Micro-expression Recognition Using Dynamic Textures on Tensor Independent Color Space

Su-Jing Wang[1,*], Wen-Jing Yan[1], Xiaobai Li[2], Guoying Zhao[2] and Xiaolan Fu[1]

1) State Key Laboratory of Brain and Cognitive Science,
Institute of Psychology, Chinese Academy of Sciences, Beijing, 100101, China
2) Department of Computer Science and Engineering, University of Oulu, P. O. Box 4500, FI-90014, Finland
{wangsujing, yanwj}@psych.ac.cn; {lxiaobai, gyzhao}@ee.oulu.fi; fuxl@psych.ac.cn

Abstract—**Micro-expression is a brief involuntary facial expression which reveals genuine emotions and helps detect lies. It intrigues psychologists and computer scientists' (especially on computer vision and pattern recognition) interests due to its promising applications in various fields. Recent research reveals that color may provide useful information for expression recognition. In this paper, we propose a novel color space model, Tensor Independent Color Space (TICS), for enhancing the performance of micro-expression recognition. An 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 in the independent color components can get higher accuracy than that in RGB. In addition, we define a set of Regions of Interest (ROIs) based on Facial Action Coding System (FACS) and calculated the dynamic texture histograms for each ROI. The experiments are conducted on two micro-expression databases, CASME and CASME 2, and the results show that the performance in TICS is better than that in RGB or gray.**

I. INTRODUCTION

Micro-expression is a brief and subtle facial expression which reveals an emotion that a person tries to conceal, especially in high-stake situations [1]. Compared with ordinary facial expressions, micro-expression has two significant characteristics: short duration and low intensity [2]. The importance of micro-expression study is derived from its potential practical applications in many fields, such as clinical diagnosis, national security and interrogations [3][4][5].

Though micro-expression has potential applications in various fields, human beings are difficult to detect and recognize them. The difficulty may stem from its short duration, low intensity and fragmental action units [1][6]. Though there is a debate on the duration, the generally accepted upper limit duration is 0.5 second [7]. The micro-expressions are usually very subtle since individuals are trying to control and repress. In addition, micro-expressions usually present only part of the action units of full-stretched facial expressions [8]. These three characteristics contribute to the difficulty of detecting and recognizing them. In order to improve the human's performance in recognizing micro-expression, Ekman [9] developed the Micro-Expression Training Tool (METT), which trains people

to better recognize the seven categories of micro-expressions. However, Frank [5] found out that the performance in detecting and recognizing micro-expressions peaked around 40% even with the help of METT. Along with the rapid development of computational methods, researchers turn to the field of computer vision and pattern recognition to search for automatic micro-expression recognition.

Up to now, there are only a limited number of researches on automatic micro-expression recognition. Polikovsky *et al.* [10] used 3D-gradient descriptor for micro-expression recognition. Wang *et al.* [11] 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-expressiones. Pfister *et al.* [12] utilized a temporal interpolation model (TIM) [13] based on Laplacian matrix to normalize the frame numbers of spontaneous micro-expression video clips. Then, the LBP-TOP [14] is used to extract the dynamic texture features of micro-expressions and multiple kernel learning is used to classify the features.

However, these methods didn't take color into account. Color is a fundamental aspect of human perception, and its effects on cognition and behavior have attracted many interests of generations of researchers [15]. Recent research efforts revealed that color may provide useful information for face recognition. Wang *et al.* [16] presented a Tensor Discriminant Color Space (TDCS) model which used a 3rd-order tensor to represent a color facial image. To be more robust for the noise, they [17] used elastic net to make TDCS sparse and proposed Sparse Tensor Discriminant Color Space (STDCS). Lajevardi and Wu [18] treated a color facial expression image as a 3rd-order tensor and showed that the perceptual color spaces (CIELab and CIELuv) are better for facial expression recognition than other color spaces.

In this paper, we propose a novel color space model, Tensor Independent Color Space (TICS). In this color space, color components are as independent as possible. The dynamic textures of micro-expressions are extracted from the color components. The independent dynamic textures lead to the better performance of TICS.

II. TENSOR FUNDAMENTALS

In this section, we introduce a general knowledge of tensor. For details, please refer to [19]. A tensor is a multidimensional array, or more formally, a N-order tensor is an element of the tensor product of N vector spaces. The order of a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is N. An element of $\mathcal{A}$ denoted by $\mathcal{A}_{i_1 i_2 \ldots i_N}$, where $1 \leq i_n \leq I_n$, $n = 1, 2, \ldots, N$. The mode-n unfolding matrix of $\mathcal{A}$, denoted by $\mathbf{A}_{(n)} \in \mathbb{R}^{I_n \times (I_1 \times \cdots \times I_{n-1} \times I_{n+1} \times \cdots \times I_N)}$. The mode-$n$ product of a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ by a matrix $\mathbf{U} \in \mathbb{R}^{J_n \times I_n}$ is a $(I_1 \times \cdots \times I_{n-1} \times J_n \times I_{n+1} \times \cdots \times I_N)$-tensor of which the entries are given by:

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

By using tensor decomposition, any tensor $\mathcal{A}$ can be expressed as the product:

$$\mathcal{A} = \mathcal{S} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \cdots \times_N \mathbf{U}_N \tag{2}$$

where $\mathcal{S} = \mathcal{A} \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T \cdots \times_N \mathbf{U}_N^T$ is called the core tensor, $\mathbf{U}_n, n = 1, 2, \ldots, N$ is an orthogonal matrix and contains the ordered principal components for the mode-n.

III. TENSOR INDEPENDENT COLOR SPACE (TICS)

A micro-expression color video clip is naturally represented by a fourth-order tensor, where mode-1 and mode-2 of a tensor are facial spatial information, mode-3 of tensor is the temporal information and mode-4 of tensor is the color space information. For instance, a color micro-expression video clip with the 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 the frames and $I_4 = 3$ has 3 components corresponding to R, G and B in RGB space. However, the R, G and B components are correlated. We first transformed the three correlated components into a series of uncorrelated components $TICS_1$, $TICS_2$ and $TICS_3$, then extracted the dynamic texture features from each uncorrelated component. That can get better results.

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

$$\begin{aligned} \mathcal{Y}_i &= \mathcal{X}_i \times_4 \mathbf{U}_4^T, \\ i &= 1, 2, \ldots, M. \end{aligned} \tag{3}$$

such that the components of mode-4 of $\mathcal{Y}_i$ are as independent as possible. In order to obtain $\mathbf{U}_4$, we use ICA[1] to decorrelate the RGB color space. M fourth-order tensor $\mathcal{X}_i$ are concatenated to a fifth-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 $\mathbf{F}_{(4)}$ corresponding to the three components in RGB space, respectively.

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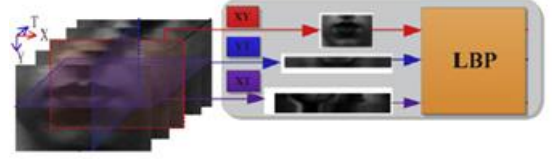

Fig. 1. Illustration of a spatiotemporal volume of a video, the XY plane (original frames) and the resulting temporal planes for LBP feature extraction.

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

$$\Sigma_F = \mathbf{U}_4^{-1} \nabla \mathbf{U}_4^{-T} \tag{4}$$

where $\nabla \in \mathbb{R}^{3 \times 3}$ 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 independent three components. The $\mathbf{U}_4$ in Eq. (4) can be derived using Comon's ICA algorithm by calculating mutual information and high-order statistics.

IV. LBP DESCRIPTION FROM THREE ORTHOGONAL PLANES

Local Binary Patterns (LBP) [20] operator was extended to a dynamic texture operator, where the dynamic LBP description from three orthogonal planes (LBP-TOP) of a space time volume was formed.

Fig. 1 shows the spatiotemporal volume of a video. It also illustrates the XY plane and the resulting XT and YT planes from a single row of 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 thought that XT and YT planes encode the vertical and horizontal motion patterns respectively.

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

$$LBP(x_c, y_c, t_c) = \sum_{p=0}^{P_{plane}-1} s(g_p - g_c) 2^p \tag{5}$$

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 sampling points: $(x_c - R_x \sin(2\pi p/P_{xt}), y_c, t_c - R_t \cos(2\pi p/P_{xt}))$ on 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 YT plane. R_d is the radius of the ellipse to direction of the 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. The 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.
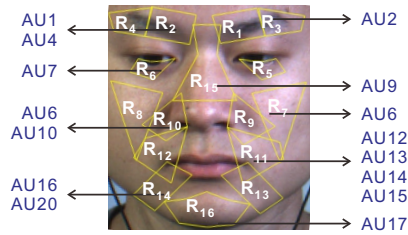
Fig. 2. The template face and 16 ROIs.

Fig. 3. The process of micro-expression registration.

V. ACTION UNIT AND REGION OF INTEREST

The Facial Action Coding System (FACS) [21] is an objective method for quantifying facial movement based on a combination of 57 elementary components. These elementary expressions, known as action units (AUs) and action descriptors (ADs), can be seen as the *phonemes* of facial expressions: words are temporal combinations of phonemes. Similar to facial expressions, micro-expressions 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 the template face into 16 Regions of Interest (ROIs). These ROIs are not exactly corresponding to the AUs. Since there are overlaps between some AUs, the ROIs were modified to be more independent between each other. The template face is used not only to draw ROIs but also to avoid the large variations in the spatial appearances of micro-expressions. The detailed can be found in Section VI.

Fig 2 shows the template face, the 16 ROIs and the AUs corresponding to the ROIs. Fox example, ROI R_1 (or R_2) corresponds to AU1 and AU4 which represent the movements of inner eyebrows. Table I lists the 16 ROIs, the corresponding AUs and the facial movements. The ROIs are drawn to exclude some noise, such as the nose tip and the eye ball movement.

TABLE I
ROIs, THE CORRESPONDING AUs AND THE FACIAL MOVEMENTS. [a]

ROIs	AUs	Facial Movements
R_1, R_2	AU1, AU4	inner eyebrows
R_3, R_4	AU2	outer eyebrows
R_5, R_6	AU7	lower eyelid
R_7, R_8	AU6	cheeks
R_9, R_{10}	AU6, AU10	side of the nose
R_{11}, R_{12}	AU12, AU13, AU14, AU15	mouth corner
R_{13}, R_{14}	AU16, AU20	side of the chin
R_{15}	AU9	nose root
R_{16}	AU17	chin

[a]ROIs are not exactly corresponding to the AUs. since there are overlaps between some AUs, the ROIs were modified to be more independent between each other.
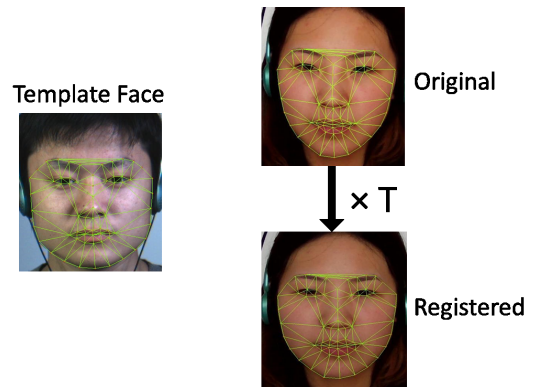
VI. LBP-TOP ON TICS FOR MICRO-EXPRESSION RECOGNITION

LBP-TOP is a dynamic texture operator and can extract not only appearance information but also motion information. It is already successfully used for expression recognition [14] and micro-expression recognition [12]. However, only gray video clips were used in [14] and [12]. Recent research shows expression recognition gets better performance in the perceptual color spaces [18]. Motivated by these researches, we propose an novel idea to use LBP-TOP on Tensor Independent Color Space (TICS) for micro-expression recognition.

To address the large variations in the spatial appearance of faces, all faces were normalized to a template face by registering 68 facial landmark points detected using the Active Shape Model (ASM). First, we selected a frontal face image M with neutral expression as the template, and the coordinates of 68 landmarks of the template face were detected by ASM as ψ_M. Second, for a sample micro-expression clip we detected the 68 facial landmarks on its first frame as ψ_{f1}, and estimated the 2D geometric transformation of the template face and the current given sample face as: $\psi_M = T\psi_{f1}$, where T is the transformation matrix. Third, we registered the sample face to the template by applying the transformation T to all frames of the micro-expression clip. Because there is no much head movement in the video clip, the transformation T can be used to all the frames in the same video clip. The sizes of each frame of samples are normalized to 163×134 pixels.

The process is illustrated in Fig. 3. As the figure shows, all sample faces have the same spatial appearance as the template face after the registration process, which facilitates the comparison of local LBP-TOP features for the classification of micro-expressions.

The frame number of each sample is normalized to I_3 by using linear interpolation, thus each sample was normalized to an fourth-order tensor $\mathcal{X}^{163 \times 134 \times I_3 \times 3}$. Its mode-4 includes 3 color components (R, G and B) in RGB color space. TICS is done to transform mode-4 from RGB into a new color space, in which three color components are as independent as possible. Fig. 4 illustrates the color components in RGB color space and TICS color space. LBP-TOP is used to extract dynamic texture

features from the $TICS1$, $TICS2$ and $TICS3$ components, respectively. Fig. 4 also shows LBP codes on XT plane in the color components. The LBP codes of R, G and B color components all are 01110000. However, the LBP codes of $TICS1$, $TICS2$ and $TICS3$ color components are 11111000, 00001111 and 1111000, respectively. So, the performance in TICS color space is better than that in RGB color space. After extracting LBP codes, the histograms of LBP on each ROI are calculated and concatenated into a vector as inputs for classifier.

VII. EXPERIMENTS

A. CASME

The Chinese Academy of Sciences Micro-Expression (CASME) database [2][22] includes 195 spontaneous facial micro-expressions recorded by two different 60 fps cameras. These samples were selected from more than 1500 facial expressions. The selected micro-expressions either have a total duration less than 500 ms or an onset duration (time from onset frame to apex frame) less than 250 ms. These samples are coded with the onset, apex and offset frames, furthermore tagged with AUs. In this database, micro-expressions are classified into 7 categories (happiness, surprise, disgust, fear, sadness, repression and tense). Fig. 5 is an example.

CASME database is divided into two sets: Set A and Set B. The samples in Set A were recorded by BenQ M31 consumer camera with 60 fps, and the resolution 1280×720 pixels. The participants were recorded in natural illumination. The samples in Set B were recorded by Point Grey GRAS-03K2C industrial camera with 60fps, and the resolution 640×480 pixels. The participants were recorded in a room with two LED lights. Set B captured by industrial cameras, the color depth of which is lower than that of consumer cameras, are more suitable for color space transformation. So, Set B is used in the following experiments.

In the experiments, we merged the 7 categories into 4 classes. Such classification may be more easily applied in practice. Positive contains happy micro-expression, which indicates "good" emotion for the individual. Negative contains disgust, sadness and fear, which are usually reflected as "bad" emotions. Surprise 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, thus were categorized into another class. We selected 97 samples from Class B. In 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 by using linear interpolation. So, each sample was normalized to an fourth-order tensor with the size of $163 \times 134 \times 70 \times 3$.

In the experiments, we compared the micro-expression recognition accuracies in TICS, RGB and gray. A support vector machine (SVM) classifier was selected and used the linear kernel as the kernel function. We used the leave-one-subject-out cross-validation in the experiments. For LBP-TOP, the radii in axes X and Y (marked as R_x and R_y) were set as

1 and the radii in axes 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 all were set as 4 or 8. The uniform pattern and the basic LBP were used in LBP coding. The results are listed in Table II.

		TICS	RGB	GRAY
$R_t = 2$	$P = 4$, uniform pattern	**57.7320**	52.5773	51.5464
	$P = 4$, basic LBP	**57.7320**	52.5773	51.5464
	$P = 8$, uniform pattern	**59.7938**	54.6392	52.5773
	$P = 8$, basic LBP	**59.7938**	53.6082	51.5464
$R_t = 3$	$P = 4$, uniform pattern	**61.8557**	55.6701	53.6082
	$P = 4$, basic LBP	**61.8557**	55.6701	53.6082
	$P = 8$, uniform pattern	**61.8557**	54.6392	54.6392
	$P = 8$, basic LBP	**60.8247**	54.6392	54.6392
$R_t = 4$	$P = 4$, uniform pattern	**60.8247**	54.6392	54.6392
	$P = 4$, basic LBP	**60.8247**	54.6392	54.6392
	$P = 8$, uniform pattern	**57.7320**	57.7320	54.6392
	$P = 8$, basic LBP	**60.8247**	55.6701	54.6392

Among the three color spaces, the performance in TICS got the best in each case. The recognition accuracy in TICS color space is equal to that in RGB color space when $R_t = 4$, $P = 8$ and the uniform pattern is used. In most cases, The performance of the uniform pattern is the same with that of the basic LBP, but the code length of the uniform pattern is far shorter than that of the basic LBP. In addition, the accuracies with $P = 8$ are not better than the those with $P = 4$ in many cases. Therefore, we used the uniform pattern and set P as 4 in the following experiments.

B. CASME2

The CASME2 [23] database includes 246 spontaneous facial micro-expressions recorded by a 200 fps camera. These samples were selected from more than 2,500 facial expressions. Compared with CASME, the database is improved in increased sample size, fixed illumination, and higher resolution (both temporal and spatial). The selected micro-expressions in this database either had a total duration less than 500 ms or an onset duration (time from onset frame to apex frame) less than 250 ms. These samples are coded with the onset and offset frames, as well as tagged with AUs and emotions. Fig. 6 is an example. There are 5 classes of the micro-expressions in this database: happiness (32 samples), surprise (25 samples), disgust (60 samples), repression (27 samples) and tense (102 samples).

In this database, 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 using linear interpolation. The size of each frame is normalized to 163×134 pixels. So, each sample was normalized to an fourth-order tensor with the size of $163 \times 134 \times 150 \times 3$.

To estimate the performance in TICS color space, we compared the micro-expression recognition performance among TICS, RGB and gray color spaces. The radii in axes X and Y
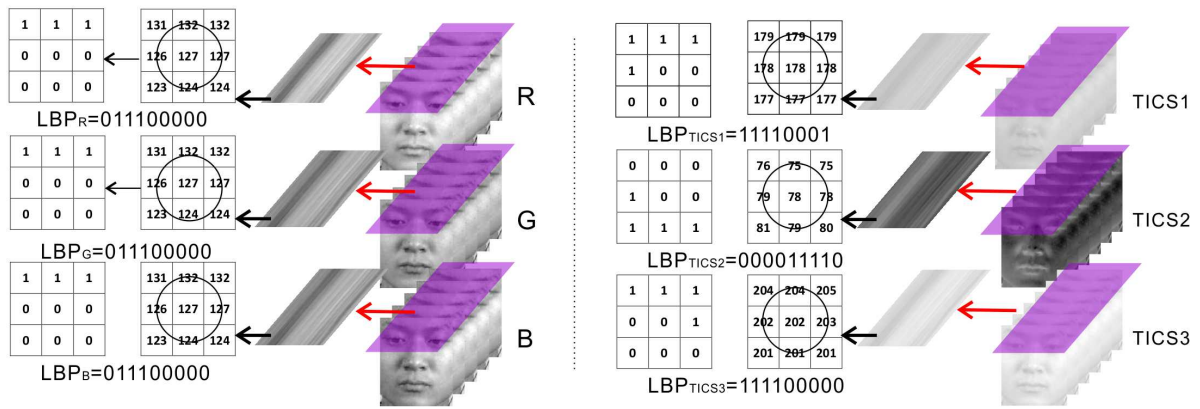
Fig. 4. Illustration of R, G, and B color components, the various components generated by TICS and corresponding LBP-TOP codes on XT plane.
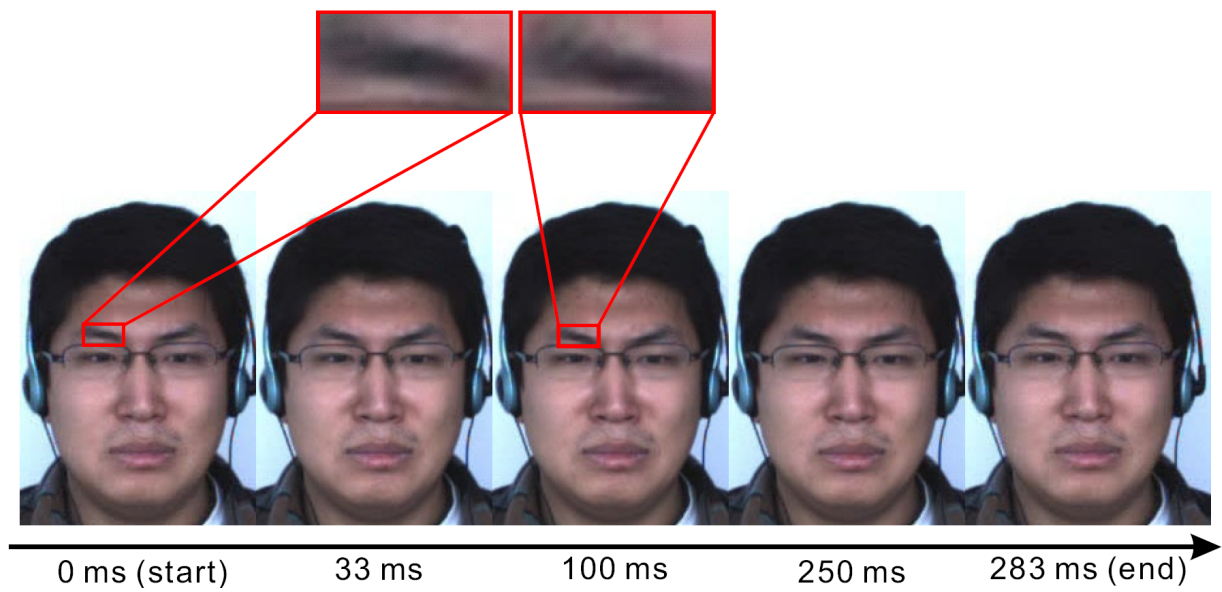

Fig. 5. An demonstration of the frame sequence in a micro-expression. The apex frame presents at about 100 ms. The AUs for this micro-expression is 4+10, which indicates disgust.
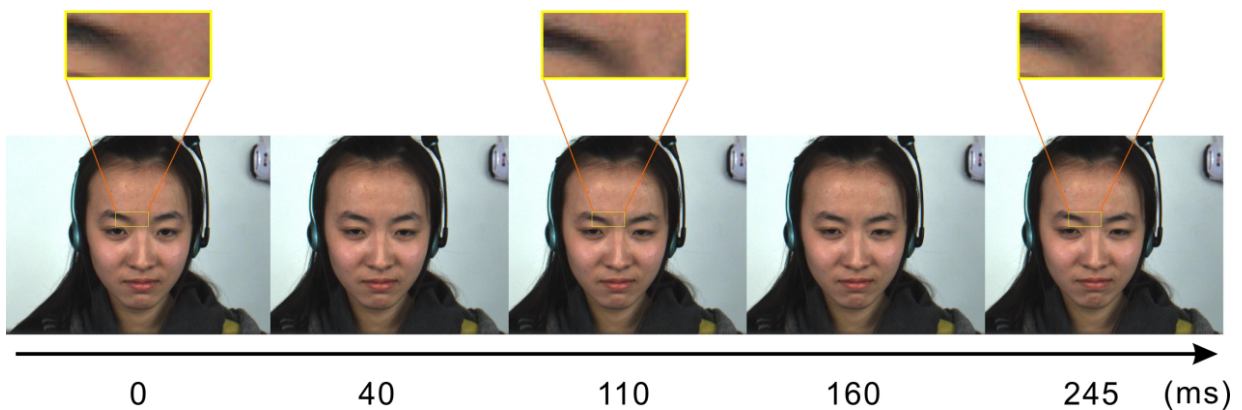

Fig. 6. A demonstration of the frame sequence in a micro-expression. The apex frame presents at about 110 ms. The AUs for this micro-expression is 4+9 (with AU 17 kept almost unchanged), which indicates disgust. The three rectangles above the images show the right inner brow (AU 4) in *zoom in* mode.

were assigned various values from 1 to 4. To avoid too much combinations of parameters, we made $R_x = R_y$. The radii in axes T were assigned various values from 2 to 4. The number of neighboring points in the XY, XT and YT planes were all set as 4. The uniform pattern was used in LBP coding. The other settings are the same with the previous experiments. The results are listed in Table III.

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

Parameters	TICS	RGB	GRAY
$R_x = 1, R_y = 1, R_t = 2$	54.0650	**54.4715**	52.4390
$R_x = 1, R_y = 1, R_t = 3$	**56.0976**	**56.0976**	53.2520
$R_x = 1, R_y = 1, R_t = 4$	**58.5366**	54.8780	54.0650
$R_x = 2, R_y = 2, R_t = 2$	**57.3171**	52.8845	55.2846
$R_x = 2, R_y = 2, R_t = 3$	**56.5041**	54.8780	54.4715
$R_x = 2, R_y = 2, R_t = 4$	**58.1301**	56.0976	56.9106
$R_x = 3, R_y = 3, R_t = 2$	54.4715	**54.8780**	51.2195
$R_x = 3, R_y = 3, R_t = 3$	**54.0650**	52.8455	52.8455
$R_x = 3, R_y = 3, R_t = 4$	**55.6911**	54.4715	52.8455
$R_x = 4, R_y = 4, R_t = 2$	**54.8780**	52.8455	51.2195
$R_x = 4, R_y = 4, R_t = 3$	**55.6911**	52.4390	50.4065
$R_x = 4, R_y = 4, R_t = 4$	**54.0650**	53.6585	53.6585

From the table, the performances in TICS color space got the best in most cases. For the various parameters, the best accuracy in TICS color space is 58.6366% and occurs in $R_x = 1, R_y = 1$ and $R_t = 4$. While the best accuracies in RGB and gray color space are 56.0976% and 56.9106%, respectively. The amount of information in RGB color space is three times as much as that in gray. However, the accuracy in RGB color space is sometimes worse than that in gray. This is derived from the lot of redundant information in RGB color space in general. The redundancy is an obstruction of further improvement of the accuracy in RGB color space. As the redundancy is removed from TICS color space, the accuracy in TICS color space is better in general.

VIII. CONCLUSION

We have presented an novel color space TICS and extracted LBP-TOP features of micro-expression. In this paper, we also proposed a set of ROIs and calculated the histograms of LBP-TOP on each ROI as inputs for SVM. The experiments on two micro-expression databases revealed that the performance in TICS is better than that in RGB or gray.

ACKNOWLEDGMENT

REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] 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.

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

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

[7] 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.

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

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

[10] 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.

[11] 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.

[12] 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.

[13] Z. Zhou, G. Zhao, and M. Pietikainen, "Towards a practical lipreading system," in *2011 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2011, pp. 137–144.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

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

[20] 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.

[21] P. Ekman and W. V. Friesen, *Facial action coding system: A technique for the measurement of facial movement*. CA: Consulting Psychologists Press, 1978, vol. 12.

[22] W.-J. Yan, S.-J. Wang, Y.-J. Liu, Q. Wu, and X. Fu, "For micro-expression recognition: Database and suggestions," *Neurocomputing*, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925231214002227

[23] 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.