Micro-expression Recognition Under Low-resolution Cases

Guifeng Li[1], Jingang Shi[2], Jinye Peng[1,*] and Guoying Zhao[1,2]
[1]*Department of Information Science and Technology, Northwest University, Xi'an, China*
[2]*Center for Machine Vision and Signal Analysis, University of Oulu, Oulu, Finland*

Keywords: Micro-expression Recognition, Surveillance Video, Low-resolution, Super-resolution, Fast LBP-TOP.

Abstract: Micro-expression is an essential non-verbal behavior that can faithfully express the human's hidden emotions. It has a wide range of applications in the national security and computer aided diagnosis, which encourages us to conduct the research of automatic micro-expression recognition. However, the images captured from surveillance video easily suffer from the low-quality problem, which causes the difficulty in real applications. Due to the low quality of captured images, the existing algorithms are not able to perform as well as expected. For addressing this problem, we conduct a comprehensive study about the micro-expression recognition problem under low-resolution cases with face hallucination method. The experimental results show that the proposed framework obtains promising results on micro-expression recognition under low-resolution cases.

1 INTRODUCTION

Facial expression is one of the most significant reflection for human emotion information. In recent years, the analysis of facial expression has attracted more and more attention in the field of computer vision. Different from the general facial expressions, micro-expression is spontaneous, involuntary and is an instinct of human beings. Commonly, the duration of micro-expression is only from 1/25 to 1/2 second, while the intensity of expression is also at a low level. Because of the above characteristics, micro-expression analysis becomes very challenging. The application of micro-expression analysis is pervasive and could be utilized in the security, judicial system, clinical diagnosis, public management and auxiliary education. At present, researchers have collected a variety of high-quality datasets to evaluate micro-expression recognition algorithms, and achieved reasonable recognition accuracies on these datasets. In the real world, however, very often we can have only low-quality video clips from traditional devices *e.g.* surveillance cameras or closed-circuit television. Therefore, it is especially crucial to conduct micro-expression analysis in such harsh cases.

The main tasks of micro-expression analysis are spotting and recognition. The spotting is to find the fragments of the micro-expression from the input video sequence, and the recognition is to determine the types of micro-expression fragment (*e.g.*, happiness, disgust, repression or surprise). Because the research in this paper focuses on the recognition of micro-expression, it is assumed that the video clips mentioned here are already spotted and only contain micro-expression.

In recent years, researchers have proposed many methods for spontaneous micro-expression recognition. Pfister et al. (Pfister et al., 2011) proposed the first automatic micro-expression recognition system on spontaneous micro-expression dataset, which utilized temporal interpolation model (TIM) (Zhou et al., 2011) together with Multiple Kernel Learning (MKL) to capture the main variation of image sequences, and employed the Local Binary Patterns from Three Orthogonal Planes (LBP-TOP) (Zhao and Pietikäinen, 2007) to extract the feature descriptors and Random Forest (RF) as the classifiers. Since then, many researchers have developed various versions and variants of descriptors to improve the micro-expression recognition accuracies (Ruiz-Hernandez and Pietikäinen, 2013; Wang et al., 2014b; Wang et al., 2014c; Wang et al., 2015). Wang et al. (Wang et al., 2014a) employed Tensor Independent Color Space (TICS) to extract discriminative features for recognition. Ngo et al. (Le Ngo et al., 2014) utilized the Selective Transfer Machine (STM) for micro-expressions image sequences pre-processing in order to solve imbalance and different facial morphology in the dataset. Liong et al. (Liong et al., 2014) obtained the subtle displacement of the faces within a temporal interval from op-

Li, G., Shi, J., Peng, J. and Zhao, G.
Micro-expression Recognition Under Low-resolution Cases.
DOI: 10.5220/0007373604270434
In *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2019)*, pages 427-434
ISBN: 978-989-758-354-4

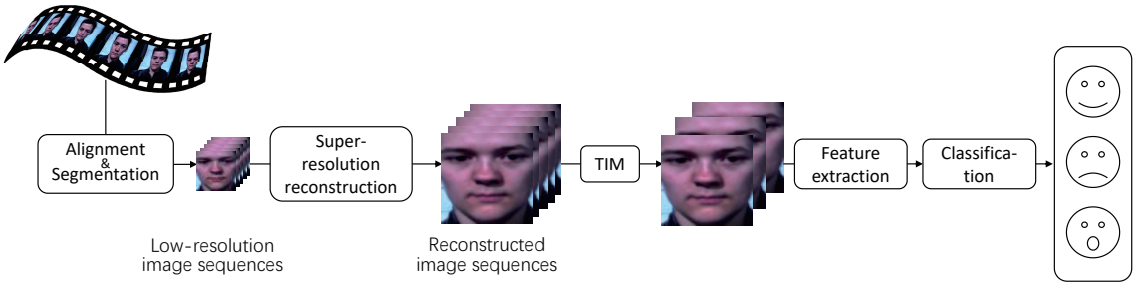

Figure 1: The flowchart of the proposed recognition system.

tical strain magnitudes and assigned different weights to the local features to form a new feature. Lu et al. (Lu et al., 2014) introduced Delaunay-based temporal coding model (DTCM) to normalize the image sequences in the spatiotemporal domain. Oh et al. (Oh et al., 2015) acquired multi-scale monogenic signals through the Riesz wavelet transform, and combined the features of magnitude, phase, and orientation for micro-expression recognition. Liu et al. (Liu et al., 2016) developed the Main Directional Mean Optical-flow (MDMO) to explore the discriminative features from micro-expression. It simultaneously considers local statistic motion and spatial position information by the use of a robust optical flow model. Li et al. (Li et al., 2017) extended Histograms of Oriented Gradients (HOG) and Histograms of Image Gradient Orientation (HIGO) on three orthogonal planes, and proposed HOG-TOP and HIGO-TOP based on the idea of LBP-TOP. Xu et al. (Xu et al., 2017) exploited optical flow estimation to conduct pixel-level alignment in the chosen granularity for the micro-expression image sequences, and obtained principal optical flow direction, as the subtle facial dynamic feature descriptor. The proposed micro-expression descriptor is also called the Facial Dynamics Map. He et al. (He et al., 2017) introduced a multi-task mid-level feature learning method for feature extraction, which has the ability to obtain more discriminative mid-level features with more generalization ability. Recently, some researchers utilized the novel popular deep learning algorithm to learn deep features for micro-expression recognition, but the results are far away from satisfactory. The primary reason is that deep learning algorithm requires a lot of training samples, but the scale of the current dataset is extremely limited (Patel et al., 2016).

Although recent micro-expression recognition algorithm achieves reasonable results, the performance highly depends on the quality of facial video clip. Once the quality of facial video clip used for recognition is poor (such as the low-resolution), the above algorithm will not work well. The reason mainly lies in two aspects: (i) The low-resolution images lose a lot of detail information, which induces the difficulty to extract the available features from the low-resolution image sequences (Lei et al., 2011). (ii) The low-resolution images are not homogeneous with the high-resolution ones (*e.g.*, different resolution and different clarity), which prevents us to directly use the low-resolution images as input in the testing phase. In the real world, the facial image captured from the surveillance video usually only accounts for a small part of the entire picture. For example, the SMIC dataset for micro-expression recognition has a facial resolution of 190×230. However, the captured facial image sequences by the surveillance videos are often under the resolution of 50×50 (or lower). This means that the previous micro-expression recognition methods cannot be directly utilized to deal with the low-resolution case. Therefore, micro-expression recognition research under low-resolution cases is vital and challenging.

To solve the above problems, we perform the research of low-resolution micro-expression recognition with the help of recent facial hallucination methods. We first hallucinate the low-quality facial image sequence to recover the lost dynamic characteristics. Then, we conduct the traditional micro-expression recognition methods to explore low-quality micro-expression recognition. We evaluate the performance of micro-expression recognition accuracies under different resolutions to investigate the relationship between resolution and recognition accuracy. Generally, the target of this paper is to make a comprehensive study about the influence of resolution in micro-expression recognition, while also develop a framework to deal with micro-expression recognition task in low-quality condition.

The rest of this paper is organized as follows. In Section 2 we describe our micro-expression recognition framework. Section 3 presents experimental results and discussion. Conclusions are drawn in Section 4.

2 THE PROPOSED RECOGNITION FRAMEWORK

The low-resolution micro-expression recognition process include image sequence pre-processing, super-resolution reconstruction, feature extraction and classification, as shown in Figure 1. More details of description are as follows:

2.1 Low-resolution Datasets

There are several datasets for micro-expression recognition, such as SMIC (Li et al., 2013) and CASME II (Yan et al., 2014). However, all of these datasets are high-definition image sequences acquired by professional cameras in specific circumstances. Figure 2 shows two frames from a video clip of the SMIC-HS dataset. We can find subtle changes in facial expressions within the red box. In particular, the movement in the area of the white ellipse and the position of the white arrow are more obvious. If the image resolution is too low, these details are hard to be noticed.

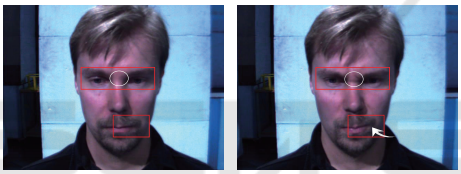

Figure 2: Two frames from a video clip of the SMIC-HS dataset.

Since there is no low-resolution image sequence in the existing spontaneous micro-expression datasets, we use image deterioration processing to obtain simulated low-resolution micro-expression image sequences. In the paper (Wang et al., 2014d), the low-resolution images are divided into three categories: small size, poor quality, and small size & poor quality. We consider the third type of images (small size & poor quality) that is closer to the situation of the real applications as the simulated image.

In the image reconstruction task, low-resolution image sequences are obtained by blurring, down-sampling, and noising processes from high-resolution image sequences (Shi et al., 2018):

$$\boldsymbol{L} = \boldsymbol{DBH} + \boldsymbol{n} \tag{1}$$

where $\boldsymbol{D}$ and $\boldsymbol{B}$ are down-sampling and blurring respectively, $\boldsymbol{H}$ is high-resolution image, $\boldsymbol{n}$ represents the additive noise, and $\boldsymbol{L}$ is low-resolution image.

2.2 Image Pre-processing

In our proposed framework, the pre-processing mainly includes three steps: face alignment, face segmentation, and TIM. There are natural pose variation and involuntary movement in raw collected videos. At the same time, micro-expression video clips are collected from different participants, with different gender, age, and ethnics. Therefore, to avoid the interference of the above-mentioned non-expression factors, it is indispensable to conduct the face alignment and face segmentation procedure.

We select a frame with frontal face and neutral expression from a particular segment as the canonical template, and manually locate the position of two eyes. Then, the Active Shape Model (ASM) is carried out to detect 68 facial landmarks (Cootes et al., 1995). The relationship between 68 facial landmarks of canonical frame and 68 facial landmarks of other frames is established by the Local Weighted Mean (LWM) (Goshtasby, 1988), and the micro-expression images are aligned to the canonical frame to minimize interference caused by non-expression factors.

Micro-expression videos have various length, from 4 frames to 50 frames (if captured by a camera with 100 fps). To solve the problem of different lengths in video clips, Li et al. (Li et al., 2017) used the TIM algorithm to map all the frames of a sequence onto a curve, sample the newly synthesised facial images with a fixed interval, and finally obtained the same pre-defined sequence length. The experimental results show that the algorithm has improved the recognition accuracy. Figure 3 shows the mapping process of TIM (Zhou et al., 2011).

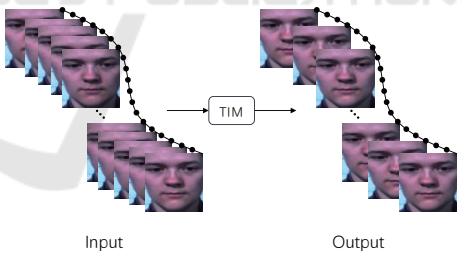

Figure 3: The input is the original image sequences, and the output is the TIM interpolated image sequences.

2.3 Super-resolution Reconstruction

Low-resolution images and high-resolution images are heterogeneous in both quality and resolution. The micro-expression recognition method of high-resolution image sequences cannot be directly applied to low-resolution image sequences. In Section 2.1, we present the procedure of generating low-resolution images from high-resolution images.

To reconstruct high-resolution images, the paper (Shi et al., 2018) proposed a novel face hallucination algorithm. It combines the patch-based regularization

term and the pixel-based regularization term to constrain the objective function. The reconstructed high-resolution image $\boldsymbol{H}$ can be obtained by minimizing the following objective function:

$$f(\boldsymbol{H}) = \|\boldsymbol{L} - \boldsymbol{DBH}\|_2^2 + \alpha \boldsymbol{F}_{patch} + \eta \boldsymbol{F}_{pixel} + \lambda \boldsymbol{F}_{penalty} \quad (2)$$

where the first item on the right side is the reconstruction error, and the last three items are patch-based regularization terms, pixel-based regularization terms, and penalty terms respectively. Figure 4 illustrates the whole hallucination task. For details, please refer to (Shi et al., 2018).

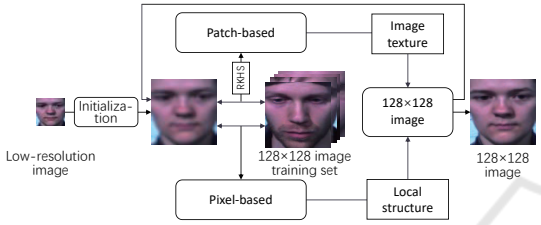

Figure 4: Super-resolution Reconstruction.

2.4 Micro-expression Recognition

As shown in Figure 1, the micro-expression recognition is mainly divided into two parts: feature extraction and classification. In the previous micro-expression analysis methods (Pfister et al., 2011; Wang et al., 2014b; Wang et al., 2014c; Li et al., 2017; He et al., 2017), the researchers show the advantages of LBP-TOP and its variants as feature descriptors.

Different from the traditional LBP feature that mainly focuses on a single image, LBP-TOP can capture dynamic variations in both spatial and temporal domains, which is essential for micro-expression recognition. We first divide the whole facial image sequence into several cuboids, such as $5 \times 5 \times 1$, $8 \times 8 \times 2$, etc., where the first two parameters determine the number of the blocks in spatial domain, and the last parameter is the number of segments in the temporal direction. Each cuboid can be considered as a new unit. The LBP features are extracted from three different orthogonal planes (XY, XT, and YT planes) in the new unit. We traverse all the cuboids to obtain the LBP-TOP features of the image sequence and then LBP-TOP features of each cuboid are concatenated. As shown in Figure 5.

In the classification part, we use the linear support vector machine (LSVM) (Chang and Lin, 2011) as the classifier. To make a fair comparison, we employ the leave-one-subject-out protocol in the experiments. According to the micro-expression labels provided by the dataset publisher, we classify the samples from

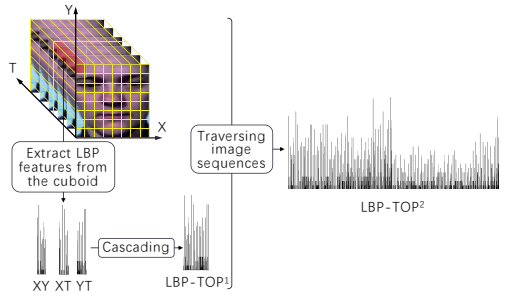

Figure 5: LBP-TOP feature extraction of image sequences. XY, XT, and YT refer to the XY plane, the XT plane, and the YT plane, respectively. LBP-TOP^1 refers to the LBP-TOP feature of a single cuboid. LBP-TOP^2 refers to the LBP-TOP feature of the entire image sequence.

SMIC into three categories (positive, negative, and surprised), and samples from CASME II into five categories (happiness, surprise, repression, disgust, and others).

3 EXPERIMENT AND ANALYSIS

We now present the experiments and results on three different spontaneous micro-expression datasets, *i.e.*, the SMIC-HS, SMIC-subHS and CASME II. The experimental parameter setting and result analysis will be discussed in the following subsections.

Table 1: The summary of used datasets.

	SMIC-HS	SMIC-subHS	CASME II
Micro-Clips	164	71	247
Participants	16	8	26
Classes	3	3	5

3.1 Pre-processing

SMIC-HS and SMIC-subHS are two subsets of SMIC. The SMIC-HS dataset contains 164 spontaneous micro-expression clips from sixteen participants, which are categorized into three categories: positive (51 clips), negative (70 clips), and surprise (43 clips). The SMIC-subHS dataset (Li et al., 2017) is a subset of SMIC-HS and contains only the last eight participants. The number of micro-expression clips from each of the first eight subjects varies a lot, three subjects have contributed almost half micro-expression samples of the whole set, which could affect the leave-one-subject-out performance, while the later eight subjects's (SMIC-subHS) clips number are more evenly distributed. And the number of positive, negative and surprise clips is 28, 23 and 20, respectively in the SMIC-subHS dataset. Meanwhile, the CASME II dataset contains twenty-six participants

belonging to five different categories: surprise (25 clips), happiness (32 clips), others (99 clips), disgust (64 clips), and repression (27 clips). Table 1 shows a summary of the datasets used in the experiments. The facial resolution of the high-resolution image is set as 128×128 in the experiment. The images of 128×128 resolution are downsampled by 2, 4 and 8 times to obtain low-resolution images (as shown in Figure 6). This means that we evaluate the low-resolution facial image sequences of three various levels (*e.g.*, 16×16, 32×32, 64×64) in the micro-expression recognition tasks.

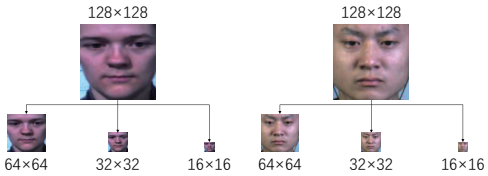

Figure 6: (a) SMIC-HS/SMIC-subHS low-resolution image, (b) CASME II low-resolution image.

3.2 Reconstruction

In this section, the low-resolution image is reconstructed into a high-resolution image using the method proposed in the paper (Shi et al., 2018), which is briefly introduced in Section 2.3. Tables 2-3 list the average peak signal to noise ratio (PSNR) and structural similarity (SSIM) index of the reconstructed image sequences for different datasets under various resolutions. Here, we use S64, S32, and S16 to name the reconstructed image sequences from resolution 64×64, 32×32, and 16×16, respectively.

Table 2: The average PSNR (dB) indexs of the reconstructed image sequences for the SMIC-HS, SMIC-subHS and CASME II datasets under the different resolution.

PSNR (dB)	16×16	32×32	64×64
SMIC-HS	31.25	37.67	44.30
SMIC-subHS	31.67	38.26	43.22
CASME II	31.80	36.49	37.83

Table 3: The average SSIM indexs of the reconstructed image sequences for the SMIC-HS, SMIC-subHS and CASME II datasets under the different resolutions.

SSIM	16×16	32×32	64×64
SMIC-HS	0.9397	0.9775	0.9883
SMIC-subHS	0.8970	0.9346	0.9424
CASME II	0.9439	0.9761	0.9882

As shown in Tables 2 and 3, the quantitative indicator (PSNR/SSIM) of the reconstructed facial image sequences is proportional to the resolution of input facial image sequences. For example, in the SMIC-HS dataset, the PSNR index of S16 is 31.25dB, which is 6.42dB lower than S32, and 13.05dB less than S64. For the SSIM index, S16 achieves the value of 0.9397, which is 0.0378 inferior to S32, and 0.0486 lower than S64. Besides, Figure 7 presents the visual performance of the reconstructed image sequences, which also indicates the same conclusion as the above viewpoint.

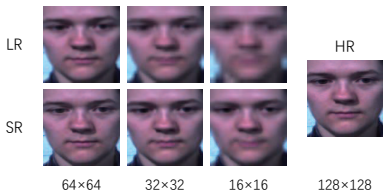

Figure 7: Comparison of reconstruction results at different resolutions.

3.3 Recognition

To normalize the duration of the video clips, the frame number of the video clips is interpolated to 10 frames by TIM algorithm (Li et al., 2017) as introduced in Section 2.2. We apply the fast LBP-TOP (Hong et al., 2016) to divide the video clips into different cuboids and extract the LBP-TOP feature of each cuboid to constitute a complete feature, where uniform mapping is used, the radius r is set to $r = 2$, and the number of neighboring points p is set to $p = 8$. We use the leave-one-subject-out protocol to conduct the experiments, i.e., use all the samples of one subject as the testing set, and the samples from all the other subjects as the training set. We employ LSVM as the classifier, where penalty coefficient $c = 1$.

3.3.1 Recognition of Low-resolution Image Sequences

In this subsection, we present a baseline for the performance of micro-expression recognition of low-resolution image sequences. To adapt the testing samples of various resolutions, we downsample the training set from 128×128 to the corresponding resolution (i.e., the same with testing samples) in order to conduct the classification procedure. Notice that the downsampling operation causes the lack of discriminative features for micro-expression. The following experiments also show that the recognition accuracy dramatically decreases under very low resolution.

Figure 8 shows the recognition accuracy of different resolution image sequences at different datasets. Here, we use L64, L32, and L16 to name the low-resolution image sequences respectively. From Figure 8, it is found that the recognition accuracy of

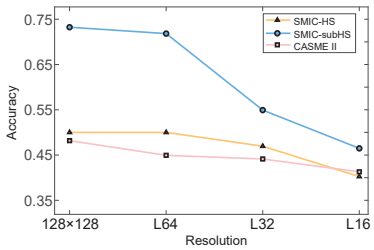

Figure 8: The recognition accuracy of different resolution image sequences at different datasets.

the SMIC-SubHS dataset (the blue fold line) dramatically decreases when the resolution of input image sequences reduces from 64×64 to 32×32. Meanwhile, we can see that the accuracy of low-resolution image sequences (*e.g.*, L16) is relatively low. This phenomenon indicates that it is hard to acquire satisfactory results with low-resolution image sequences. The main reason is that the low-resolution causes the lack of high-frequency information and texture details in describing the micro-expression.

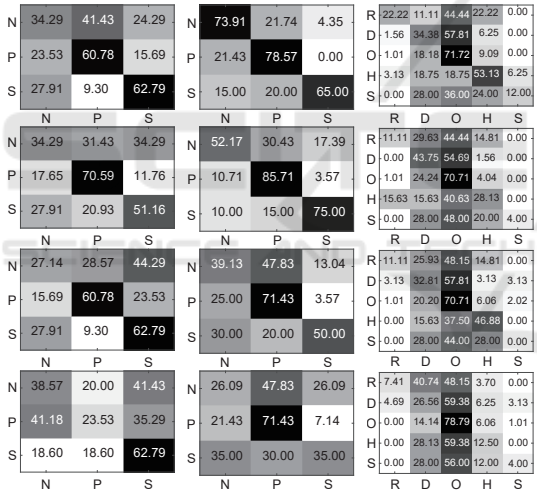

Figure 9: Confusion matrix of recognition accuracy of image sequences with different resolution. In the horizontal direction, Left: SMIC-HS, Middle: SMIC-subHS, Right: CASME II. In the vertical direction, from top to bottom, there are 128×128, 64×64, 32×32, 16×16, respectively. P: positive, N: negative, S: surprise, R: repression, D: disgust, O: others, H: happiness.

Figure 9 shows the confusion matrix for the classification results of low-resolution image sequences and also the performance under 128×128 resolution as the reference. We can find that when the resolution of the image sequences is 128×128 in the SMIC-SubHS dataset (the second column of Figure 9), the confusion matrix is more concentrated on the diagonal, which indicates that the micro-expression recognition method performs well. However, when the resolution of the image sequences decreases, the confusion matrix is gradually to become poor. We can also find that the proportion of misclassification in SMIC-HS (the first column of Figure 9) and CASME II (the third column of Figure 9) is more than the SMIC-subHS, and the recognition accuracy shown in Figure 8 is also relatively low. For SMIC-HS, the main reason for the above problems may be that the last eight subjects (the SMIC-subHS dataset) have a more balanced distribution than the first eight subjects. For CASME II, it is mainly because of the imbalance distribution and excessive categories. For example, the number of video clips from the class OTHERS accounts for 40.08% in the CASME II dataset.

3.3.2 Performance of the Proposed Framework

In this subsection, we carry out the experiments of our proposed framework on three datasets. The testing set is first reconstructed to the resolution of 128×128, and then we conduct the classification in the high-resolution space. According to empirical analysis, we select the optimal parameters for LBP-TOP. Table 4 gives the recognition accuracy together with the corresponding parameter setting of blocksize.

We can see from Table 4 that the experimental results have been significantly improved. For example, in the SMIC-subHS dataset, the recognition accuracy of the image sequences under 64×64 resolution increased from 71.83% to 74.65%, with an increase of 2.82%. The recognition accuracy of S32 is 74.65%, which is 19.72% higher than L32. The recognition accuracy of S16 is 73.24%, which is 26.76% higher than L16. This shows that our method has a good improvement on the micro-expression recognition accuracy of low-resolution image sequences. We also notice that the proposed framework even obtains better results with S64 than directly utilizing the input as original 128×128 image sequences. It could be because that the samples in the original SMIC-HS/subHS dataset suffer from apparent noises in the recording so that the face sequences actually includes redundant and noisy information.

The confusion matrix of recognition accuracy of super-resolution reconstructed image sequences is shown in Figure 10. We exhibit the recognition accuracy of image sequences with 128×128, S64, S32, and S16 according to different datasets. From Figure 10, it is found that the confusion matrix of our proposed framework is more concentrated on the diagonal than Figure 9. Particularly in the 16×16 at the SMIC-HS dataset (bottom left), the recognition accuracy on POSITIVE has significantly improved. Additionally, we can see from the results of SMIC-subHS dataset (the second column) that the proportion of NEGA-

Table 4: Comparison of recognition accuracy of different resolution images on different datasets. High-resolution means the image sequences with the resolution of 128×128. Super-resolution Reconstruction represents the low-resolution image sequences are reconstructed to the 128×128 by the facial hallucination method. Low-resolution denotes the low-resolution image sequence before reconstruction. $X \times Y \times T$ expresses the number of horizontal, vertical and temporal cuboids.

Accuracy (%)	High-resolution	Super-resolution Reconstruction			Low-resolution		
	128×128	64×64	32×32	16×16	64×64	32×32	16×16
SMIC-HS	50.00 ($8 \times 8 \times 2$)	52.44 ($5 \times 5 \times 2$)	51.83 ($5 \times 5 \times 2$)	51.83 ($5 \times 5 \times 2$)	50.00 ($6 \times 6 \times 5$)	46.95 ($6 \times 6 \times 5$)	40.24 ($3 \times 3 \times 6$)
SMIC-subHS	73.24 ($5 \times 5 \times 2$)	74.65 ($5 \times 5 \times 2$)	74.65 ($5 \times 5 \times 2$)	73.24 ($8 \times 8 \times 3$)	71.83 ($6 \times 6 \times 2$)	54.93 ($6 \times 6 \times 2$)	46.48 ($4 \times 4 \times 1$)
CASME II	48.18 ($7 \times 7 \times 3$)	48.18 ($7 \times 7 \times 5$)	44.53 ($7 \times 7 \times 3$)	42.92 ($7 \times 7 \times 5$)	44.94 ($7 \times 7 \times 1$)	44.13 ($4 \times 4 \times 2$)	41.30 ($2 \times 2 \times 5$)

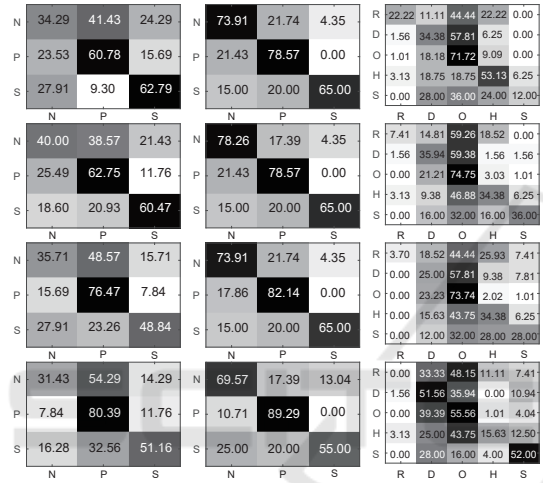

Figure 10: Confusion matrix of recognition accuracy of super-resolution reconstructed image sequences. In the horizontal direction, Left: SMIC-HS, Middle: SMIC-subHS, Right: CASME II. In the vertical direction, from top to bottom, there are 128×128, S64, S32, and S16, respectively.

TIVE misclassified into POSITIVE are significantly reduced, while the proportion of POSITIVE correctly classified are also substantially improved. Unfortunately, even though the results of the CASME II dataset (the third column) have improved, the classification of each category is still very poor, and they are generally falsely classified into OTHERS. Perhaps it is because OTHERS includes all the other types of micro-expressions excluding surprise, happiness, disgust, and repression, so it has mixed categories. In summary, from the comparison between Figure 10 and Figure 9, we see that the proposed framework can obtain very promising performance boosting for low-resolution micro-expression recognition.

4 CONCLUSIONS

In this paper, we give a comprehensive study about the task of low-resolution micro-expression recognition problem. We use blurring and downsampling model to produce and simulate the low-resolution micro-expression facial image sequences. We reconstruct the high-quality facial image sequences by employing facial hallucination method on each frame, which enhances the local details and amplifies the low-quality image sequences to the high resolution ones. Then, we utilize fast LBP-TOP to extract the dynamic features and recognize the micro-expression by SVM classifier. The experimental results illustrate that the proposed framework performs well on publicly available micro-expression datasets (SMIC-HS, SMIC-subHS, and CASME II) on the low-resolution micro-expression recognition problem. In the future, we will focus on the research of deep features for micro-expression recognition under low-resolution cases.

ACKNOWLEDGEMENTS

G. Li is visiting the University of Oulu when the manuscript is completed. His visit is supported by the Graduate Academic Visiting Program of Northwest University and University of Oulu. J. Shi and G. Zhao have been supported by Academy of Finland, Tekes Fidipro Program (1849/31/2015), Tekes Project (3116/31/2017), Infotech, National Natural Science Foundation of China (61772419) and Tekniikan Edistamissaatio Foundation. J. Peng has been supported by Changjiang Scholars and Innovation Research Team in University (IRT_17R87), National Key R&D Program of China (2017YFB1402103), National Nature Science Foundation of China (41427804) and Shaanxi provincial Key R&D Program (2017ZDXM-G-20-1).

The asterisk indicates corresponding author.

REFERENCES

Chang, C.-C. and Lin, C.-J. (2011). Libsvm: a library for support vector machines. In *ACM transactions on intelligent systems and technology (TIST)*. ACM.

Cootes, T. F., Taylor, C. J., Cooper, D. H., and Graham, J. (1995). Active shape models-their training and application. In *Computer vision and image understanding*. Elsevier.

Goshtasby, A. (1988). Image registration by local approximation methods. In *Image and Vision Computing*. Elsevier.

He, J., Hu, J.-F., Lu, X., and Zheng, W.-S. (2017). Multi-task mid-level feature learning for micro-expression recognition. In *Pattern Recognition*. Elsevier.

Hong, X., Xu, Y., and Zhao, G. (2016). Lbp-top: a tensor unfolding revisit. In *Asian Conference on Computer Vision*. Springer.

Le Ngo, A. C., Phan, R. C.-W., and See, J. (2014). Spontaneous subtle expression recognition: Imbalanced databases and solutions. In *Asian conference on computer vision*. Springer.

Lei, Z., Ahonen, T., Pietikäinen, M., and Li, S. Z. (2011). Local frequency descriptor for low-resolution face recognition. In *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*. IEEE.

Li, X., Pfister, T., Huang, X., Zhao, G., and Pietikäinen, M. (2013). A spontaneous micro-expression database: Inducement, collection and baseline. In *Automatic face and gesture recognition (fg), 2013 10th ieee international conference and workshops on*. IEEE.

Li, X., Xiaopeng, H., Moilanen, A., Huang, X., Pfister, T., Zhao, G., and Pietikäinen, M. (2017). Towards reading hidden emotions: A comparative study of spontaneous micro-expression spotting and recognition methods. In *IEEE Transactions on Affective Computing*. IEEE.

Liong, S.-T., See, J., Phan, R. C.-W., Le Ngo, A. C., Oh, Y.-H., and Wong, K. (2014). Subtle expression recognition using optical strain weighted features. In *Asian conference on computer vision*. Springer.

Liu, Y.-J., Zhang, J.-K., Yan, W.-J., Wang, S.-J., Zhao, G., and Fu, X. (2016). A main directional mean optical flow feature for spontaneous micro-expression recognition. In *IEEE Transactions on Affective Computing*. IEEE.

Lu, Z., Luo, Z., Zheng, H., Chen, J., and Li, W. (2014). A delaunay-based temporal coding model for micro-expression recognition. In *Asian conference on computer vision*. Springer.

Oh, Y.-H., Le Ngo, A. C., See, J., Liong, S.-T., Phan, R. C.-W., and Ling, H.-C. (2015). Monogenic riesz wavelet representation for micro-expression recognition. In *Digital Signal Processing (DSP), 2015 IEEE International Conference on*. IEEE.

Patel, D., Hong, X., and Zhao, G. (2016). Selective deep features for micro-expression recognition. In *Pattern Recognition (ICPR), 2016 23rd International Conference on*. IEEE.

Pfister, T., Li, X., Zhao, G., and Pietikäinen, M. (2011). Recognising spontaneous facial micro-expressions. In *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE.

Ruiz-Hernandez, J. A. and Pietikäinen, M. (2013). Encoding local binary patterns using the re-parametrization of the second order gaussian jet. In *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*. IEEE.

Shi, J., Liu, X., Zong, Y., Qi, C., and Zhao, G. (2018). Hallucinating face image by regularization models in high-resolution feature space. In *IEEE Transactions on Image Processing*. IEEE.

Wang, S.-J., Yan, W.-J., Li, X., Zhao, G., and Fu, X. (2014a). Micro-expression recognition using dynamic textures on tensor independent color space. In *2014 22nd International Conference on Pattern Recognition (ICPR)*. IEEE.

Wang, S.-J., Yan, W.-J., Zhao, G., Fu, X., and Zhou, C.-G. (2014b). Micro-expression recognition using robust principal component analysis and local spatiotemporal directional features. In *Workshop at the European conference on computer vision*. Springer.

Wang, Y., See, J., Phan, R. C.-W., and Oh, Y.-H. (2014c). Lbp with six intersection points: Reducing redundant information in lbp-top for micro-expression recognition. In *Asian Conference on Computer Vision*. Springer.

Wang, Y., See, J., Phan, R. C.-W., and Oh, Y.-H. (2015). Efficient spatio-temporal local binary patterns for spontaneous facial micro-expression recognition. In *PloS one*. Public Library of Science.

Wang, Z., Miao, Z., Wu, Q. J., Wan, Y., and Tang, Z. (2014d). Low-resolution face recognition: a review. In *The Visual Computer*. Springer.

Xu, F., Zhang, J., and Wang, J. Z. (2017). Microexpression identification and categorization using a facial dynamics map. In *IEEE Transactions on Affective Computing*. IEEE.

Yan, W.-J., Li, X., Wang, S.-J., Zhao, G., Liu, Y.-J., Chen, Y.-H., and Fu, X. (2014). Casme ii: An improved spontaneous micro-expression database and the baseline evaluation. In *PloS one*. Public Library of Science.

Zhao, G. and Pietikäinen, M. (2007). Dynamic texture recognition using local binary patterns with an application to facial expressions. In *IEEE transactions on pattern analysis and machine intelligence*. IEEE.

Zhou, Z., Zhao, G., and Pietikäinen, M. (2011). Towards a practical lipreading system. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE.