

LBP with Six Intersection Points: Reducing Redundant Information in LBP-TOP for Micro-expression Recognition

Yandan Wang¹(✉), John See², Raphael C.-W. Phan¹, and Yee-Hui Oh¹

¹ Faculty of Engineering,
Multimedia University (MMU), Cyberjaya, Malaysia
{yandvn,yeehui716}@gmail.com, raphael@mmu.edu.my

² Faculty of Computing and Informatics,
Multimedia University (MMU), Cyberjaya, Malaysia
johnsee@mmu.edu.my

Abstract. Facial micro-expression recognition is an upcoming area in computer vision research. Up until the recent emergence of the extensive CASMEII spontaneous micro-expression database, there were numerous obstacles faced in the elicitation and labeling of data involving facial micro-expressions. In this paper, we propose the Local Binary Patterns with Six Intersection Points (LBP-SIP) volumetric descriptor based on the three intersecting lines crossing over the center point. The proposed LBP-SIP reduces the redundancy in LBP-TOP patterns, providing a more compact and lightweight representation; leading to more efficient computational complexity. Furthermore, we also incorporated a Gaussian multi-resolution pyramid to our proposed approach by concatenating the patterns across all pyramid levels. Using an SVM classifier with leave-one-sample-out cross validation, we achieve the best recognition accuracy of 67.21 %, surpassing the baseline performance with further computational efficiency.

1 Introduction

Facial (macro-)expression recognition is a popular research area that has seen tremendous advancement in the past few decades. Indeed, macro-expression recognition research has reported accuracies of over 90 % for the six basic facial expressions (i.e. anger, disgust, surprise, fear, sadness and happiness).

In contrast, facial *micro-expression* recognition has recently seen more emphasis in the computer vision community, and addresses a more challenging research problem than its macro-expression counterpart.

A micro-expression is defined as a brief facial movement that reveals an emotion that a person tries to conceal [2]. Micro-expressions are distinctly different from macro-expressions in the aspect of its short duration and occurrence as a response towards a presented emotional stimuli. Its imperceptibility to the naked eyes is the primary motivation towards achieving machine detection and recognition of micro-expressions. There is also a notable lack of well-established

databases due to difficulties in proper elicitation and labeling of micro-expression data. In current literature (and to our best knowledge), there are only two spontaneous micro-expression databases, i.e. SMIC [4] and CASME [10]/CASMEII [12] (both of these CASME variants can be seen as one as the former is a subset of the latter), while there are very few other works to date on automatic recognition of spontaneous micro-expressions.

Local Binary Patterns (LBP) is widely used in facial expression recognition [8] due to its ability to derive local statistical patterns that exhibit invariance towards illumination changes and simplicity in computation. In order to cope with dynamic textures and events across spatio-temporal dimensions, the classic LBP descriptor was extended to a volume-based LBP (VLBP) and LBP from three orthogonal planes (LBP-TOP) [14].

Among all the available work, Yan et al. [12] reported a baseline performance of up to 63.41% accuracy for a 5-class classification task on their own CASMEII database, adopting LBP-TOP and SVM for feature extraction and classification respectively. The CASMEII has since superseded the original CASME database with the inclusion of more subjects and a higher sampling rate that is able to capture detailed facial muscle movements. While the LBP-TOP is an effective descriptor for dynamic textures, there are redundant pattern information within the overlapping orthogonal planes. This redundancy contributes to an increase in computational complexity, and also intuitively results in a less discriminative set of features.

In this paper, we propose Local Binary Pattern with Six Intersection Points (LBP-SIP), a computationally lightweight descriptor based on the LBP-TOP. In LBP-SIP, the unique distinct points that lie on the three intersecting lines of the three orthogonal planes are considered for computing the spatio-temporal patterns. The proposed descriptor is then incorporated in a multi-resolution Gaussian pyramid by concatenating the feature histograms of all four pyramid levels. Our proposed method is able to consistently match or outperform LBP-TOP in various aspects in addition to the computational efficiency it brings. Using SVM classifier with leave-one-sample-out cross validation, we obtain the best recognition accuracy of 67.21%.

The rest of the paper is organized as follows: Sect. 2 reviews some recent methods in this area of research. Our proposed approach is presented in Sect. 4. Then we show our detailed experimental results in Sect. 5. Finally, the conclusion is given in Sect. 6.

2 Related Work

In dealing with dynamic textures that evolve over space-time dimensions, the LBP-TOP remains a popular choice of feature extraction for various applications such as texture recognition [13], face spoofing [1], gait and action recognition [3, 5], and facial expression recognition [14, 15].

Following its conception, several works were proposed to improve upon its effectiveness and robustness. Shan and Gritti [7] proposed to learn discriminative

LBP-Histogram bins which are able to provide a more compact yet discriminative representation for facial expressions. In the similar vein, Zhao et al. [15] extended the usage of LBP-TOP to multi-resolution space while utilizing AdaBoost to learn and select the most prominent expression-related features from different blocks and slices. To increase its robustness against view-based variations in texture, rotation-invariant descriptors [16] computed from the LBP-TOP features were proposed, to a good measure of success. Interestingly, the use of a multi-resolution pyramid of LBPs [6] was also found to be beneficial in extracting dominant structures in textures.

A majority of these methods are tailored towards texture recognition and *macro-expression* recognition, while very scant attention is given to address the challenging task of recognizing subtle facial *micro-expressions*. The use of LBP-TOP for feature description and SVM as classifier provides the baseline performance for the recently-proposed SMIC [4] and CASMEII [12] datasets, the latter obtaining a good accuracy rate of up to 63.41%. As pointed out in [11], research in micro-expression recognition is still at an early stage with very few reported works to date. In a recent method, Wang et al. [9] proposed an efficient technique that uses discriminant tensor subspace for feature extraction and extreme learning machine (ELM) for classification. Experiments on the CASME micro-expression dataset showed some promising results (46.9%, up from baseline of 41% reported in [10]), especially when high-order tensors are applied.

In this paper, we improve the baseline performance in CASMEII by introducing an efficient and robust descriptor that trims the excess redundancy in feature patterns, resulting in a more compact and well-formed representation.

3 LBP-Three Orthogonal Planes (LBP-TOP)

In this section, we describe the key idea of the LBP-TOP descriptor. Given a pixel c located at (x_c, y_c) , its LBP code is computed as:

$$LBP_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} s(i_p - i_c) \times 2^p \quad (1)$$

where

$$s(x) = \begin{cases} 1 & \text{if } x \geq 0, \\ 0 & \text{if } x < 0, \end{cases} \quad (2)$$

i_c denotes the intensity of the central pixel c , P denotes the total number of neighbours of c parametrized by the radius of the neighbourhood R , while i_p indexed by p denotes the intensity of the neighbouring pixels. Then, the histogram of all r LBP patterns, $H_{LBP}(r)$ is computed for all pixels in an image, describing the LBP texture features of that image.

LBP-TOP computes the local spatio-temporal patterns based on LBP. More precisely, the LBP-TOP feature is constructed by the concatenation of LBP histograms on three orthogonal planes - XY, XT and YT respectively as shown in Fig. 1(a). The XT and YT planes contain the temporal transition information

pertaining to the facial movement displacement e.g. how eyes, lips, muscle or eyebrows change over time. They are the stack of columns and rows of pixels respectively. In contrast, the XY plane (a frame itself) contains only spatial information which includes both expression and identity information of a face appearance.

However, by deeper inspection of the LBP-TOP formulation, we observe that not all the neighbour points on three orthogonal planes respectively used to compute the LBP code (feature pattern) are distinctively different. In fact, when all three planes are considered in totality, some points are used more than once in the LBP-TOP computation of the center pixel, thus leading to redundant differencing and thresholding computations when the LBP codes are computed. Therefore, to compute more compactly while preserving the essential pattern information, we propose to uniquely compute the spatial and temporal two groups of neighbour points only in order to obtain the spatio-temporal LBP patterns. More precisely, we only consider the six distinct neighbour points on the three intersecting lines formed by the three orthogonal planes as shown in Fig. 1(b). The details of the proposed approach are elaborated in the next section.

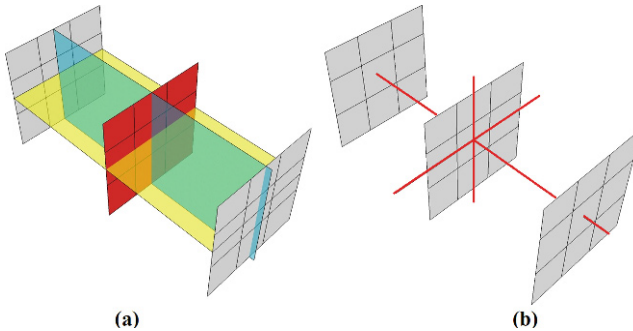


Fig. 1. (a) Three orthogonal planes in the LBP-TOP computation, and (b) Three intersecting lines crossing over the center pixel, formed by the three orthogonal planes, are considered to obtain the six distinct neighbour points surrounding the center point.

4 Proposed LBP-Six Intersection Points (LBP-SIP)

To further extend the LBP-TOP, we propose a more compact and efficient form while preserving much of robustness and the essential information that describes the dynamic textures. By closer examination of the neighbour points on all three orthogonal planes, the key property lies in the uniqueness of these neighbours.

Given a center pixel c in spatial location $(x_{c,t}, y_{c,t})$ at time t , its LBP-TOP feature can be denoted as $LBP-TOP_{P_{XY}, P_{XT}, P_{YT}, R_X, R_Y, R_T}(x_{c,t}, y_{c,t})$ with six parameters; the P parameters denote the number of neighbours in each of the planes, while the R parameters denote the radii in each of the axes. Three LBP

histograms, one for each of the planes (XY, XY, YT) are concatenated to form the final LBP-TOP feature histogram, i.e. $H_{LBP-TOP} = H_{LBP,\pi}(\pi = XY, XT, YT)$.

In detail, the LBP-TOP neighbours of the pixel $(x_{c,t}, y_{c,t})$ are given as (without loss of generality, we consider only 4-neighbours for the proposal of this method):

- XY plane:
 $(x_{c,t}, y_{c,t} + R_Y), (x_{c,t} + R_X, y_{c,t}), (x_{c,t}, y_{c,t} - R_Y), (x_{c,t} - R_X, y_{c,t})$
- XT plane:
 $(x_{c,t} + R_X, y_{c,t}), (x_{c,(t+R_T)}, y_{c,(t+R_T)}), (x_{c,t} - R_X, y_{c,t}), (x_{c,(t-R_T)}, y_{c,(t-R_T)})$
- YT plane:
 $(x_{c,t}, y_{c,t} + R_Y), (x_{c,(t+R_T)}, y_{c,(t+R_T)}), (x_{c,t}, y_{c,t} - R_Y), (x_{c,(t-R_T)}, y_{c,(t-R_T)})$

Note that each neighbour pixel is used more than once in the computation of LBP-TOP, and that there are uniquely only 6 distinct neighbour points within the set of intersecting planes.

Building on this, we propose an LBP descriptor with reduced set of spatio-temporal neighbourhood points derived from the intersection of the three orthogonal planes. To better describe the method, consider the example in Fig. 2. In the original LBP-TOP computation, we compute for the central pixel C lying on the XY plane (in the middle frame) with 4 neighbour points considered on each plane that are the 4-neighbour points set $\{D, E, F, G\}$ for XY plane (middle frame), $\{E, A, G, B\}$ for XT plane (red plane), and $\{D, A, F, B\}$ for YT plane (blue plane). Observing that from these three point sets, every point is used twice to compute the resulting pattern.

Therefore, we propose to reduce the computational complexity by discarding the redundant *intersection points*. From Fig. 2, we can clearly see that the three orthogonal planes produce three intersecting lines (AB, DF, EG), all crossing over the center point C . Regardless of the radius between the center point and the original neighbour points, there are only six unique neighbour points on the intersection lines surrounding the center point—the six intersection points. Concisely from this example,

$$XY \cap XT \cap YT = \{A, B, D, E, F, G\} \quad (3)$$

Intuitively, these 6 unique neighbour points carry enough information to describe the spatio-temporal textures that center upon point C . Geometrically, we can view the new neighbour point set in two groups representing both spatial and temporal texture information. Firstly, we regard points $\{D, E, G, F\}$ as the spatial neighbour set while the two end points $\{A, B\}$ along the temporal axis (that is on the intersection line of the XT and YT planes) make up the temporal neighbour set. As such, the final feature histogram consists of two concatenated histograms of length $2^4 + 2^2 = 20$.

In contrast, the three orthogonal planes of the LBP-TOP produce a concatenated feature histogram of length $2^4 \times 3 = 48$, more than two times that

of the proposed LBP-SIP. In terms of computational complexity, this results in a much compact feature length (or dimension). This is much desirable as high-dimensional feature spaces often suffer from the curse of dimensionality whereby the represented data becomes increasingly sparse, affecting classification ability.

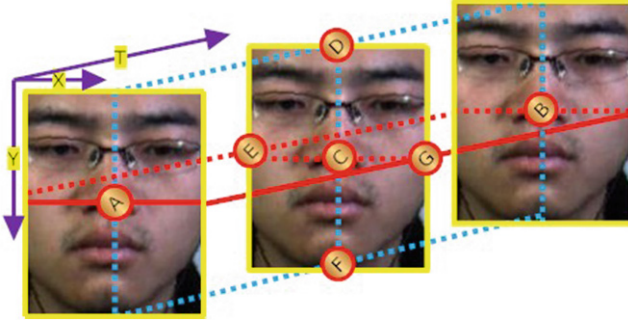


Fig. 2. The three orthogonal planes (XY, XT, YT) shown in different colors, and the intersecting points that are the neighbour points of point C shared by all three planes (A, B, D, E, F, G) (Color figure online).

Consistent with the previously described methods, we formally denote the proposed descriptor as: $LBP - SIP_{P_{XY}, R_X, R_Y, R_T}(x_{c,t}, y_{c,t})$ where P_{XY} is fixed to 4, leaving only the radii parameters free. Good values for R_X , R_Y and R_T have been reported in [12] by empirical means.

5 Experiments

In this section, we present our experimental results. Firstly, We describe the dataset we used for our experiments in Sect. 5.1. In Sect. 5.3, we compare the performance of the proposed LBP-SIP with LBP-TOP through a multi-resolution Gaussian pyramid to examine the robustness of the methods across scale. In Sect. 5.4, we further demonstrate the robustness of the intuitive spatial and temporal neighbour point grouping exemplified in our proposed method. Finally, the complexity of both LBP-TOP and LBP-SIP methods are analyzed in Sect. 5.5.

5.1 Dataset

Experiments were conducted on two recently-proposed datasets—SMIC [4] and CASMEII [12]. We intensively test our proposed method on the CASMEII dataset, since it is the more comprehensive dataset between the two. In addition, we also substantiate our proposed idea by testing on the SMIC dataset for the basic case without using Gaussian pyramids (see Sect. 5.2).

SMIC [4] consists of both micro and macro expression videos. In this paper, we focus on the micro-expression only. A high speed (HS) camera (PixelINK

PL-B774U, 640×480) of 100 fps was used to record the short duration of micro-expressions. 20 participants (164 videos) participated in the recording experiment. Only 3 micro-expression classes (positive, surprise and negative) are provided.

CASMEII [12] is the most extensive spontaneous micro-expression dataset to date, and it is publicly made available by the Chinese Academy of Science (CAS). Due to the lack of samples in some expression classes, the CAS team suggested a baseline experimental setup of 5 expression classes (Happiness, Disgust, Surprise, Repression, Tense) with a total of 247 different spontaneous micro-expression videos used for experiment. The micro-expression samples were recorded with high-speed camera (at 200 fps) from 26 participants with higher face resolution of around 280×340 pixels (original resolution). These samples were selected from nearly 3,000 elicited facial movements with their onset and offset frames coded. The Action Units (AU) and emotions are also properly marked and labelled according to the FACS coding system. The selection procedure was implemented as some samples are too subtle to be coded or labelled by the naked eye. This enforces the nature of micro-expressions and the obvious difficulties in creating a micro-expression database.

In our experiments, we strive to achieve consistency with the baseline work [12]. The smaller version of the cropped faces are used without frame size (X-Y dimension) normalization and video length (T dimension) normalization. This is possible as the descriptors tested (LBP-TOP, LBP-SIP) can accommodate different spatial and temporal scales. We also employ the same number of classes used. The details of the experimental results are shown in the following subsections.

5.2 Baseline Comparison: LBP-SIP vs LBP-TOP

We test our proposed LBP-SIP against LBP-TOP on both CASMEII and SMIC databases. We use 5×5 block partition and set the radii to $\{R_X, R_Y, R_T\} = \{1, 1, 4\}$, corresponding to the best results in [12]. Table 1 shows the results. From the results shown in Table 1, the LBP-SIP is clearly superior on all accounts.

On an Intel Core i7 machine with 8GB RAM, the average feature extraction time per video on CASMEII dataset for LBP-TOP is 18.289s, while ours took 15.888s. The recognition time for LBP-TOP is 0.584s per video while ours took 0.208s per video (an improvement of ≈ 2.8 times).

Table 1. Comparison of LBP-TOP and LBP-SIP on the CASMEII and SMIC datasets using SVM and RBF kernels for SVM

	CASMEII		SMIC	
	LBP-TOP (%)	LBP-SIP (%)	LBP-TOP (%)	LBP-SIP (%)
Linear	62.75	63.56	60.98	64.02
RBF	65.99	66.40	60.98	62.80

5.3 LBP-SIP vs LBP-TOP on a Gaussian Pyramid

From previous works, Zhao et al. [14] conducted intensive experiments on LBP-TOP for facial expression recognition, recommending that the neighbourhood radii takes the values $\{R_X, R_Y, R_T\} = \{1, 1, 2\}$. Meanwhile, Yan et al. [12] empirically showed that the best values applied to facial micro-expression recognition are $\{R_X, R_Y, R_T\} = \{1, 1, 4\}$ though the result is not significantly better than with $\{1, 1, 2\}$. Hence, for ease of comparison with the best known works, we consider both settings in this experiment.

We use a Gaussian pyramid to downsample every single image frame into 4 levels, where level 0 denotes the original size of image. Let the original frame resolution be $w \times h$ (width and height), and pyramid level be l which ranges from 0 to 3 for our case (as shown in Fig. 3). Applying Gaussian low pass filtering, which is a smoothing process at each level l results in a frame resolution of $(w \times h)/2^l$. In other words, the image size at each level will be half that of the previous level.

To better visualize the effect on different levels of the multi-resolution Gaussian pyramid, we normalize the processed images to the size of 163×134 , as shown in the second row of Fig. 3. The third row shows the LBP coding at different resolution levels of the pyramid.

We use 5×5 blocks partition to compute the LBP-TOP and LBP-SIP feature histograms. Due to different pre-processing applied (e.g. image sequence normalization), our accuracy rate for LBP-TOP appears to be slightly different from the reported baseline [12]. As such, we maintain the same pre-processing for all our experiments on both LBP-TOP (as our baseline) and LBP-SIP to ensure comparisons can be made under fair conditions. We then follow through the rest of the recognition process using the SVM classifier with leave-one-sample-out cross validation.

Tables 2 and 3 show the performance of LBP-TOP and LBP-SIP using features derived from the different levels of the Gaussian pyramid as well as the concatenated features of all levels, with the temporal radius R_T of 2 and 4 respectively. In Table 2, in the linear case, we can see that LBP-TOP outperforms LBP-SIP at some levels, while at some other levels LBP-SIP slightly outperforms LBP-TOP. There is little difference between the two. On the other hand, LBP-SIP always outperforms LBP-TOP when the nonlinear RBF kernel is applied.

The performance results shown in Table 3 where $R_T = 4$ can be better visualized in Figs. 4(a) and (b), showing the linear and RBF kernel respectively. In Fig. 4(a), it is obvious that the performance of LBP-SIP (marked with triangular points) on the Linear kernel is almost consistently above or superposing the LBP-TOP line (marked with circular points) through all individual pyramid levels and the concatenated levels.

Overall, it may seemed that there is no significant difference between LBP-TOP and LBP-SIP in terms of accuracy (except when the RBF kernel is used), but the increase in computational efficiency and robustness in high dimensionality (concatenated feature) are promising advantages for practical purposes.

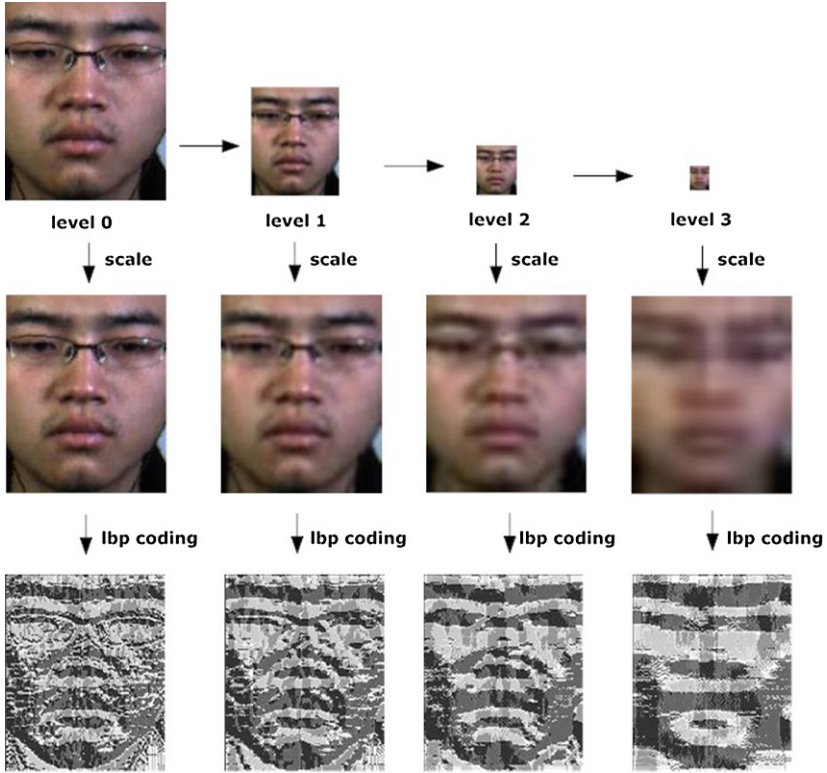


Fig. 3. LBP coding at different levels of a Gaussian pyramid

5.4 Spatial and Temporal Grouping of Neighbour Points

We further demonstrate the intuitiveness of considering a group of 4 intersecting points as the spatial LBP pattern on XY plane and the 2 remaining intersecting points as the temporal LBP pattern ($XY+2$), as compared to the two other possible grouping permutations that can arise from the 6 unique neighbour points. Two other combination of groupings are tested out: spatial grouping on XT plane and two points along the Y dimension ($XT+2$), and spatial grouping on YT plane and two points along the X dimension ($YT+2$). An illustration of how the six unique neighbour points can be grouped is shown in Fig. 5. For simplicity, we compare the recognition performance of the different neighbour groupings with $R_T = 4$ (since it was shown earlier to be better temporal radius).

Table 4 clearly shows that irrespective of the choice of kernels used, the $XY+2$ grouping outperforms the $XT+2$ and $YT+2$ groupings in most of the evaluated cases. We observe that the spatial and temporal grouping of neighbour points on the $XY+2$ setting is more robust across different levels of the Gaussian pyramid. The $XY+2$ grouping is also the most intuitive considering that the spatial

Table 2. Comparison of LBP-TOP and LBP-SIP on different level of Gaussian pyramid with $R_X = 1, R_Y = 1, R_T = 2$

	Linear (%)		RBF (%)	
	LBP-TOP	LBP-SIP	LBP-TOP	LBP-SIP
Level0	57.89	61.13	59.11	63.16
Level1	62.75	61.54	62.75	65.59
Level2	60.73	57.09	58.70	59.51
Level3	57.89	57.49	57.09	58.30
All levels (concatenated)	65.59	63.97	64.78	64.78

Table 3. Comparison of LBP-TOP and LBP-SIP on different level of Gaussian pyramid with $R_X = 1, R_Y = 1, R_T = 4$

	Linear (%)		RBF (%)	
	LBP-TOP	LBP-SIP	LBP-TOP	LBP-SIP
Level0	60.73	62.75	65.99	65.18
Level1	61.94	62.75	62.75	65.59
Level2	65.99	65.18	66.40	67.21
Level3	63.16	63.16	64.78	62.75
All levels (concatenated)	66.80	67.21	67.61	67.21

Table 4. Comparison of LBP-SIP for different groupings of neighbour points on different level of Gaussian pyramid with $R_X = 1, R_Y = 1, R_T = 4$

	Linear (%)			RBF (%)		
	XY+2	XT+2	YT+2	XY+2	XT+2	YT+2
Level0	62.75	61.94	60.73	65.18	61.13	63.56
Level1	62.75	60.73	61.13	65.59	63.97	64.37
Level2	65.78	57.49	67.20	67.21	62.75	66.40
Level3	63.16	62.35	57.89	62.75	61.94	61.13
All levels (concatenated)	67.21	65.59	66.40	67.21	66.40	65.18

pattern that resides on the XY plane is akin to a classic LBP pattern while the temporal pattern straddles across the T axis.

5.5 Complexity of LBP-TOP vs LBP-SIP

LBP-TOP. For each video sample, the use of 4 neighbour points on each of the 3 orthogonal planes result in $4 \times 3 \times w \times h \times l$ number of computations where $w \times h$ is the spatial resolution of the image (frame) and l is the length of the video

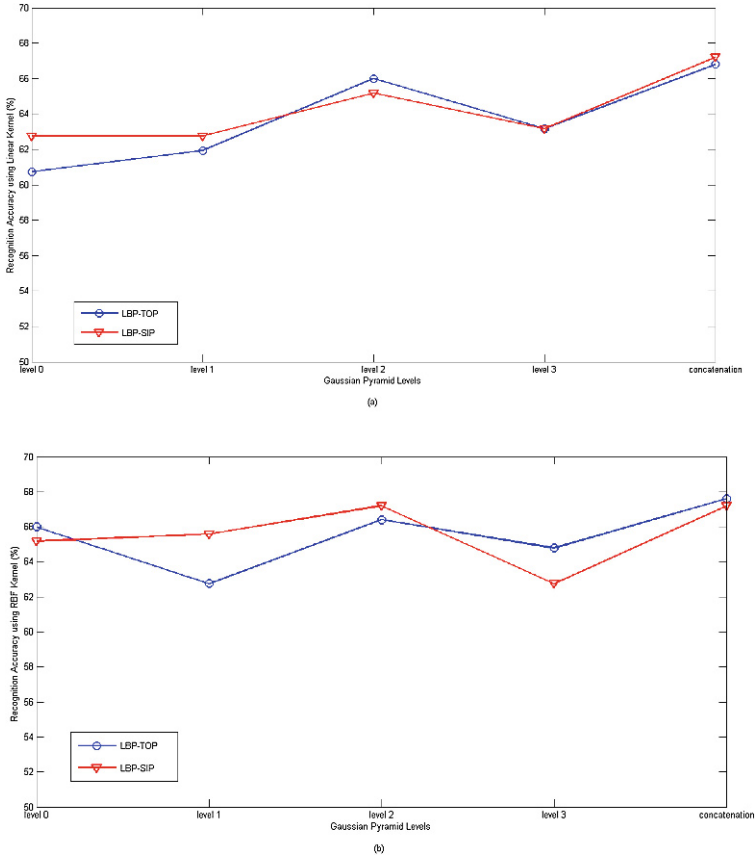


Fig. 4. The comparison between LBP-TOP and LBP-SIP using SVM linear in (a) and RBF kernel in (b) with $R_T = 4$

(i.e. number of frames). The length of the entire concatenated feature histogram is $2^4 \times 3 \times (5 \times 5) = 1200$ dimensions, where $2^4 \times 3$ is the dimensionality of feature in a single block partition with three orthogonal planes while 5×5 gives the number of block partitions applied (to the XY plane).

LBP-SIP. In the proposed LBP-SIP approach, there are only 6 unique points derived from the three intersecting lines formed by the three orthogonal planes. We separate these neighbour points into two groups; namely a spatial LBP group that consists of 4 points along the spatial XY plane, and a temporal LBP group containing the remaining 2 points along the T axis. This results in $(4 + 2) \times w \times h \times l$ computations, which is half the number of computations required for the LBP-TOP. Furthermore, the dimensionality is reduced by 2.4 times from the LBP-TOP approach to a compact size of $(2^4 + 2^2) \times (5 \times 5) = 500$ dimensions, which is sufficient to represent the essential spatio-temporal patterns while maintaining a competitive performance.

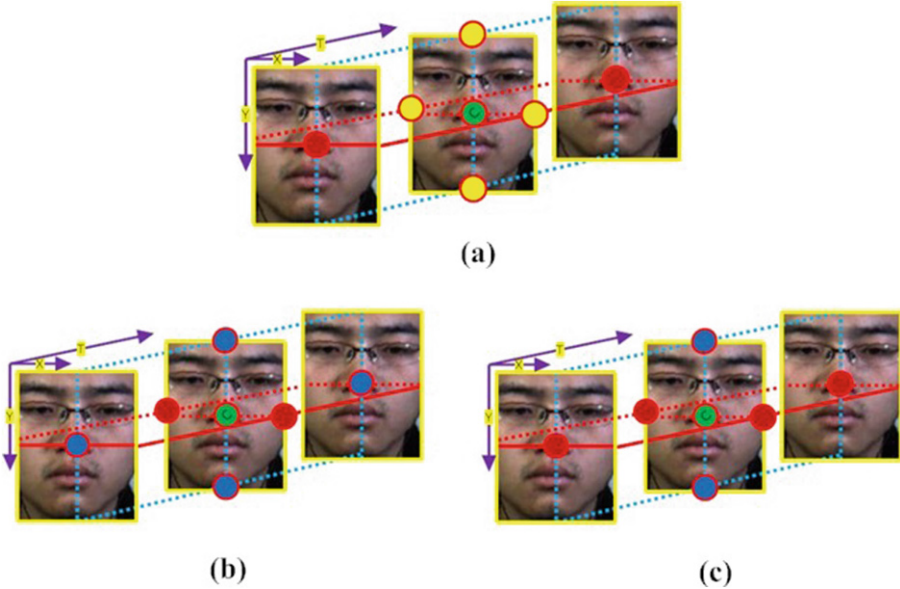


Fig. 5. The various histogram groupings of neighbours surrounding the center point c , denoted in green color: (a) $XY+2$ grouping, where the four yellow colored neighbour points in the XY plane are grouped as spatial neighbour points to correspond to a histogram, and the two red colored points are grouped as the temporal neighbour points to correspond to a second histogram (b) $YT+2$ where the four blue colored points in YT plane are one group of neighbour points and the two red color are another group of neighbour points (c) $XT+2$ where the four red colored points in XT are one group of neighbour points and the two blue color are another group of neighbour points (Color figure online)

6 Conclusion

In this paper, we propose LBP-SIP as a more compact and efficient formulation over LBP-TOP. Instead of considering all three orthogonal planes which contain redundant points, we propose a reduced set of unique spatio-temporal neighbour points derived from the intersecting lines of the three orthogonal planes. LBP-SIP is then computed through a multi-resolution Gaussian pyramid by concatenating feature patterns from every level to improve on the task of facial micro-expression recognition. In various experiments conducted on the recently proposed CASMEII database, the LBP-SIP has consistently matched or outperformed the LBP-TOP in accuracy terms while exhibiting stability and robustness in high feature dimensionality. Most noteworthy here is its computational efficiency and a clear reduction in the length of feature histogram without deterioration in performance.

Acknowledgement. We thank the Chinese Academy of Sciences for access to the CASMEII micro-expression database and Su-Jing Wang for providing more details on

their CASMEII work [12]. We also thank the anonymous reviewers for their constructive comments. This research work is funded by the TM Grant under project UbeAware.

References

1. de Freitas Pereira, T., Anjos, A., De Martino, J.M., Marcel, S.: *LBP – TOP* based countermeasure against face spoofing attacks. In: Park, J.-I., Kim, J. (eds.) *ACCV Workshops 2012, Part I. LNCS*, vol. 7728, pp. 121–132. Springer, Heidelberg (2013)
2. Ekman, P.: Lie catching and microexpressions. In: Martin, C.W. (ed.) *The Philosophy of Deception*, pp. 118–133. Oxford University Press, Oxford (2009)
3. Kellokumpu, V., Zhao, G., Li, S.Z., Pietikäinen, M.: Dynamic texture based gait recognition. In: Tistarelli, M., Nixon, M.S. (eds.) *ICB 2009. LNCS*, vol. 5558, pp. 1000–1009. Springer, Heidelberg (2009)
4. Li, X., Pfister, T., Huang, X., Zhao, G., Pietikainen, M.: A spontaneous microexpression database: inducement, collection and baseline. In: *10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, pp. 1–6 (2013)
5. Mattivi, R., Shao, L.: Human action recognition using LBP-TOP as sparse spatio-temporal feature descriptor. In: Jiang, X., Petkov, N. (eds.) *CAIP 2009. LNCS*, vol. 5702, pp. 740–747. Springer, Heidelberg (2009)
6. Qian, X., Hua, X.S., Chen, P., Ke, L.: PLBP: an effective local binary patterns texture descriptor with pyramid representation. *Pattern Recogn.* **44**(10), 2502–2515 (2011)
7. Shan, C., Gritti, T.: Learning discriminative LBP-histogram bins for facial expression recognition. In: *BMVC*, pp. 1–10 (2008)
8. Shan, C., Gong, S., McOwan, P.W.: Facial expression recognition based on local binary patterns: a comprehensive study. *Image Vis. Comput.* **27**(6), 803–816 (2009)
9. Wang, S.J., Chen, H.L., Yan, W.J., Chen, Y.H., Fu, X.: Face recognition and micro-expression recognition based on discriminant tensor subspace analysis plus extreme learning machine. *Neural Process. Lett.* **39**(1), 25–43 (2014)
10. Yan, W.J., Wu, Q., Liu, Y.J., Wang, S.J., Fu, X.: CASME database: a dataset of spontaneous micro-expressions collected from neutralized faces. In: *10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, pp. 1–7 (2013)
11. Yan, W.J., Wang, S.J., Liu, Y.J., Wu, Q., Fu, X.: For micro-expression recognition: database and suggestions. *Neurocomputing* **136**, 82–87 (2014)
12. Yan, W.J., Li, X., Wang, S.J., Zhao, G., Liu, Y.J., Chen, Y.H., Fu, X.: CASME II: an improved spontaneous micro-expression database and the baseline evaluation. *PloS One* **9**(1), e86041 (2014)
13. Zhao, G., Pietikainen, M.: Local binary pattern descriptors for dynamic texture recognition. In: *18th International Conference on Pattern Recognition (ICPR)*, vol. 2, pp. 211–214. IEEE (2006)
14. Zhao, G., Pietikainen, M.: Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(6), 915–928 (2007)
15. Zhao, G., Pietikainen, M.: Boosted multi-resolution spatiotemporal descriptors for facial expression recognition. *Pattern Recogn. Lett.* **30**(12), 1117–1127 (2009)
16. Zhao, G., Ahonen, T., Matas, J., Pietikainen, M.: Rotation-invariant image and video description with local binary pattern features. *IEEE Trans. Image Process.* **21**(4), 1465–1477 (2012)