Subtle Expression Recognition using Optical Strain Weighted Features

Sze-Teng Liong[1], John See[2], Raphael C.-W. Phan[3], Anh Cat Le Ngo[3], Yee-Hui Oh[3], KokSheik Wong[1]

[1] Faculty of Computer Science & Information Technology, University of Malaya, Kuala Lumpur, Malaysia
szeteng1206@hotmail.com
[2] Faculty of Computing & Informatics, Multimedia University, Cyberjaya, Malaysia
johnsee@mmu.edu.my
[3] Faculty of Engineering, Multimedia University, Cyberjaya, Malaysia
raphael@mmu.edu.my, lengoanhcat@gmail.com, yeehui716@gmail.com

Abstract. Optical strain characterizes the relative amount of displacement by a moving object within a time interval. Its ability to compute any small muscular movements on faces can be advantageous to subtle expression research. This paper proposes a novel optical strain weighted feature extraction scheme for subtle facial micro-expression recognition. Motion information is derived from optical strain magnitudes, which is then pooled spatio-temporally to obtain block-wise weights for the spatial image plane. By simple product with the weights, the resulting feature histograms are intuitively scaled to accommodate the importance of block regions. Experiments conducted on two recent spontaneous micro-expression databases– CASMEII and SMIC, demonstrated promising improvement over the baseline results.

1 Introduction

Facial based emotion recognition attracts research attention both in the computer vision and psychology community. Six basic facial expressions which are commonly considered are happy, surprise, anger, sad, fear and disgust [1]. Contributing to this interest in emotion recognition is the increased research into affective computing, i.e. the ability for software and machines to react to human emotions as they are performing their tasks.

Facial *micro-expressions* were discovered by Ekman [2] in 1969 when he analyzed the interview video of a patient stricken with depression who tried to commit suicide. According to Ekman, micro-expressions cannot be controlled by humans and they are able to reveal concealed emotions. Micro-expressions occur at a high speed (within one twenty-fifth to one fifth of a second) and they are usually involuntary facial expressions [3]. The fact that they occur in a short duration and potentially in only one part of the face makes it hard to detect them with naked eye in real-time conversations.

There are various applications that support why micro-expressions are important to be analysed, such as clinical diagnosis, national security and interrogation [4–6]. To date, detection of micro-expressions is still a great challenge to researchers in the field of computer vision due to its extremely short duration and low intensity.

Optical strain is the relative amount of deformation of an object [7]. It is able to calculate any small changes on the facial expression, including small muscular movements on the face. In this paper, we propose a new optical strain weighting scheme that utilizes the block-based optical strain magnitudes to extract weighted spatio-temporal features for subtle micro-expression recognition. Firstly, the optical strain map images are computed and normalized from the optical strain magnitudes. Then, the spatial plane (XY) is partitioned into $N \times N$ non-overlapping blocks, where spatio-temporal pooling is applied to obtain a single magnitude for each block. The histograms obtained from the feature extractor are then multiplied with the optical strain weights to form the final feature histogram.

2 Related Work

Optical strain patterns justify its superiorty over the raw image in face recognition as the computation of the magnitudes is based on biomechanics. It is also robust to the lighting condition, heavy make up and under camouflage [8, 9]. In [10], Shreve et al. used the optical strain technique to automatically spot macro- and micro- expressions on facial samples. They could achieve 100% accuracy in detecting seven micro-expressions in the USF dataset. However the micro-expressions in the database are posed expression rather than spontaneous ones.

Two years later, [11] an extensive testing was carried on two larger datasets (Canal-9 [12] and found videos [13]), containing a total of 124 micro-expressions by implementing a modified algorithm to spot the micro-expressions. To overcome the noises caused by irrelevant movements on the face, some parts of the face were masked. The face was partitioned into eight regions for the optical strain magnitude to be calculated locally. They extended the work by modifying the algorithms [14]. However, they mentioned that the background and some parts of the face should be masked to avoid the inaccurate optical flow values affect the spotting accuracy.

Block-based method in feature extraction process is widely used in detecting or recognizing micro-expressions, as demonstrated in [15–18]. The face image is partitioned into multiple $N \times N$ non-overlapping or overlapping blocks. The Local Binary Pattern with Three Orthogonal Planes (LBP-TOP) histograms in each block are computed and concatenated into a single histogram. By doing so, the local information of facial expression at its spatial location are taken into account.

Pooling is a method to decrease the number of features (lower dimension) in image recognition. If all the features are extracted, it may result in overfitting.

Spatial pooling summarizes the values in the neighbouring locations to achieve better robustness to noise [19]. In [20], Philippe. et. al. demonstrated several combinations of temporal pooling over a time period and it has been proven to improve the performance of automatic annotation and ranking music audio.

Gaussian filter is one of the effective and adaptive filters to remove Gaussian noises on an image [21]. To track the action units (AUs) on facial expressions using Facial Action Coding System (FACS) [22], a 5 x 5 Gaussian filter is applied to smooth the images and different sizes of gradient filter are used on different regions of face [23]. In [24], an adaptive Gaussian filter is used to reduce the noises on images in order to compute the illumination change of one person or Expression Ratio Image (ERI) resulted from deformation of the person's face.

To analyze the micro-expressions through a recognition system, it is necessary to have a database to act as a test data set for the researchers to be able to compare the results. There are plenty of facial expression databases available for evaluation [25]. However, there are only a few well-established databases for micro-expressions. This brings to an even bigger obstacle in classifiying the micro-expressions and training the detection algorithms. For example, the micro expressions are posed rather than spontaneous in USF-HD [26] and Polikovsky's database [27]. On the other hand, there are insufficient videos in YorkDDT [28] and SMIC [17] databases.

3 Motion and Feature Extraction

3.1 Optical Flow

Optical flow specifies the velocity of each image pixel's movement between adjacent frames [29]. Computation of differential optical flow is by measuring the spatio and temporal changes of intensity to find a matching pixel in the next frames [30]. As this estimation method is highly sensitive to any changes in brightness, hence it is assumed that all temporal intensity changes are due to motion only. There are threee assumptions to measure the optical flow. First, brightness constancy, where the brightness intensity of moving objects between two image frames are assumed to remain constant. Second, spatial coherence, where the pixels in a small image window are assumed to be originating from the same surface and having similar velocity. Third, temporal persistence, where it assumes objects changes gradually over time. The optical flow gradient equation is often expressed as:

$$\nabla I \bullet \boldsymbol{p} + I_t = 0, \tag{1}$$

where *I(x,y,t)* is the image intensity function at point (x, y) at time t. $\nabla I =$ *(I_x, I_y)* is the spatial gradients and I_t denotes the temporal gradient of the intensity function. $\boldsymbol{p} = [p = d_x/d_t, q = d_y/d_t]^T$ represents the horizontal and vertical motion vector.

3.2 Optical Strain

Using optical strain in identifying deformable results performs better than optical flow [31] as it can well distinguish the time interval of the occuring of micro-expressions. A deformable object can be described in two dimensional space by using a displacement vector $\mathbf{u} = [u, v]^T$. Assuming that the moving object is in small motion, the finite strain tensor can be represented as:

$$\varepsilon = \frac{1}{2}[\nabla \mathbf{u} + (\nabla \mathbf{u})^T] \tag{2}$$

or in an expanded form:

$$\varepsilon = \begin{bmatrix} \varepsilon_{xx} = \frac{\partial u}{\partial x} & \varepsilon_{xy} = \frac{1}{2}(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}) \\ \varepsilon_{yx} = \frac{1}{2}(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}) & \varepsilon_{yy} = \frac{\partial v}{\partial y} \end{bmatrix} \tag{3}$$

where $(\varepsilon_{xx}, \varepsilon_{yy})$ are normal strain components and $(\varepsilon_{xy}, \varepsilon_{yx})$ are shear strain components.

The magnitude of the optical strain can be computed as follows:

$$\varepsilon = \sqrt{{\varepsilon_{xx}}^2 + {\varepsilon_{yy}}^2 + {\varepsilon_{xy}}^2 + {\varepsilon_{yx}}^2} \tag{4}$$

An *optical strain map* (OSM) provides a visual representation of the motion intensity for each pixel in a video frame. To visualize the OSM, the optical strain magnitudes for each point (x, y) in image space at time t can be normalized to intensity values 0-255. By observing the OSM, we can clearly notice regions in the image frame that contain the most prominent (large values) or least prominent (small values) motion in terms of spatial displacement. To obtain a summed OSM for the entire sequence, all the individual generated OSMs can be summed across the temporal dimension. This accumulates all motion displacements in the whole sequence, a pooling operation that will be discussed later in Subsection 4.2. Fig. 1 shows a sample optical strain map image (for two adjacent frames), and a summed optical strain map image (for all frames, temporal sum pooled) after applying intensity normalization.

3.3 Block-based LBP-TOP

Block-based LBP-TOP is implemented by partitioning each frame of the video into $N \times N$ non-overlapping blocks then concatenate them into a single histogram. Fig. 2 shows the process of extracting the features from three orthogonal plane for one block volume and concatenate them into a histogram. The feature histogram of block-based LBP-TOP [15] can be defined as follows:

$$H_{i,j,c,b} = \sum_{x,y,t} I f_c(x, y, t) = b,\ b = 0, \ldots, n_c - 1;\ c = 0, 1, 2;\ i, j \in 1 \ldots N \tag{5}$$

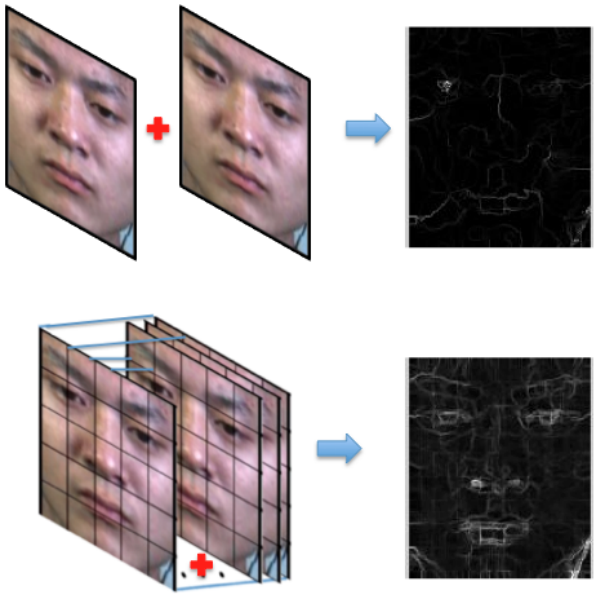

Fig. 1. Example of optical strain map for two image frames *(top row)* and for all the frames in sequence *(bottom row)* for a tense micro-expression

where n_c is the number of different labels produced by the LBP operator in the cth plane ($c = 0 : XY, 1 : XT$ and $2 : YT$), $f_c(x, y, t)$ is the LBP code of the central pixel (x, y, t) in c-th plane, $x \in \{0, \ldots, X-1\}, y \in \{0, \ldots, Y-1\}, t \in \{0, \ldots, T-1\}$,

$$I\{A\} = \begin{cases} 1, & \text{if } A \text{ is true;} \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

The histogram is normalized to get a coherent description:

$$\overline{H}_{i,j,c,b} = \frac{H_{i,j,c,b}}{\sum_{k=0}^{n_c-1} H_{i,j,c,k}} \tag{7}$$

We denote LBP-TOP parameters by $LBP\text{-}TOP_{P_XY,P_XY,P_YT,R_X,R_Y,R_T}$ where the P parameters indicate the number of neighbor points for each of the three orthogonal planes, while the R parameters denote the radii along the X, Y, and T dimensions of the descriptor.

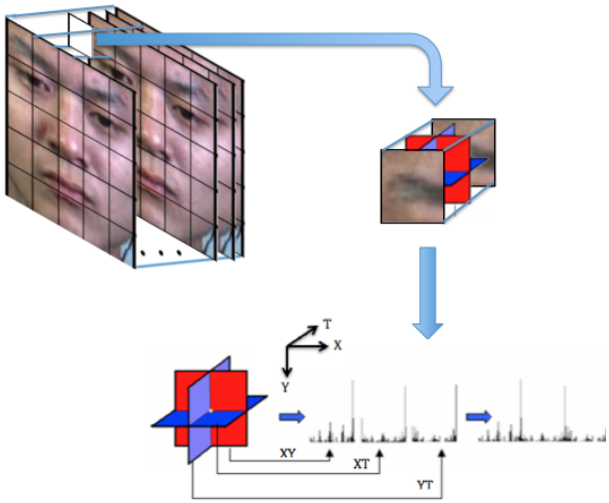

Fig. 2. Block-based LBP-TOP: Feature extraction from three orthogonal planes for one block volume

4 Proposed Algorithm

4.1 Block-wise Optical Strain Magnitudes

The magnitude of optical strain for each pixel is very small. Much of the surrounding pixels that contain very little flow corresponds to very minute values. As such, we hypothesize that using the optical strain map magnitudes directly as features or by extraction of LBP patterns for classification may result in a loss of essential information from its original image intensity values.

However, optical strain maps provide valuable motion information between successive frames, more so in the case of subtle expressions that may be difficult to distinguish at the feature level. In this paper, we propose a new technique that uses optical strain information as a weighting function for the LBP-TOP feature extractor. This is because pixels with a large displacement in space (large optical strain magnitude) indicate large motion at that particular location and vice versa. Hence, we can increase (or decrease) the importance of the extracted features by placing more (or less) emphasis through the use of weights.

To obtain the optical strain magnitudes, first, horizontal and vertical optical flow vectors, (p, q) are calculated for each image frames in a video [32]. Then optical strain magnitude, ε of each pixel for each frame in a video will be computed.

4.2 Spatio-temporal Pooling

Spatial sum pooling is applied on each optical strain image, where each of the strain map image will first be partitioned into $N \times N$ non-overlapping blocks,

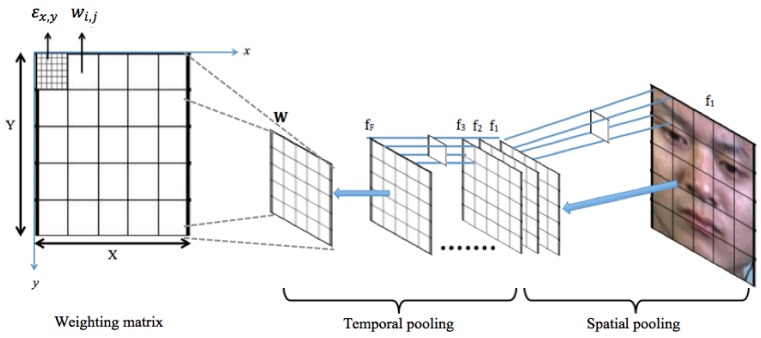

Fig. 3. Spatial-temporal sum pooling of a strain image divided into 5×5 non-overlapping blocks

then all the pixels in that particular block will be summed up. Spatial sum pooling can be computed for each block in an image as follows:

$$s_{i,j} = \sum_{y=(j-1)H+1}^{jH} \sum_{x=(i-1)L+1}^{iL} \varepsilon_{x,y}, \; i,j \in 1 \ldots N \tag{8}$$

where (i,j) and (X,Y) are the block's coordinate and width and height of the frame in (x,y). L and H are equal to X/N and Y/N respectively. Temporal sum pooling is then performed by summing up the resultant optical strain magnitudes of each block from the first frame, f_i to the last frame, f_F.

Hence, for each video, a weight matrix $W = \{w_{i,j}\}_{i,j=1}^{N}$ is formed using spatial-temporal sum pooling (process illustrated in Fig. 3) where each block weight value is given by

$$w_{i,j} = \sum_{t=1}^{F} s_{i,j} = \sum_{t=1}^{F} \sum_{y=(j-1)H+1}^{jH} \sum_{x=(i-1)L+1}^{iL} \varepsilon_{x,y} \tag{9}$$

4.3 Obtaining Block Weights for XY Plane Histogram

Subsequently, all weight matrices are max-normalized to increase the significance of each weighting value. As optical strain magnitudes only describe the expression details in spatial information, the weighting values should be effective on the XY plane only. As such, the resultant histogram is obtained by multiplying the histogram bins of the XY plane with the weighting values, as illustrated in Fig. 4. The new feature histogram is given as:

$$G_{i,j,c,b} = \begin{cases} w_{i,j}\overline{H}_{i,j,c,b}, & \text{if } c = 0 \\ \overline{H}_{i,j,c,b}, & \text{else} \end{cases} \tag{10}$$

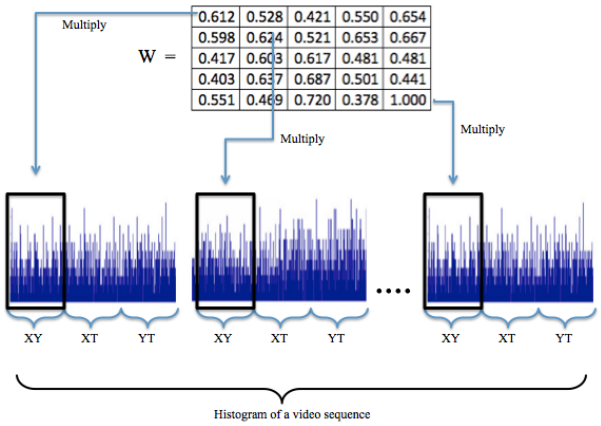

Fig. 4. Multiplication of weighting matrix to X-Y plane of histogram bins

5 Experiments

5.1 Subtle Expression Databases

There are only a few known subtle or micro-expression databases available due to numerous difficulties in the creation process; proper elicitation of stimuli and ground-truth labelling. To evaluate our proposed methods, we consider two of the most recent and comprehensive databases: CASMEII [16] and SMIC (Spontaneous Micro-expression Database) [17]. The databases are recorded under constrained lab condition and all the images have been preprocessed with face registration and alignment.

CASMEII consists of 26 candidates (mean age of 22.03 years), containing 247 spontaneous and dynamic micro-expression clips. The videos are recorded using Point Grey GRAS-03K2C camera with a frame rate of 200*fps* and a spatial resolution of 280 × 340 pixels. There are 5 micro-expression classes (tense, repression, happiness, disgust and surprise) and selection was done by two coders then marked based on the AUs, participants' self report as well as the content of the clip episodes. Each sample contains the ground-truth of onset and offset frames, emotions labeled and AUs represented. The baseline performance reported in CASMEII for 5-category classification is 63.41%. This was obtained using a block-based LBP-TOP consisting of 5×5 blocks. Support Vector Machine (SVM) was used as classifier with leave-one-out cross validation (LOOCV).

SMIC contains 164 micro-expression samples from 16 participants (mean age of 28.1 years). The camera used to capture participant's face is a high speed camera (PixeLINK PL-B774U) with 100*fps* and a resolution of 640 × 480 pixels. There are three classes of micro-expressions: positive (happy), negative (sad, fear, digust) and surprise. The micro-expressions are selected by two coders based on participants' self report and the suggestion by [2] to view the video

frame-by-frame with increasing speed. The reported baseline 3-class recognition performance for SMIC is 48.78% using polynomial kernel of degree six in SVM classifier based on leave-one-subject-out cross-validation (LOSOCV) setting. All image frames from each video are first interpolated to ten frames by temporal interpolation model (TIM) [18], while features were extracted using $LBP\text{-}TOP_{4,4,4,1,1,3}$ with block size of 8×8.

5.2 Pre-processing

Gaussian Filtering. Since the motions characterized by the subtle facial expressions are very fine and we are using the cropped and resampled frames for both databases, it is likely that the presence of unwarranted noise from the acquisition or down-sampling process might be incorrectly identified as fine facial motions. Thus, as a feasible pre-processing step, all the images are filtered by 5×5 pixel Gaussian filter ($\sigma = 0.5$) to suppress the background noise present in the images. The filter size and standard deviation value are empirically determined. Fig. 5 shows the difference of an image before and after filtering.

Fig. 5. Sample image from CASMEII before (left) and after (right) applying Gaussian filter

Noise Block Removal. The two bottom corner blocks (bottom left and bottom right) are removed entirely from consideration in the feature histogram by setting their respective weights to zero, i.e. $\{w_{N,1}, w_{N,N}\} = 0$. This results in only the remaining $N^2 - 2$ weights to be effective on the XY-plane histograms. The reason of removing these 2 blocks from consideration is because there are unexpectedly high optical strain magnitudes that do not correspond to the desired facial movements but are rather unfortunately caused by background/clothing texture noise or wirings from the headset worn by the participants. This problem is consistent across both CASMEII and SMIC datasets, as can be clearly seen in Fig.6. Analogously, the authors of [11] and [14] applied masking technique at consistently noisy regions of the face that unnecessarily affect the optical strain, such as eyes (blinking) and mouth (opening/ closing) regions.

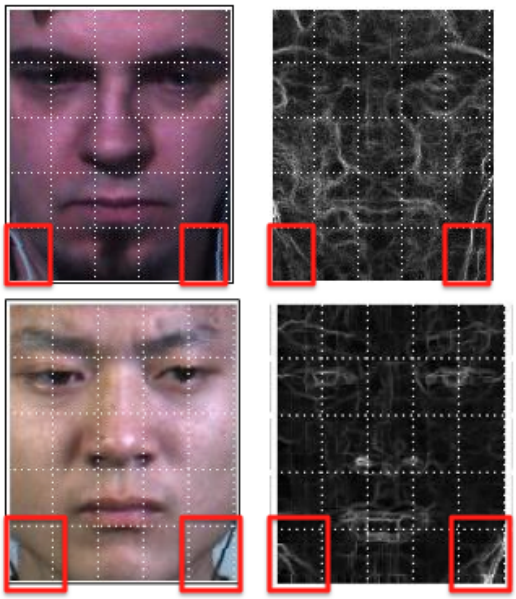

Fig. 6. Top row: sample image from SMIC (left) and its optical strain map image (right). Bottom row: sample image from CASMEII (left) and its optical strain map image (right).

5.3 Results and Discussions

Experiments were conducted on both CASMEII and SMIC databases based on carefully configured settings in order to validate the effectiveness of our proposed method in improving recognition of subtle facial expressions. In our experiments, we performed classification using SVM with leave-one-out cross-validation (LOOCV) in CASMEII and leave-one-subject-out cross-validation (LOSOCV) in SMIC in order to appropriately compare with the baselines reported in the original CASMEII and SMIC papers. In our work, CASMEII is evaluated using linear and RBF kernel, whereas SMIC uses linear, RBF and polynomial kernel with degree six. There are two ways to calculate the classification performance in LOSOCV approach, which are *macro-* and *micro*-averaging. Macro-averaged results are the average accuracy of per-subject results. Micro-averaged results are the average accuracy across all individual results (per sample) which can be obtained from the confusion table that summarizes the overall performance.

To establish our baseline evaluation, the standard methods employed by the original authors of CASMEII and SMIC [16, 18]— LBP-TOP for feature extraction and SVM for classification, were used. For CASMEII, we opt for the best reported configuration, that is $LBP\text{-}TOP_{4,4,4,1,1,4}$. As for SMIC, we used both $LBP\text{-}TOP_{4,4,4,1,1,3}$ and $LBP\text{-}TOP_{4,4,4,1,1,4}$. CASMEII baseline used the block configuration of 5×5 blocks, whereas SMIC used 8×8 blocks.

Table 1. Accuracy results (%) on CASMEII database based on LOOCV

Methods	SVM kernel:	RBF	Linear
Baseline: LBP-TOP		63.97	61.94
OSW-LBP-TOP		**65.59**	**62.75**

Table 2. Accuracy results (%) on CASMEII database based on LOOCV with pre-processing (PP)

Methods	SVM kernel:	RBF	Linear
Baseline: LBP-TOP (with PP)		63.56	63.97
OSW-LBP-TOP (with PP)		**66.40**	62.75

In our experiments, we evaluated our proposed Optical Strain Weighted (OSW) LBP-TOP method (denoted as **OSW-LBP-TOP** in table of results) against the baseline method of **LBP-TOP**. Apart from that, we also examined the method with pre-processing, which filters all the images using Gaussian filter and removes two specific "noise blocks" that are contributing to surplus motions unrelated to facial expressions. For the basic weighted method, all $(N \times N)$ weight coefficients are multiplied with the respective histogram bins of the XY plane. The tables shows the recognition accuracy of the evaluated methods for both CASMEII and SMIC, using SVM classifier with leave-one-out cross-validation (LOOCV) and leave-one-subject-out cross-validation (LOSOCV) respectively.

Generally, the recognition capabilities of the LBP-TOP descriptor demonstrated encouraging signs of improvement when the features are weighted using the proposed scheme. The pooled optical strain magnitudes as block weights intuitively increases the classification accuracy. Crucially, more weightage is assigned to blocks that exhibit more movements, and vice versa, so that the significance of each block histogram can be scaled accordingly. The OSW-LBP-TOP method, with pre-processing obtained the best CASMEII result of 66.4% (RBF kernel), an increase of 2.84% over the baseline. It managed to achieve 65.59% (RBF kernel) without pre-processing, an increase of 1.62% over the baseline. The recognition results of CASMEII are illustrated in Table 1 and 2.

On the other hand, the OSW-LBP-TOP method is consistently superior in the SMIC database. With the $LBP\text{-}TOP_{4,4,4,1,1,3}$ setting from the original paper [17], we are able to obtain an improvement of 3.6% (linear and RBF kernel) without pre-processing and 4.49% (polynomial kernel) of increment with pre-processing, as shown in Table 3 and Table 4 respectively. However, we discovered that with parameters $LBP\text{-}TOP_{4,4,4,1,1,4}$, we are able to generate better baselines while the proposed OSW-LBP-TOP method performed even better with pre-processing. An increment of 1.83% (polynomial kernel) and 5.13% (RBF ker-

Table 3. Accuracy results (%) on SMIC database using $LBP\text{-}TOP_{4,4,4,1,1,3}$ based on LOSOCV

Methods	SVM kernel:	Macro RBF	Macro Linear	Macro Poly	Micro RBF	Micro Linear	Micro Poly
Baseline		43.11	43.11	51.63	43.29	43.29	48.78
OSW-LBP-TOP		**46.71**	**46.71**	**51.70**	**46.34**	**46.34**	**49.39**

Table 4. Accuracy results (%) on SMIC database using $LBP\text{-}TOP_{4,4,4,1,1,3}$ based on LOSOCV with pre-processing (PP)

Methods	SVM kernel:	Macro RBF	Macro Linear	Macro Poly	Micro RBF	Micro Linear	Micro Poly
Baseline (with PP)		44.06	44.06	48.94	42.07	42.07	46.34
OSW-LBP-TOP (with PP)		**47.17**	**47.17**	**53.43**	**46.34**	**46.34**	**50.00**

nel) were achieved in cases of without and with pre-processing, as tabulated in Table 5 and Table 6 respectively.

The improvement in accuracy is apparent on both databases, albeit the fact that the choice of SVM kernel seems to play an equally important role as well. Notably, the OSW-LBP-TOP methods easily outperform the CASMEII baseline result when the RBF kernel is used for the SVM classifier. In the case of SMIC when the OSW-LBP-TOP methods are used, all the three kernels consistently produced improved results. This is an interesting finding that requires further investigation as to how these weights impact and alter the sample distribution to the advantage of specific linear or nonlinear (RBF in this case) kernel types.

Another observation that is worth highlighting for subtle micro-expression research is that sufficient attention should be given to deal with the impact of noise on the recognition performance. The addition of essential pre-processing steps to suppress image noise and remove the noisy blocks are able to produce better results. This can be attributed to the discarding of the histogram bins (set to zero) or features that belong to those noisy regions of the image.

6 Conclusion

In this paper, we have presented a novel method for recognizing subtle expressions in video sequence. The proposed optical strain weighted feature extraction method for subtle expression recognition is able to achieve 66.4% accuracy for a five-class classification on CASMEII database and a 57.71% accuracy for a three-class classification on SMIC database. However, due to the subtlety of facial micro-expressions, the presence of image noise is a challenging problem that requires attention. For future works, the weighting scheme can be extended to the classifier kernel distances to further increase the effectiveness and robustness

Table 5. Accuracy results (%) on SMIC database using $LBP\text{-}TOP_{4,4,4,1,1,4}$ based on LOSOCV

Methods	SVM kernel:	Macro RBF	Macro Linear	Macro Poly	Micro RBF	Micro Linear	Micro Poly
Baseline		55.65	55.65	57.63	51.83	51.83	51.83
OSW-LBP-TOP		**57.34**	**57.34**	**57.71**	**53.05**	**53.05**	**53.66**

Table 6. Accuracy results (%) on SMIC database using $LBP\text{-}TOP_{4,4,4,1,1,4}$ based on LOSOCV with pre-processing (PP)

Methods	SVM kernel:	Macro RBF	Macro Linear	Macro Poly	Micro RBF	Micro Linear	Micro Poly
Baseline (with PP)		51.66	51.66	55.04	47.56	47.56	50.00
OSW-LBP-TOP (with PP)		**56.79**	**56.09**	**57.54**	**53.05**	**52.44**	**53.05**

in the classification stage. Also, noise suppression schemes can also be introduced reduce the impact of noisy textures.

References

1. Ekman, P., Friesen, W.V.: Constants across cultures in the face and emotion. Journal of personality and social psychology **17(2)** (1971) 124
2. Ekman, P.: Lie catching and microexpressions. The philosophy of deception (2009) 118–133
3. Porter, S., ten Brinke, L.: Reading between the lies identifying concealed and falsified emotions in universal facial expressions. Psych. Science **19** (2008) 508–514
4. Frank, M.G., Herbasz, M., Sinuk, K., Keller, A., Kurylo, A., Nolan, C.: I see how you feel: Training laypeople and professionals to recognize fleeting emotions. In: Annual meeting of the International Communication Association, Sheraton New York, New York City, NY. (2009)
5. O'Sullivan, M., Frank, M.G., Hurley, C.M., Tiwana, J.: Police lie detection accuracy: The eect of lie scenario. Law and Human Behavior **33.6** (2009) 530–538
6. Frank, M.G., Maccario, C.J., Govindaraju, V. In: Protecting Airline Passengers in the Age of Terrorism. ABC-CLIO (2009)
7. D'hooge, J., Heimdal, A., Jamal, F., Kukulski, T., Bijnens, B., Rademakers, F., ..., Sutherland, G.R.: Regional strain and strain rate measurements by cardiac ultrasound: principles, implementation and limitations. European Journal of Echocardiography **1.3** (2000) 154–170
8. Shreve, M., Manohar, V., Goldgof, D., Sarkar, S.: Face recognition under camouflage and adverse illumination. In: Biometrics: Theory Applications and Systems (BTAS), 4th IEEE Int. Conf. on. (2010) 1–6
9. Manohar, V., Goldgof, D., Sarkar, S.: Facial strain pattern as a soft forensic evidence. In: Applications of Computer Vision (WACV). (2007)
10. Shreve, M., Godavarthy, S., Manohar, V., Goldgof, D., Sarkar, S.: Towards macro- and micro-expression spotting in video using strain patterns. In: Applications of Computer Vision (WACV). (2009) 1–6

11. Shreve, M., Godavarthy, S., Goldgof, D., Sarkar, S.: Macro-and micro-expression spotting in long videos using spatio- temporal strain. In: Automatic Face, Gesture Recognition and Workshops. (2011) 51–56
12. Vinciarelli, A., Dielmann, A., Favre, S., Salamin, H.: Canal9: A database of political debates for analysis of social interactions. In: In Affective Computing and Intelligent Interaction and Workshops. (2009) 1–4
13. Ekman, P. In: Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage. W. W. Norton and Company (2009) WW Norton and Company.
14. Shreve, M., Brizzi, J., Fefilatyev, S., Luguev, T., Goldgof, D., Sarkar, S.: Automatic expression spotting in videos. Image and Vision Computing **32(8)** (2014) 476–486
15. Zhao, G., Pietikainen, M.: Dynamic texture recognition using local binary patterns with an application to facial expressions. Pattern Analysis and Machine Intelligence, IEEE Transactions **20.6** (2007) 915–928
16. Yan, W.J., Wang, S.J., Zhao, G., Li, X., Liu, Y.J., Chen, Y.H., Fu, X.: Casme ii: An improved spontaneous micro-expression database and the baseline evaluation. PLoS ONE **9** (2014) e86041
17. Pfister, T., Li, X., Zhao, G., Pietikainen, M.: Recognising spontaneous facial micro-expressions. In: Computer Vision (ICCV). (2011) 1449–1456
18. Li, X., Pfister, T., Huang, X., Zhao, G., Pietikainen, M.: A spontaneous micro-expression database: Inducement, collection and baseline. In: Automatic Face and Gesture Recognition. (2013) 1–6
19. Boureau, Y.L., Ponce, J., LeCun, Y.: A theoretical analysis of feature pooling in visual recognition. (2010) 11–118 Machine Learning (ICML-10).
20. Hamel, P., Lemieux, S., Bengio, Y., Eck, D.: Temporal pooling and multiscale learning for automatic annotation and ranking of music audio. In: International Society for Music Information Retrieval Conference. (2011) 729–734
21. Forsyth, D.A., Ponce., J.: Computer vision: a modern approach. Prentice Hall (2002)
22. Ekman, P., Friesen, W.V.: Facial action coding system. (1978)
23. Lien, J.J.J., Kanade, T., Cohn, J.F., Li, C.C.: Detection, tracking, and classification of action units in facial expression. Robotics and Autonomous Systems **31.3** (2000) 131–146
24. Liu, Z., Shan, Y., Zhang, Z.: Expressive expression mapping with ratio images. In: Computer graphics and interactive techniques. (2001) 271–276
25. Anitha, C., Venkatesha, M.K., Adiga, B.S.: A survey on facial expression databases. Int. Journal of Engineering Science and Technology **2.10** (2010) 5158–5174
26. Yan, W.J., Wang, S.J., Liu, Y.J., Wu, Q., Fu, X.: For micro-expression recognition: Database and suggestions. Neurocomputing **136** (2014) 82–87
27. Polikovsky, S., Kameda, Y., .O.Y.: Facial micro-expressions recognition using high speed camera and 3d-gradient descriptor. (In: Crime Detection and Prevention)
28. Warren, G., Schertler, E., Bull, P.: Detecting deception from emotional and unemotional cues. Journal of Nonverbal Behavior **33.1** (2009) 59–59
29. Barron, J.L., Thacker, N.A.: Tutorial: Computing 2d and 3d optical flow (2005) Imaging Science and Biomedical Eng. Div., Medical School, Univ. of Manchester.
30. Jain, R., Kasturi, R., Schunck, B.G.: Machine vision. Volume 5. McGraw-Hill Education (1995)
31. Godavarthy, S.: Microexpression spotting in video using optical strain. Masters thesis, University of South Florida (2010)
32. Sun, D., Roth, S., Black, M.J.: Secrets of optical flow estimation and their principles. In: Computer Vision and Pattern Recognition. (2010) 2432–2439