Multimed Tools Appl
https://doi.org/10.1007/s11042-018-6133-z

Facial expression classification using salient pattern driven integrated geometric and textual features

Ruiqi Li[1] · **Jing Tian**[1] · **Matthew Chin Heng Chua**[1]

Received: 24 January 2018 / Revised: 9 April 2018 / Accepted: 9 May 2018

Abstract Facial expression classification aims to recognize human emotion via face images. The major challenge of facial expression classification is how to extract discriminative features from the human face images to differentiate various emotions. To tackle this challenge, a new feature extraction approach is proposed in this paper. The proposed approach defines a new set of salient patterns at the facial keypoint locations. This is in contrast to conventional approaches that either represent the whole face image as a regular grid or use local image patches centered at all facial key point locations. Driven by the proposed salient patterns, both the geometric and textual features are extracted and then concatenated and further incorporated into a machine learning framework to perform facial expression classification. The proposed approach is evaluated in the well-known CK+ benchmark dataset to demonstrate its superior performance.

Keywords Facial expression classification · Geometric features · Textual features · Salient patterns

1 Introduction

Facial expression classification has significant applications in diverse domains, such as health care, driver safety, criminal investigation, and surveillance. Automatic recognizing facial expression is a challenging task and active area in computer vision research [22].

✉ Jing Tian
tianjing@nus.edu.sg

Ruiqi Li
liruiqi@u.nus.edu

Matthew Chin Heng Chua
mattchua@nus.edu.sg

[1] Institute of Systems Science, National University of Singapore, Singapore 119615, Singapore

Published online: 23 May 2018

Facial expression recognition studies in the literature can be classified into two main categories: static images and dynamic image sequences [4, 20]. The first kind of approach comprises only information about the single input image, while the second kind of approach studies temporal information of images to recognize expressions based on more than one frames. Typically, motion information is extracted from sequential expression images, such as the movement distance and direction of feature points; then they are modeled using the optical flow method or statistical method [26]. The focus of this paper is on single static image facial expression recognition in the first category, since it is computationally efficient to use a single image to recognize facial expression [4, 20].

The first challenge of facial expression classification is how to perform feature extraction with the aim to extract discriminative features from the input face images so that various types of facial expression can be recognized [4, 20]. Two popular feature extraction methods are geometric feature (changes in distances between feature points caused by the variety of expression) and textual feature (changes in the local intensity pattern during a facial expression manifestation) [1, 3, 5, 6, 14]. Geometry-based features describe the shape of the face and its components, such as mouth or eyebrow, whereas textual-based features describe the texture of the face caused by expression. For example, Acevedo et al. proposes a geometric feature based on areas and angles of triangles formed by facial landmarks for facial expression recognition [1]. Kotsia et al. uses geometry-based classification approach to predict facial expression [14]. Ghimire et al. proposes to use normalized central moments from specific facial regions that are obtained using incremental search approach [6], which is further extended to address the facial expression video problem [7]. A set of Euclidean distances between the landmarks are used in [5]. The facial angle and the uniform local binary pattern are integrated to form a hybrid features in [3].

The second challenge of facial expression classification is on feature classification, where the objective is to decide the facial expression of the target image by calculating the similarity of the image and many known images, then assign the facial expression label of the most similar image to the target image [4, 20]. A recognition method of elastic graph matching has been proposed to find the most similar person, then the facial expression label of this person is used to recognize the expression [8]. An augmented Gabor feature vector is studied with different similarity measures in [15]. Recently, advanced machine learning techniques, such as deep learning algorithm [19] and deep random forest algorithm [25], have been studied in this domain. Zhang et al. optimizes facial landmark detection together with a bunch of correlated sub-tasks, like head pose estimation and facial attribute inference, then constructs a tasks-constrained deep convolutional neural network to jointly detect facial expressions with a series of related tasks [27]. A 3D convolutional neural networks with deformable action parts constraints is used in [16]. Kim et al. trains several convolutional neural networks as committee members and combine their decisions using a hierarchical architecture of the committee with exponentially-weighted decision fusion [12].

The aforementioned conventional approaches rely on the detected facial keypoint locations to extract features. Typically, they study the whole face region in a regular grid or focus on specific local image patches centered at facial keypoint locations. All keypoint locations are treated equally; however, it overlooks that the impacts of different points imposed to facial expression recognition are different. Motivated by the fact that certain salient pattern features have more contributions in recognizing facial expression, a new feature extraction approach is proposed in this paper. The proposed approach has two key stages. First, a new set of salient patterns is proposed by designing hand-crafted templates at salient facial keypoint locations. Then, steered by the proposed salient patterns, both geometric and textual features are extracted and integrated together to form a set of features. Second, the

concatenated features are further incorporated into a conventional machine learning classifier (say, the conventional support vector machine is used in this paper) to perform facial expression classification.

The rest of this paper is organized as follows. In Section 2, the description of our proposed approach for the facial expression recognition is presented, which includes the proposed salient patterns and the feature extraction of integrated geometric and textual features. The proposed approach is evaluated in Section 3 using benchmark dataset and compared with other state-of-the-art approaches. Finally, Section 4 concludes this paper.

2 Proposed facial expression classification approach

The proposed facial expression classification approach is presented in this section. An overview of the proposed approach is illustrated in Fig. 1. The proposed approach relies on the facial keypoints detected on the face image to extract proposed features. For that, the facial keypoints (landmarks) are detected first using the well-known *Dlib* library [13] to obtain 68 landmarks on the face, such as eyes, eyebrows, nose, mouth and jaw, and assign index to each landmark. It is critical to estimate keypoint positions accurately, since the error incurred in localization would cumulate in the succeeding feature extraction steps. Next, a set of hand-crafted templates are proposed at salient facial keypoint locations. Then, both geometric and textual features are extracted based on the proposed salient patterns. They are finally integrated together into conventional machine learning classifier to recognize facial expression. The details of the proposed approach are presented in following sections, respectively.

2.1 Proposed salient patterns

The physical part of the face remains unchanged despite the variety of different facial expression. Inspired by this, only a subset of local facial keypoints is used for extracting geometric and appearance features for learning facial expressions [24]. For example, the face local region around mouth and eyes are selected as these regions carry the most discriminating information for learning facial expressions.

To address the challenge of the selection of facial keypoints for analyzing facial expression, a set of salient patterns at different locations is proposed in this paper as follows. Fig. 2a shows the mask design based on the outlines of each region and Fig. 2b indicates the corresponding sliding matrices for each salient pattern. As seen from Fig. 2, all of these matrices targeting at different regions are designed using the same size (say, 5×5 used for the purpose of illustration). However, the binary values are different within each matrix according to the region it analyzes. Each landmark locates in the center position of the matrix, which is shown as the red circle in the matrix. Only the pixels labeled as 1 in the

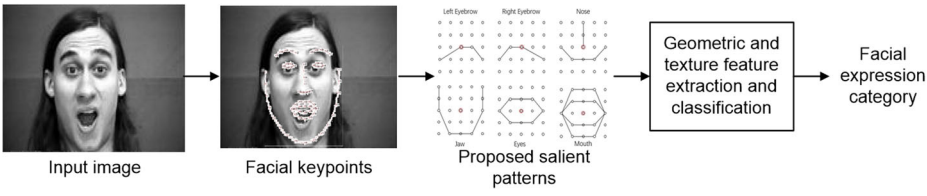


Fig. 1 An overview of the proposed facial expression classification approach

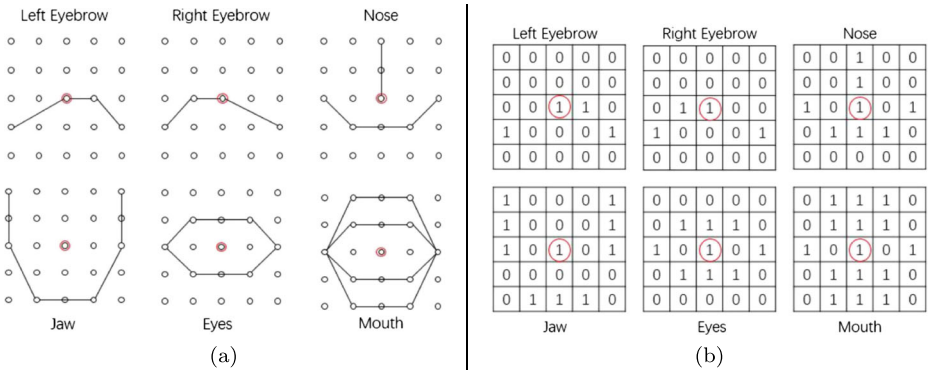
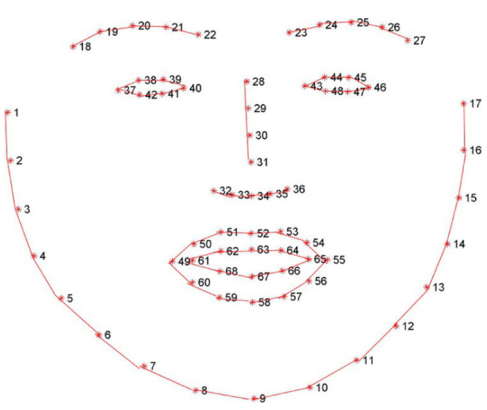


Fig. 2 The proposed salient patterns (**a**) and their respective matrices (**b**). (A color version of this Figure if available in the electronic version)

surrounding pixels are analyzed. For example, for the left eyebrow region, the keypoints included in the left eyebrow area are from the 18-th keypoint to the 22-nd keypoint. The sliding matrix first stays on the 18-th keypoint, which is the red circle in the left eyebrow mask. Assuming that the position coordinate of this keypoint is (x, y), then according to the surrounding 24 values distribution in the left eyebrow sliding matrix, the pixels with $(x-2, y-1)$, (x, y), $(x+1, y)$ and $(x+2, y-1)$ coordinates will be selected and their corresponding pixel values will be extracted. The matrix then continues to slide to the next (say, 19-th) keypoint and repeats the same operation until the pixel value of the 22-nd keypoint is processed.

The contributions of the usage of the propose salient patterns (as illustrated in Fig. 2) are clarified as follows. First, it imposes different contributions of different regions of facial image. On the contrary, the whole face image is studied as a regular grid in the conventional approach, where all facial keypoint locations are treated with the same importance levels. Second, the design of such salient pattern provides a fairly flexible framework to be



(a)

Region	Keypoints index
left eyebrow	$\{18, \ldots, 22\}$
left eye	$\{37, \ldots, 42\}$
right eyebrow	$\{23, \ldots, 27\}$
right eye	$\{43, \ldots, 48\}$
inner mouth	$\{61, \ldots, 60\}$
outer mouth	$\{49, \ldots, 68\}$
jaw	$\{1, \ldots, 17\}$
nasal bridge	$\{28, \ldots, 31\}$
nose	$\{28, \ldots, 36\}$

(b)

Fig. 3 **a** An illustration of geometric features extracted based on keypoint locations; **b** A list of keypoints indices used in calculating various geometric features

extended into other configurations of patterns, although a fixed size of 5×5 is used in the proposed approach and illustrated here.

The proposed approach has fundamental differences with conventional methods [5, 8]. In other geometric feature based models, they use geometric features of all pixels or 68 key point pixels as the features, but in our proposed method, we design a silent feature extraction template to extract geometric features, which demonstrated as the red lines in Fig. 3 of the revised paper. First we separate face regions like: eyebrows, eyes, lip, nose and jaw, then calculate the distance changes between landmarks within one face region, because based on the analysis of the muscle changes on face, when human make different facial expressions, the distance changes between landmarks within one face region are much significant than that in rest regions. The proposed approach not only can extract more important geometric features upon face, but also can reduce a lot of computational time.

2.2 Feature extraction

2.2.1 Geometric feature

With the landmarks coordinates, the next step is to calculate the distance among landmarks. If a pair-wise distance is calculated for all 68 keypoints, then a 68×68 distance matrix is obtained, which not only needs lots computational time but also contains many not-so-important distance features for modeling. Based on the analysis of the muscle changes on face, when human make different facial expressions, the distance changes between landmarks within one face region are much significant than that in rest regions.

Motivated by this, only the distance within each region is used in the proposed approach, which is shown as Fig. 3a, where there are totally 68 keypoints have been extracted and each landmark has its own index. Based on these keypoints, 9 lines are constructed including left eye, left eyebrow, right eye, right eyebrow, nasal nose, nose, jaw, inner mouth, outer mouth. Their respective keypoints indices are summarized in Fig. 3b. Then, the distance-based feature is calculated using Euclidean distance as follows.

Given two keypoints (denoted as l_1 and l_2) and their respective coordinates are denoted as (x_1, y_1) and (x_2, y_2), respectively, the distance $D(\cdot)$ between these two keypoints l_1 and l_2 is first calculated as [17]

$$D(l_1, l_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}. \tag{1}$$

Then, for each red line illustrated in Fig. 3b, its total distance along this line is proposed to be calculated as

$$D_k = \sum_{i,j \in \Omega_k} D(l_i, l_j), \tag{2}$$

in which k is the region index defined as $k = \{1, 2, \ldots, M\}$, M is the number of lines defined in the proposed approach, that is $M = 9$, the distance function $D(\cdot)$ is defined in (1), i and j belong to the set of keypoints Ω_k for the k-th line defined as in Fig. 3b. By such way, totally 9 geometric features are calculated and merged together to form the proposed geometric feature as

$$f_g = \{D_1, D_2, \ldots, D_M\}, \tag{3}$$

which will be integrated with the textual features together for the next feature classification step.

2.2.2 Textual feature

Textual features quantify the perceived texture of an image and give us information about the spatial arrangement of colors or intensities in selected areas of an image. Textual features are represented by the gray-scale distribution of pixels and their neighborhood as the local texture information. Texture features are not simple pixel-based features, they need to be calculated in the region containing multiple pixels. For that, the conventional *local binary patterns* (LBP) technique is used in the proposed approach, due to its computational efficiency. Although, various modified versions of LBP have been used in facial image analysis [9], the original LBP [21] is used in the proposed approach in this paper.

The LBP technique processes the key pixels that are already extracted in the previous section by thresholding neighborhood of each pixel and outputting the result as a series of binary number. To be more specific, it studies the neighborhood at the given pixel location and compares the surrounding pixels with it. If the surrounding pixel is larger than the central key pixel value, then the position of the pixel is marked as 1, otherwise it is marked as 0.

First, given the pixel location with the coordinate (x_i, y_i), its pixel intensity value is denoted as $I(x_i, y_i))$, the LBP code at this location (denoted as $LBP_{R,P}(x_i, y_i))$ is calculated based on its neighborhood, where P is the number of neighbors and R is the radius of the neighborhood,

$$LBP_{P,R}(x_i, y_i) = \sum_{n=0}^{P-1} s(I(x_n, y_n) - I(x_i, y_i)) \cdot 2^n, \tag{4}$$

where (x_n, y_n) is the neighboring pixel centered at the position (x_i, y_i), the function $s(\cdot)$ is the sign function, that means $s(x)$ is 1, if $x \geq 0$ and 0 otherwise. Then, the calculated LBP code goes through histogram processing and normalization by first counting the frequency of each LBP value obtained in each region, then drawing the corresponding histogram to present the distribution of them.

Then, along the red line defined in Fig. 3b, the LBP code at these keypoints are calculated and further construct the histogram feature for the k-th line as

$$H_k = \{h_k(n)\}, \tag{5}$$

where $h_k(n)$ is the n-bin of the histogram, and $h_k(n) = \sum_c I(LBP_{P,R}(x_i, y_i) = n), i = 0, 1, \ldots, 2^P - 1$, $I(A) = 1$ if A is true; otherwise, $I(A) = 0$. Finally, the proposed textual feature is obtained as

$$f_t = \{H_1, H_2, \ldots, H_M\}, \tag{6}$$

where M is the total number of keypoints contained by 9 lines defined in Fig. 3b.

2.3 Feature classification

With the above extracted features, the next step is to build a classifier to classify the facial expressions for face image based on the features. Due to the fairly small size of the dataset used in facial expression classification, the *support vector machines* (SVM) is potential to be used in this domain [10]. The SVM classifiers maximize the hyper plane margin between classes [2], in which the optimal parameter selection is performed based on the grid search strategy. Furthermore, the SVM is mainly designed for binary classification, a multi-layer SVM classifier is built, in which each layer has a binary SVM classifier, based on one-vs-rest strategy, each layer will select one facial expression image as positive label and the rest facial expression image as negative label and 7 layers in total.

To be more specific, the aforementioned geometric feature (defined in (3)) and the textual feature (defined in (6)) are concatenated to form the hybrid features $\mathbf{f} = \{f_i, i = 1, 2, \ldots, N\}$, which have the corresponding classes $y_i = \pm 1$, the classification cost function can be formulated as $g(\mathbf{f}) = sign(\mathbf{w}^T\mathbf{f} + b)$, where $\mathbf{w}^T\mathbf{f} + b = 0$ denotes the optimal classifier, b is the bias, and $\mathbf{w}$ is the weight vector. This optimization problem can be solved by a constrained optimization technique [2].

3 Experimental results

3.1 Experimental setup

To evaluate the performance of the proposed facial expression approach, the benchmark dataset *Extended Cohn-Kanade Dataset*(CK+) [17] is used in the experiments. This database is one of the most comprehensive in the current facial expression research. In this dataset, there are 7 categories of facial expressions, including *anger*, *contempt*, *disgust*, *fear*, *happiness*, *sadness*, *surprise*. It contains 593 sequences recorded by 123 persons. Each sequence consists of images from the neutral frame to the peak expression frame. However, not all sequences have their corresponding facial expression ground truth labels; there are only 327 images have labels. Therefore, only the images with ground truth labels are used to evaluate the performance of various facial expression approach. These images include 45 images for *anger*, 18 image for *contempt*, 59 image for *disgust*, 25 image for *fear*, 69 image for *happiness*, 28 image for *sadness*, 83 image for *surprise*. Several snapshots of this dataset is presented in Fig. 4.

In this paper, we have done two experiments to evaluate the performance of the proposed approach. In both these experiments, we only evaluate the proposed approach with conventional approaches that belong to same geometric category, for a fair performance evaluation. The parameters of the proposed approach are set to be $P = 8$, $R = 3$. In each experiment, the dataset is split into training dataset and test dataset with a ratio of 75 : 25. To provide an objective performance evaluation, two performance criterions: *accuracy* and *F1-score* are used in the experiments. The higher accuracy and/or F1-score value indicate better performance of the facial expression classification approach.

The first one is a simple experiment that uses 6 classes of facial expression including *anger*, *disgust*, *fear*, *happiness*, *sadness*, *surprise* in total 309 images to classify the facial expression. That means, the category *contempt* is excluded in this experiment, as what have

Fig. 4 Several test images of the CK+ dataset [17] used in the experiment

been done in the literature [5, 8, 11, 17]. For this experiment, the proposed approach is compared with two state-of-the-art geometric feature-based models. The first approach utilizes a SVM model based on geometric attributes calculated with the *Point Distribution Model* [5]. The second approach is a one-vs-one classifier by using appearance features of selected facial patches proposed in [8].

The second experiment is a more challenging experiment that include all 7 categories in the experiment. For this experiment, Two widely-spread models as selected as the baseline to be compared with the proposed model, including *Active Appearance Model* approach [17], which is proposed along with the CK+ dataset as a reliable baseline, and the *Conditional Random Field* model, which uses geometric features [1].

3.2 Experimental results

For the first simple experiment, which uses 6 categories of facial expression, the performance of the proposed model and baseline models is shown as Table 1. As seen from Table 1, the proposed model using the integrated geometric and textual features, which uses both distance based features and pixel values based features as the input to the SVM classifier, has achieved 96.10% accuracy and 96.67% F1-score. Compared with the two baselines, the proposed approach has shown superior performance; it has raised 3.89% accuracy and 3.72% F1-score of the model, which only uses distance based features. For further investigating the prediction performance on each facial expression, we also present the confusion matrix performance in Table 2. As seen from Table 2, the proposed approach can achieve a satisfactory performance on each facial expression since all of them has reached over 80% accuracy.

For the second challenging experiment that uses all 7 categories of facial expressions, the performance of the proposed model is shown as Table 3. From the performance comparison presented in Table 3, it can be concluded that the proposed model still achieves the best performance with 86.58% accuracy and 86.22% F1-score and by adding textual features. The performance of model using geometric features is also raised, which shows that textual features can also improve the performance of facial expression classification. The confusion matrix is presented in Table 4. For each facial expression category in Table 4, the accuracy of them has decreased because of the addition of *contempt*. Also, we can see that the accuracy of contempt facial expression even only has 40%, so for the challenging experiment, we can conclude that among all the categories of facial expressions, contempt label is the hardest to be recognized.

The third experiment is to evaluate how the choice of the template size of the proposed approach can affect its performance, we have conducted experiments to compare different size of templates, as well as different type of template in Table 5. From this result, we can see that the recommended choice of 5×5 template size is fairly good, as it achieves higher

Table 1 The performance comparison of various facial expression classification approach

Approach	F1-score	Accuracy
Ref. [5]	–	94.60%
Ref. [8]	94.39%	94.09%
Proposed (geometric only)	92.95%	92.21%
Proposed (texture only)	92.69%	92.21%
Proposed (integrated geometric and texture)	96.67%	96.10%

Table 2 Confusion matrix of the proposed facial expression classification approach for 6-category classification experiment

	Anger	Disgust	Fear	Happiness	Sadness	Surprise
Anger	80%	0	0	0	20%	0
Disgust	7.14%	92.86%	0	0	0	0
Fear	0	0	83.33%	0	16.67%	0
Happiness	0	0	0	100%	0	0
Sadness	0	0	0	0	100%	0
Surprise	0	0	0	0	0	100%

Table 3 The performance comparison of various facial expression classification approach

Approach	F1-score	Accuracy
Ref. [17]	–	83.30%
Ref. [1]	–	86.70%
Proposed (geometric only)	58.14%	64.63%
Proposed (texture only)	83.97%	84.14%
Proposed (integrated geometric and texture)	86.22%	86.58%

Table 4 Confusion matrix of the proposed facial expression classification approach for 7-category classification experiment

	Anger	Contempt	Disgust	Fear	Happiness	Sadness	Surprise
Anger	100%	0	0	0	0	0	0
Contempt	0	40%	0	0	60%	0	0
Disgust	6.67%	0	93.33%	0	0	0	0
Fear	0	0	0	67.14%	14.29%	18.57%	0
Happiness	4.35%	0	0	0	95.65%	0	0
Sadness	0	20%	0	0	0	80%	0
Surprise	0	0	0	4.35%	0	0	95.65%

Table 5 The performance comparison of the proposed approach with different template sizes

Window Size	F1-Score	Accuracy
3*3(No template)	93.39%	93.30%
3*3(Designed template)	94%	87.77%
5*5(No template)	96.56%	96.16%
5*5(Designed template)	96.67%	96.10%
7*7(No template)	96.41%	95.40%
7*7(Designed template)	96.02%	96.64%

Table 6 The performance evaluation of the proposed approach using the JAFFE dataset [18]

	F1-Score	Accuracy
Proposed (texture only)	0.86	0.87
Proposed (geometric only)	0.86	0.86
Proposed (integrated geometric and texture)	0.88	0.91

F1-score than that using ordinary window or designed template with 3×3 or 7×7 window sizes, respectively.

The fourth experiment is conducted using the *Japanese Female Facial Expression* (JAFFE) dataset [18] to evaluate the performance of the proposed approach. The JAFFE database contains 213 images of 7 facial expressions (6 basic facial expressions + 1 neutral) posed by 10 Japanese female models. Each image has been rated on 6 emotion adjectives by 60 Japanese subjects. In our experiment, we use the 6 basic facial expressions as the labels to do the classification work using our model. The emotion label of JAFFE is presented as the semantic ratings. A 5 level scale was used for each of the 6 adjectives (5-high, 1-low). To make the classification result more reliable and more in line with the actual situation, we select the emotions of top two high score as the label to evaluate the accuracy. The result of the proposed approach is shown in Table 6. From this Table we can see that the proposed model has a satisfactory performance on JAFFE dataset, furthermore, using the integrated geometric and texture features as the training data still achieved the highest F1-score 0.88 and accuracy 0.91.

The last experiment is to evaluate the execution speed of the proposed approach. The proposed approach is implemented using Python programming language and run on a PC with an Intel Core i7 3.40 GHz CPU and an 8 GB RAM. The average processing time of the proposed feature extraction approach is 3.34 seconds to process a frame with a resolution of 480×640 pixels, which includes the face detection, facial key point detection, and the proposed feature extraction. Therefore, the proposed approach is potential to be used in real-time applications.

4 Discussion and conclusion

The proposed approach is potential to be applied in a few application areas, since it provides a flexible framework to incorporate other features. Furthermore, the proposed salient pattern is potential to facial video classification, where multiple frames are considered and the temporal profile of feature changes are needed [26].

The limitation of the proposed approach lie in following aspects. First, the proposed salient pattern has a fixed-size pattern. The choice of the pattern size needs further study to be adaptive to the detected facial key points. Second, the classification module of the proposed approach uses a conventional SVM method. Other state-of-the-art machine learning method, such as deep features and convolutional neural network, could be considered to be incorporate into the proposed framework. Lastly, further experiments using different datasets [23] are needed to verify the performance of the proposed approach in a wider and more extensive scenario.

In conclusion, a novel facial expression recognition approach has been proposed in this paper by using the integrated geometric and texture features, which are driven by the

proposed salient patterns. The proposed new features have been incorporated into a machine learning framework to perform facial expression classification, which is compared with conventional geometric-based approaches in the benchmark dataset.

References

1. Acevedo D, Negri P, Buemi ME, Fernández FG, Mejail M (2017) A simple geometric-based descriptor for facial expression recognition. In: IEEE Int Conf on automatic face gesture recognition, pp 802–808
2. Chang C-C, Lin C-J (2011) LIBSVM: A library for support vector machines. ACM Trans Intell Syst Technol 2(3):1–17
3. Datta S, Sen D, Balasubramanian R (2017) Integrating geometric and textural features for facial emotion classification using SVM frameworks. In: Int Conf on computer vision and image processing, Singapore, pp 619–628
4. Deshmukh S, Patwardhan M, Mahajan A (2016) Survey on real-time facial expression recognition techniques. IET Biom 5(3):155–163
5. Fernandes JA, Matos LN, Aragao MGS (2016) Geometrical approaches for facial expression recognition using support vector machines. In: Int Conf on graphics, patterns and images Sao Paulo, Brazil, pp 347–354
6. Ghimire D, Jeong S, Lee J, Park SH (2017) Facial expression recognition based on local region specific features and support vector machines. Multimedia Tools and Applications 76(6):7803–7821
7. Ghimire D, Lee J, Li Z-N, Jeong S (2017) Recognition of facial expressions based on salient geometric features and support vector machines. Multimedia Tools and Applications 76(6):7921–7946
8. Happy SL, Routray A (2015) Automatic facial expression recognition using features of salient facial patches. IEEE Trans Affect Comput 6(1):1–12
9. Huang D, Shan C, Ardabilian M, Wang Y, Chen L (2011) Local binary patterns and its application to facial image analysis: A survey. IEEE Trans Syst Man Cybern Part C Appl Rev 41(6):765–781
10. Hsieh C-C, Hsih M-H, Jiang M-K, Cheng Y-M, Liang E-H (2016) Effective semantic features for facial expressions recognition using SVM. Multimedia Tools and Applications 75(11):6663–6682
11. Jain S, Hu C, Aggarwal JK (2011) Facial expression recognition with temporal modeling of shapes. In: IEEE Int Conf on computer vision workshops, pp 1642–1649
12. Kim B-K, Roh J, Dong S-Y, Lee S-Y (2016) Hierarchical committee of deep convolutional neural networks for robust facial expression recognition. Journal on Multimodal User Interfaces 2:173–189
13. King DE (2009) Dlib-ml: a machine learning toolkit. J Mach Learn Res 10:1755–1758
14. Kotsia I, Pitas I (2007) Facial expression recognition in image sequences using geometric deformation features and support vector machines. IEEE Trans Image Process 16(1):172–187
15. Liu C, Wechsler H (2002) Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. IEEE Trans Image Process 11(4):467–476
16. Liu M, Li S, Shan S, Wang R, Chen X (2015) Deeply learning deformable facial action parts model for dynamic expression analysis. In: Asian Conf on computer vision, Singapore, pp 143–157
17. Lucey P, Cohn JF, Kanade T, Saragih J, Ambadar Z, Matthews I (2010) The extended cohn-kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression. In: IEEE Int. Conf. on computer vision and pattern recognition, San Francisco, pp 94–101
18. Lyons MJ, Budynek J, Akamatsu S (1999) Automatic classification of single facial images. IEEE Trans Pattern Ana Mach Intell 21(12):1357–1362
19. Majumder A, Behera L, Subramanian V (2018) Automatic facial expression recognition system using deep network-based data fusion. IEEE Trans on Cybernetics 28(1):103–114
20. Martinez B, Valstar MF, Jiang B, Pantic M (2017) Automatic analysis of facial actions: A survey. IEEE Trans. on Affective Computing. accepted
21. Ojala T, Pietikainen M, Maenpaa T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell 24(7):971–987
22. Sariyanidi E, Gunes H, Cavallaro A (2015) Automatic analysis of facial affect: a survey of registration, representation, and recognition. IEEE Trans Pattern Anal Mach Intell 37(6):1113–1133
23. Siddiqi MH, Ali M, Abdelrahman ME, Khan A, Banos O, Khan AM, Lee S, Choo H (2018) Evaluating real-life performance of the state-of-the-art in facial expression recognition using a novel YouTube-based datasets. Multimedia Tools and Applications 77(1):917–937

24. Takalkar M, Xu M, Wu Q, Chaczko Z (2017) A survey: Facial micro-expression recognition. Multimedia Tools and Applications :1–25. accepted
25. Yang B, Cao J-M, Jiang D-P, Lv J-D (2017) Facial expression recognition based on dual-feature fusion and improved random forest classifier. Multimedia Tools and Applications :1–23. accepted
26. Yu J, Wang Z (2017) A video-based facial motion tracking and expression recognition system. Multimedia Tools and Applications 76(13):14,653–14,672
27. Zhang Z, Luo P, Loy CC, Tang X (2014) Facial landmark detection by deep multi-task learning. In: European Conf on computer vision, Zurich, Switzerland, pp 94–108

Ruiqi Li is currently pursuing her Master of Technology(Knowledge Engineering) at the Institute of System Science, NUS since 2017. Her area of interest is in Artificial Intelligence and Robotics. She currently concentrates on researches of computer vision and natural language processing based on machine learning algorithms.

Dr. Jing Tian received the PhD degree from Nanyang Technological University, Singapore. Currently, he lectures in Institute of Systems Science at the National University of Singapore, in the areas of artificial intelligence, computer vision, and machine learning.

Dr. Matthew Chin Heng Chua is currently a lecturer and principal investigator at the NUS Institute of System Science where he is spearheading the Medical & Cybernetics Systems. He is overseeing the research programme in Smart Healthcare, Artificial Intelligence and Advanced Robotics. He has won prestigious research grants from the National Medical Research Council (NMRC) for his work in intelligent medical devices. His wide array of expertise makes him a highly sort after collaborator from various industries.