) RGB (b) TICS

(c) CIELuv (d) CIELab

Fig. 14. Scatter plots made in TICS, CIELuv, CIELab, and RGB color spaces. Note that the degree of correlation in these plots is defined by the angle of rotation of the mean axis of the point clouds, with rotations of 0 or 90 degrees indicating uncorrelated data and in-between values indicate various degrees of correlation.

$(\mathbf{T}_2, \mathbf{T}_3)$ is the best. Comparing the combination of $(\mathbf{T}_2, \mathbf{T}_3)$ with the combination of $(\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3)$ (see Table IV), we find that their results are almost the same. However, the combination of $(\mathbf{T}_2, \mathbf{T}_3)$ has a lower costs of storage and higher efficiency.

TABLE VI
MICRO-EXPRESSION RECOGNITION ACCURACIES (%) OF THREE DIFFERENT COMBINATIONS OF COMPONENT PAIRS IN TICS.

Components	$\mathbf{T}_1, \mathbf{T}_2$	$\mathbf{T}_2, \mathbf{T}_3$	$\mathbf{T}_1, \mathbf{T}_3$
$R_x = 1, R_y = 1, R_t = 2$	53.97	**56.75**	50.40
$R_x = 1, R_y = 1, R_t = 3$	54.76	**58.73**	51.19
$R_x = 1, R_y = 1, R_t = 4$	52.38	**61.51**	53.17
$R_x = 2, R_y = 2, R_t = 2$	56.75	**59.92**	51.19
$R_x = 2, R_y = 2, R_t = 3$	57.54	**61.51**	52.78
$R_x = 2, R_y = 2, R_t = 4$	55.95	**62.30**	50.79
$R_x = 3, R_y = 3, R_t = 2$	**57.14**	53.97	55.16
$R_x = 3, R_y = 3, R_t = 3$	53.97	55.16	**57.54**
$R_x = 3, R_y = 3, R_t = 4$	55.56	**56.75**	56.35
$R_x = 4, R_y = 4, R_t = 2$	55.95	**58.33**	56.75
$R_x = 4, R_y = 4, R_t = 3$	55.95	**58.73**	55.56
$R_x = 4, R_y = 4, R_t = 4$	54.37	**57.54**	56.75

VIII. CONCLUSION

We have presented a novel color space, Tensor Independent Color Space (TICS) to recognize micro-expression. In TICS, the three color components are as independent from each other as possible. The combination of LBP codes in TICS is thus more effective than that in RGB, and we used the mutual information to explain this. For the locality of LBP, we designed a set of ROIs based on action units such a result. The ROIs can remove some noises such as the nose tip. In this paper, we also showed that the performance of micro-expression recognition is better in the two perceptual color spaces. The experiments on two micro-expression databases revealed that the performances in TICS, CIELuv, and CIELab are better than those in RGB or gray, because their components are as independent from each other as possible.

REFERENCES

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

[2] P. Ekman and W. Friesen, "Nonverbal leakage and clues to deception," DTIC Document, Tech. Rep., 1969.

[3] P. Ekman, "Lie catching and microexpressions," *The Philosophy of Deception*, pp. 118–133, 2009.

[4] M. G. . Frank, M. Herbasz, A. K. K. Sinuk, and C. Nolan., "I see how you. feel: Training laypeople and professionals to recognize fleeting emotions," in *the Annual Meeting of the International Communication Association*, New York, 2009.

[5] M. OSullivan, M. Frank, C. Hurley, and J. Tiwana, "Police lie detection accuracy: The effect of lie scenario," *Law and Human Behavior*, vol. 33, no. 6, pp. 530–538, 2009.

[6] M. Frank, C. Maccario, and V. Govindaraju, *Behavior and security*. Santa Barbara, California: Greenwood Pub Group, 2009, pp. 86–106.

[7] S. Lajevardi and Z. Hussain, "Automatic facial expression recognition: Feature extraction and selection," *Signal, Image and Video Processing*, no. 6, pp. 159–169, 2012.

[8] E. A. Haggard and K. S. Isaacs, *Methods of Research in Psychotherapy*. New York: Appleton-Century-Crofts, 1966, ch. Micromomentary facial expressions as indicators of ego mechanisms in psychotherapy, pp. 154–165.

[9] P. Ekman, "Darwin, deception, and facial expression," *Annals of the New York Academy of Sciences*, vol. 1000, no. 1, pp. 205–221, 2006.

[10] S. Weinberger, "Airport security: Intent to deceive," *Nature*, vol. 465, no. 7297, pp. 412–415, 2010.

[11] W.-J. Yan, Q. Wu, J. Liang, Y.-H. Chen, and X. Fu, "How fast are the leaked facial expressions: The duration of micro-expressions," *Journal of Nonverbal Behavior*, pp. 1–14, 2013.

[12] D. Matsumoto and H. Hwang, "Evidence for training the ability to read microexpressions of emotion," *Motivation and Emotion*, vol. 35, no. 2, pp. 181–191, 2011.

[13] S. Porter and L. Ten Brinke, "Reading between the lies," *Psychological Science*, vol. 19, no. 5, p. 508, 2008.

[14] P. Ekman, "Microexpression training tool (METT)," *San Francisco: University of California*, 2002.

[15] S. Polikovsky, Y. Kameda, and Y. Ohta, "Facial micro-expressions recognition using high speed camera and 3D-gradient descriptor," in *3rd International Conference on Crime Detection and Prevention*. IET, 2009, pp. 1–6.

[16] S.-J. Wang, H.-L. Chen, W.-J. Yan, Y.-H. Chen, and X. Fu, "Face recognition and micro-expression based on discriminant tensor subspace analysis plus extreme learning machine," *Neural Processing Letters*, 2013.

[17] T. Pfister, X. Li, G. Zhao, and M. Pietikainen, "Recognising spontaneous facial micro-expressions," in *12th IEEE International Conference on Computer Vision*. IEEE, 2011, pp. 1449–1456.

[18] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 915–928, 2007.

[19] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *Image Processing, IEEE Transactions on*, vol. 19, no. 6, pp. 1635–1650, 2010.

[20] C. Zhu, C.-E. Bichot, and L. Chen, "Image region description using orthogonal combination of local binary patterns enhanced with color information," *Pattern Recognition*, 2013.

[21] S. H. Lee, H. Kim, Y. M. Ro, and K. N. Plataniotis, "Using color texture sparsity for facial expression recognition," in *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*. IEEE, 2013, pp. 1–6.

[22] R. Mehta and R. J. Zhu, "Blue or red? exploring the effect of color on cognitive task performances," *Science*, vol. 323, no. 5918, pp. 1226–1229, 2009.

[23] M. Villegas, R. Paredes, A. Juan, and E. Vidal, "Face verification on color images using local features," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2008. CVPRW '08*, 2008, pp. 1–6.

[24] C. J. Liu, "Learning the uncorrelated, independent, and discriminating color spaces for face recognition," *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 2, pp. 213–222, 2008.

[25] L. Chengjun and Y. Jian, "ICA color space for pattern recognition," *IEEE Transactions on Neural Networks*, vol. 20, no. 2, pp. 248–257, 2009.

[26] J. Yang and C. Liu, "Color image discriminant models and algorithms for face recognition," *IEEE Transactions on Neural Networks*, vol. 19, no. 12, pp. 2088–2098, 2008.

[27] S.-J. Wang, J. Yang, N. Zhang, and C.-G. Zhou, "Tensor discriminant color space for face recognition," *IEEE Transactions on Image Processing*, no. 9, pp. 2490–2501, 2011.

[28] S.-J. Wang, J. Yang, M.-F. Sun, X.-J. Peng, M.-M. Sun, and C.-G. Zhou, "Sparse tensor discriminant color space for face verification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 6, pp. 876–888, 2012.

[29] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition CVPR 2005*, vol. 1, 2005, pp. 947–954.

[30] S. M. Lajevardi and H. R. Wu, "Facial expression recognition in perceptual color space," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3721–3733, 2012.

[31] G. A. Ramirez, O. Fuentes, S. L. Crites, M. Jimenez, and J. Ordonez, "Color analysis of facial skin: Detection of emotional state," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2014 IEEE Conference on*. IEEE, 2014, pp. 474–479.

[32] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *Siam Review*, vol. 51, no. 3, pp. 455–500, 2009.

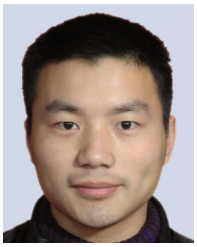
[33] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, *Digital image processing using MATLAB*. Pearson Education India, 2004.

[34] P. Comon, "Independent component analysis, a new concept?" *Signal processing*, vol. 36, no. 3, pp. 287–314, 1994.

[35] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[36] P. Ekman, W. Friesen, and J. Hager, *Facial Action Coding System (The Manual on CD Rom)*. Network Information Research Corporation, 2002.
[37] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models-their training and application," *Computer vision and image understanding*, vol. 61, no. 1, pp. 38–59, 1995.
[38] W.-J. Yan, Q. Wu, Y.-J. Liu, S.-J. Wang, and X. FU, "CASME Database: A dataset of spontaneous micro-expressions collected from neutralized faces," in *10th IEEE Conference on Automatic Face and Gesture Recognition*, 2013, pp. 1–7.
[39] P. Ekman, W. Friesen, and J. Hager, "Facs investigators guide," *A Human Face*, 2002.
[40] W.-J. Yan, X. Li, S.-J. Wang, G. Zhao, Y.-J. Liu, Y.-H. Chen, and X. Fu, "CASME II: An improved spontaneous micro-expression database and the baseline evaluation," *PLoS ONE*, vol. 9, no. 1, p. e86041, 01 2014.
[41] F. S. Khan, J. van de Weijer, S. Ali, and M. Felsberg, "Evaluating the impact of color on texture recognition," in *Computer Analysis of Images and Patterns*. Springer, 2013, pp. 154–162.
[42] T. Mäenpää and M. Pietikäinen, "Classification with color and texture: jointly or separately?" *Pattern Recognition*, vol. 37, no. 8, pp. 1629–1640, 2004.
[43] K. E. Van De Sande, T. Gevers, and C. G. Snoek, "Evaluating color descriptors for object and scene recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 9, pp. 1582–1596, 2010.

Su-Jing Wang (M'12) received the Master's degree from the Software College of Jilin University, Changchun, China, in 2007. He received the Ph.D. degree from the College of Computer Science and Technology of Jilin University in 2012. He was a postdoctoral researcher in Institute of Psychology, Chinese Academy of Sciences from 2012 to 2015. He is now an Assistant Researcher in Institute of Psychology, Chinese Academy of Sciences. He has published more than 40 scientific papers. He is One of Ten Selectees of the Doctoral Consortium at International Joint Conference on Biometrics 2011. He was called as *Chinese Hawkin* by the Xinhua News Agency. His current research interests include pattern recognition, computer vision and machine learning. He serves as an associate editor of Neurocomputing (Elsevier). For more information, visit http://sujingwang.name.



Wen-Jing Yan received his Ph.D degree from Institute of Psychology, Chinese Academy of Sciences, Beijing, China, in 2014. He is now an Assistant Professor in Department of Psychology in Wenzhou University, China. His interests include facial expression and deception. His research interests include interdisciplinary research on facial expression and affective computing.

Xiaobai Li received the B.S. degree in Peking University, Beijing, China, in 2004, and the M.S. degree in the Graduate University of Chinese Academy of Sciences, Beijing, China, in 2007. From 2011 to current, she is doing her PhD study in the Center for Machine Vision Research group, Department of Computer Science and Engineering at University of Oulu, Finland. She is involved in particular on affective computing, facial expression analysis, and Biosignal analysis from facial videos.

Guoying Zhao (SM'12) received the Ph.D. degree in computer science from the Chinese Academy of Sciences, Beijing, China, in 2005. She is currently an Associate Professor with the Center for Machine Vision Research, University of Oulu, Finland, where she has been a researcher since 2005. In 2011, she was selected to the highly competitive Academy Research Fellow position. She has authored or co-authored more than 120 papers in journals and conferences, and has served as a reviewer for many journals and conferences. She has lectured tutorials at ICPR 2006, ICCV 2009, and SCIA 2013, and authored/edited three books and four special issues in journals. Dr. Zhao was a Co-Chair of five International Workshops at ECCV, ICCV, CVPR and ACCV, and two special sessions at FG13 and FG15. She is editorial board member for Image and Vision Computing Journal, International Journal of Applied Pattern Recognition and ISRN Machine Vision. She is IEEE Senior Member. Her current research interests include image and video descriptors, gait analysis, dynamic-texture recognition, facial-expression recognition, human motion analysis, and person identification.

Chun-Guang Zhou PhD, professor, PhD supervisor, Dean of Institute of Computer Science of Jilin University. He is Jilin-province-management Expert, Highly Qualified Expert of Jilin Province, One-hundred Science-Technique elite of Changchun. And he is awarded the Governmental Subsidy from the State Department. He has many pluralities of national and international academic organizations. His research interests include related theories, models and algorithms of artificial neural networks, fuzzy systems and evolutionary computations, and applications of machine taste and smell, image manipulation, commercial intelligence, modern logistic, bioinformatics, and biometric identification based on computational intelligence. he has published over 168 papers in Journals and conferences and he published 1 academic book.

Xiaolan Fu (M'13) received her Ph. D. degree in 1990 from Institute of Psychology, Chinese Academy of Sciences. Currently, she is a Senior Researcher at Cognitive Psychology. Her research interests include visual and computational cognition: (1) attention and perception, (2) learning and memory, and (3) affective computing. At present, she is the director of Institute of Psychology, Chinese Academy of Sciences and Vice Director, State Key Laboratory of Brain and Cognitive Science.

Minghao Yang now is an associate professor in Institute of Automation, Chinese Academy of Sciences. His current research interests include Speech Generation Visualization, Multimodal Human-Interaction and Emotion Recognition. He published over 20 papers on ACM MultimediaMTAPJUMIICASSP and other important international journals or conferences. He serves as the chair or program committee member for several major conferences, including ACII 2015, NCMMSC 2015, MLSP 2011 etc. He obtained the top winner award of ACM Multimedia Audio/Visual Emotion Challenge Workshop (2015), the best paper nomination of HMMME 2015, the best paper award nominated by ACM MM Audio/Visual Emotion Challenge Workshop (2014), Beijing Science & Technology Progress Award Grade 2 (2014), and the best paper award of HMMME 2013.

Jianhua Tao (M'98) received the M.S. degree from Nanjing University, Nanjing, China, in 1996 and the Ph.D. degree from Tsinghua University, Beijing, China, in 2001. He is currently a Professor with the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing. His current research interests include speech synthesis and recognition, human-computer interaction, and emotional information processing. He has published more than 60 papers in major journals and proceedings, such as the IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, ICASSP, Interspeech, ICME, ICPR, ICCV, ICIP, etc. In 2006, he was elected as Vice-Chairperson of the ISCA Special Interest Group of Chinese Spoken Language Processing (SIG-CSLP), and Executive Committee member of the HUMAINE association. He is the Editorial Board Member for the Journal on Multimodal User Interfaces (JMUI), the International Journal of Synthetic Emotions (IJSE), and the Steering Committee Member for the IEEE TRANSACTIONS ON AFFECTIVE COMPUTING.