Facial emotion recognition using Min-Max similarity classifier

Alex Pappachen James Corresp., [1] , **Olga Krestinskaya** [1]

[1] School of Engineering, Nazarbayev University, Astana, Kazakhstan

Corresponding Author: Alex Pappachen James
Email address: apj@ieee.org

Recognition of human emotions from the imaging templates is useful in a wide variety of human-computer interaction and intelligent systems applications. However, the automatic recognition of facial expressions using image template matching techniques suffer from the natural variability with facial features and recording conditions. In spite of the progress achieved in facial emotion recognition in recent years, the effective and computationally simple feature selection and classification technique for emotion recognition is still an open problem. In this paper, we propose an efficient and straightforward facial emotion recognition algorithm to reduce the problem of inter-class pixel mismatch during classification. The proposed method includes the application of pixel normalization to remove intensity offsets followed-up with a Min-Max metric in a nearest neighbor classifier that is capable of suppressing feature outliers. The results indicate an improvement of recognition performance from 92.85% to 98.57% for the proposed Min-Max classification method when tested on JAFFE database. The proposed emotion recognition technique outperforms the existing template matching methods.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2794v1 | CC BY 4.0 Open Access | rec: 8 Feb 2017, publ: 8 Feb 2017

Facial Emotion Recognition using Min-Max Similarity Classifier

Olga Krestinskaya[1] and Alex Pappachen James[1]

[1]**School of Engineering, Nazarbayev University**

Corresponding author:
Alex James[1]

Email address: apj@ieee.org

ABSTRACT

Recognition of human emotions from the imaging templates is useful in a wide variety of human-computer interaction and intelligent systems applications. However, the automatic recognition of facial expressions using image template matching techniques suffer from the natural variability with facial features and recording conditions. In spite of the progress achieved in facial emotion recognition in recent years, the effective and computationally simple feature selection and classification technique for emotion recognition is still an open problem. In this paper, we propose an efficient and straightforward facial emotion recognition algorithm to reduce the problem of inter-class pixel mismatch during classification. The proposed method includes the application of pixel normalization to remove intensity offsets followed-up with a Min-Max metric in a nearest neighbor classifier that is capable of suppressing feature outliers. The results indicate an improvement of recognition performance from 92.85% to 98.57% for the proposed Min-Max classification method when tested on JAFFE database. The proposed emotion recognition technique outperforms the existing template matching methods.

INTRODUCTION

In recent years, the human-computer interaction challenge has led to the demand to introduce efficient facial and speech recognition systems Chao et al. (2015); Danisman et al. (2013); Ververidis and Kotropoulos (2008); Li et al. (2015); Gupta et al. (2015). Facial emotion recognition is the identification of a human emotion based on the facial expression and mimics Zhang et al. (2016). The facial emotion recognition has a wide range of appliction prospects in different areas, such as medicine Zhao et al. (2014), robotics Shih et al. (2008); P. et al. (2015), computer vision, surveillance systems Chao et al. (2015), machine learning Cruz et al. (2014), artificial intelligence, communication Latif et al. (2015); Sudha et al. (2015), psychological studies Li et al. (2015), smart vehicles P. et al. (2015), security and embedded systems Sun and An (2010).

There are two main approaches for facial expression recognition: geometry-based and appearance-based methods Danisman et al. (2013). The geometry-based methods extract main feature points and their shapes from the face image and calculate the distances between them. While, appearance-based methods focus on the face texture using different classification and template matching methods Chiranjeevi et al. (2015); Ghimire and Lee (2013). In this paper, we focus on facial emotion recognition based on template matching techniques that remains a challenging task Brunelli and Poggio (1993); Zhang et al. (2011); Wang et al. (2014). Since the orientation of pixel features are sensitive to the changes in illumination, pose, scale and other natural imaging variabilities, the matching errors tend to be high Li et al. (2015); Tang et al. (2013); Zhao et al. (2016). Pixel matching methods are known to be useful when the images has missing features because imaging matrices become sparse and feature computation process is not trivial. As facial expressions cause a mismatch of intra-class features due to their orientation variability, it is difficult to map them between the imaging templates.

Facial emotion recognition accuracy depends on the robustness of a feature extraction method to intra-class variations and classifier performance under noisy conditions and with various types of occlusions Cruz et al. (2014). Even thought a variety of approaches for the automated recognition of human expressions from the face images using template matching methods have been investigated and proposed

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2794v1 | CC BY 4.0 Open Access | rec: 8 Feb 2017, publ: 8 Feb 2017

over the last few years Chiranjeevi et al. (2015), the emotion recognition method with robust feature extraction and effective classification techniques accompanied by low computational complexity is still an open research problem Kamarol et al. (2016). Therefore, in this paper, we address the issues of matching templates through pixel normalization followed by the removal of inter-image feature outliers using a Min-Max similarity metric. We apply Gaussian normalization method with local mean and standard deviation to normalize the pixels and extract relevant face features followed by Min-Max classification method to determine an emotion class. The simulation is performed in Matlab for the Japanese Female Facial Expression (JAFFE) database Lyons et al. (1998) and the emotion recognition accuracy is calculated using leave-one-out cross-validation method.

The main contributions of this work are the following:

- We develop a simplified approach for facial emotion recognition with template matching method using Gaussian normalization, mean and standard deviation based feature extraction and Min-Max classification approach.
- We present simple and effective facial emotion recognition algorithm having low computational complexity and ability to suppress the outliers and remove intensity offsets.
- We conduct the experiments and simulations on JAFFE database to demonstrate the efficiency of the proposed approach and highlight its advantages, comparing to the other existing methods.

The paper is organized as follows. Section presents the overview of the existing methods for facial emotion recognition, their drawbacks and reasons to propose a new method. In Section , we show normalization, feature extraction and classification parts of the proposed method, present the algorithm and describe the experiments. Section contains the simulation results and comparison of the obtained results with the existing methods. In Section 1, we discuss the benefits and drawbacks of the proposed method, in comparison to the traditional methods. Section 1 concludes the paper.

BACKGROUND AND RELATED WORKS

To address the problem of facial emotion recognition, several template matching methods have been proposed in the last decades Chao et al. (2015); Shih et al. (2008); Poursaberi et al. (2012); Cheng et al. (2010); Kamal et al. (2016a). In most of the cases, the process of emotion recognition from human face images is divided into two main stages: feature extraction and classification Chao et al. (2015); Shih et al. (2008). The main aim of feature extraction methods is to minimize intra-class variations and maximize inter-class variations. The most important facial elements for human emotion recognition are eyes, eyebrows, nose, mouth and skin texture. Therefore, a vast majority of feature extraction methods focus on these features Danisman et al. (2013); Fasel and Luettin (2003). The selection of irrelevant face image features or insufficient number of them would lead to low emotion recognition accuracy, even applying effective classification methods Kamarol et al. (2016). The main purpose of the classification part is to differentiate the elements of different emotion classes to enhance emotion recognition accuracy.

The commonly used feature extraction methods include two-dimensional Linear Discriminant Analysis (2D-LDA) Shih et al. (2008); Kamal et al. (2016a), two-dimensional Principle Component Analysis (2D-PCA) Rajendran et al. (2014), Discrete Wavelet Transform (DWT) Basu et al. (2015); Shih et al. (2008); Zhang et al. (2016), Gabor based methods Hegde et al. (2016); Zhang et al. (2014) and wavelets-based techniques Poursaberi et al. (2012, 2013). In 2D-LDA method, the two-dimensional image matrix is exploited to form scatter matrices between the classes and within the class Shih et al. (2008). 2D-LDA method can be applied for facial features extraction alone or accompanied with the other feature extraction method, as DWT Shih et al. (2008); Kamal et al. (2016a). In 2D-PCA feature extraction method, the covariance matrix representation of the image is derived directly from the original image Shih et al. (2008); Rajendran et al. (2014). The size of the derived principle component matrix is smaller than the original image size that allows to decrease the amount of processing data, and consequently, reduce the required computational memory Marvadi et al. (2015). However, 2D-LDA and 2D-PCA methods applied in template matching techniques require an additional processing of the image, dimensionality reduction techniques or application of another feature extraction method to achieve higher recognition accuracy, which leads to the increase in processing time.

The other feature extraction method is DWT. This method is based on the low-pass and high-pass filtering, therefore, it is appropriate for the images with different resolution levels Shih et al. (2008). In

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2794v1 | CC BY 4.0 Open Access | rec: 8 Feb 2017, publ: 8 Feb 2017

the emotion recognition task, DWT is applied for the extraction of useful features from the face images and can be replaced with its Orthogonal Wavelet Transform (OWT) and Biorthogonal Wavelet Transform (BWT) having the advantages of orthogonality Zhang et al. (2016). Another method for facial emotion recognition is Gauss-Laguerre wavelet geometry based technique. This method represents the processed image in polar coordinates with the center at a particular pivot point. The degree of freedom is one of the advantages that Gauss-Laguerre approach provides, which in turn allows to extract of features of the desirable frequency from the images Poursaberi et al. (2012, 2013). However, DWT and Gauss-Laguerre approaches are complex and require time and memory consuming calculations.

The classification of the extracted features can be implemented using Support Vector Machine (SVM) algorithm Shih et al. (2008); Basu et al. (2015), K-Nearest Neighbor (KNN) method Poursaberi et al. (2012); Kamal et al. (2016b), Random Forest classification method Zhao et al. (2014); Wei et al. (2016) and Gaussian process Cheng et al. (2010). The SVM principle is based on non-linear mapping and identification of a hyperplane for the separation of data classes. SVM classifier is used with the application of different kernel functions, such as linear, quadratic, polynomial and radial basis functions, to optimize the SVM performance Shih et al. (2008); Basu et al. (2015). KNN approach is based on the numerical comparison of a testing data sample with training data samples of each class followed by the determination of the similarity scores. The data class is defined by K most similar data samples based on the minimum difference between train and test data Poursaberi et al. (2012); Kamal et al. (2016b). KNN and SVM classifiers are simple and widely used for emotion recognition, however these classifiers do not suppress the outliers that leads to lower recognition accuracy.

The Random Forest classification method is based on the decision making tree approach with randomized parameters Zhao et al. (2014); Wei et al. (2016). To construct the decision tree, Random Forest algorithm takes a random set of options and selects the most suitable from them Hariharan and Adam (2015). Random Forest classifier is robust and has a high recognition rate for the images of large resolution Jia et al. (2016). The drawback of Random Forest classifier is its computational complexity. The other classification method is the Gaussian process approach. Gaussian process is based on the predicted probabilities and can be used for facial emotion recognition without application of feature selection algorithms. The Gaussian process allows a simplified computational approach, however has a smaller emotion recognition rate, comparing to the other methods Cheng et al. (2010).

Even thought the a number of methods for feature extraction and classification have been proposed, there is a lack of template matching methods that allow to achieve high recognition accuracy with minimum computational cost. Therefore, the method that we propose has a potential to reduce the computational complexity of facial emotion recognition operation and increase the recognition accuracy due to the simplicity of the algorithm, effective feature extraction and ability to suppress outliers. The proposed algorithm can be implemented using small computational capacity devices keeping facial emotion recognition operation fast and accurate.

METHODOLOGY

The main idea of the proposed method is to extract the spatial change of standardized pixels in a face image and detect the emotion class of the face using a Min-Max similarity Nearest Neighbor classier. The images from the JAFFE database Lyons et al. (1998) are used for the experiments. This database contains 213 images of 10 female faces comprising 6 basic facial expressions and neutral faces. The original images from the database have the size of 256×256 pixels and in our experiments they are cropped to a size of 101×114 pixels retaining only the relevant information of the face area. A block diagram of the proposed method is shown in Fig. 1.

Pre-processing

Illumination variability introduces the inter-class feature mismatch resulting in the inaccuracies in the detection of emotion discriminating features from the face images. Therefore, image normalization is essential to reduce the inter-class feature mismatch that can be viewed as intensity offsets. Since the intensity offsets are uniform within a local region, we perform Gaussian normalization using local mean and standard deviation. The input image is represented as $x(i,j)$, and $y(i,j)$ is the normalized output image, where i and j are the row and column number of the processed image. The normalized output image is calculated by Eq. 1, where μ is a local mean and σ is a local standard deviation computed over a window of $N \times N$ size.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2794v1 | CC BY 4.0 Open Access | rec: 8 Feb 2017, publ: 8 Feb 2017

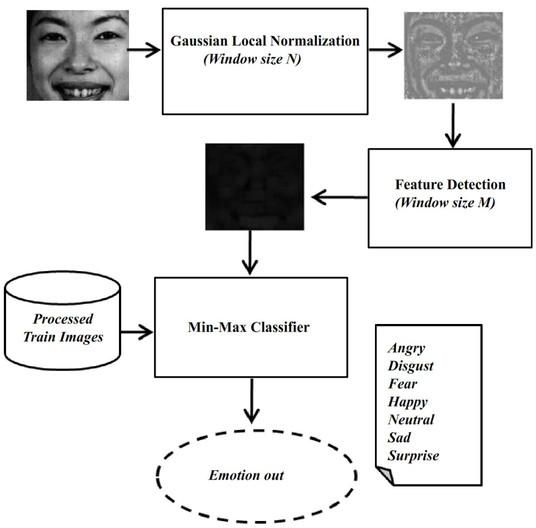


Figure 1. Outline of the proposed emotion recognition system

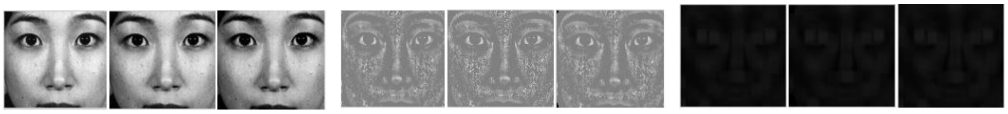
$$y(i,j) = \frac{x(i,j) - \mu(i,j)}{6\sigma(i,j)} \tag{1}$$

The parameters μ and σ are calculated using Eq. 2 and 3 with $a = (N-1)/2$.

$$\mu(i,j) = \frac{1}{N^2} \sum_{k=-a}^{a} \sum_{h=-a}^{a} x(k+i, h+j) \tag{2}$$

$$\sigma(i,j) = \sqrt{\frac{1}{N^2} \sum_{k=-a}^{a} \sum_{h=-a}^{a} [x(k+i, h+j) - \mu(i,j)]^2} \tag{3}$$

An example image from the JAFFE database with three different lighting conditions is shown in Fig. 2 (a). As the JAFFE database does not contain the face images with different illumination conditions, the illumination change was created by adding and subtracting the value of 10 from the original image. Fig. 2 (b) illustrates the respective images after Gaussian local normalization. Irrespective of the illumination conditions, the three locally normalized images appear similarly with the minimum pixel intensity variation.



(a) (b) (c)

Figure 2. (a) Sample image from JAFFE database with different lighting conditions obtained by adding and subtracting a value of 10 from the original image.(b) Normalized images of above sample images obtained by performing Gaussian normalization using local mean and local standard deviation taken over a window of size N=11. (c) Feature detected images from the normalized image by performing local standard deviation using a window of size M=11.

Feature detection

The feature parts useful for the facial emotion recognition are eyes, eyebrows, cheeks and mouth regions. In this experiment, we perform the feature detection by calculating local standard deviation of normalized image using a window of M×M size. Eq. 4 is applied for the feature detection with $b = (M-1)/2$.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2794v1 | CC BY 4.0 Open Access | rec: 8 Feb 2017, publ: 8 Feb 2017

$$w(i,j) = \sqrt{\frac{1}{M^2}\sum_{k=-b}^{b}\sum_{h=-b}^{b}[y(k+i,h+j)-\mu'(i,j)]^2} \tag{4}$$

In Eq. 4 the parameter μ' refers to the mean of the normalized image $y(i,j)$ and can be calculated by Eq. 5.

$$\mu'(i,j) = \frac{1}{M^2}\sum_{k=-b}^{b}\sum_{h=-b}^{b} y(k+i,h+j) \tag{5}$$

Fig. 2 (c) shows the results of feature detection corresponding to the normalized images.

Emotion Classification

For the recognition stage, we propose a Min-Max similarity metric in a Nearest Neighbor classifier framework. This method is based on the principle that the ratio of the minimum difference to the maximum difference of two pixels will produce a unity output for equal pixels and an output less than unity for unequal pixels. The proposed method is described in Algorithm 1. The algorithm parameter $trainlen$ refers to the number of train images, N corresponds to the normalization window size, and M indicates the feature detection window size. Each cropped image is of $m \times n$ pixel dimension. The parameter *train* is a feature array of $trainlen \times (m \times n)$ size, where each row corresponds to the processed train images. After normalization and feature detection, test images are stored in to a vector *test* of $1 \times (m \times n)$ size. A single test image is compared pixel-wise with processed train images of all the classes in the feature array using the proposed Min-Max classifier:

$$s(i,j) = [\frac{min[train(i,j), test(1,j)]}{max[train(i,j), test(1,j)]}]^{\alpha}, \tag{6}$$

where a parameter α controls the power of exponential to suppress the outlier similarities. Outliers are the observations that come inconsistent with the remaining observations and are common in real-time image processing. The presence of outliers may cause misclassification of an emotion, since sample maximum and sample minimum are maximally sensitive to them. In order to remove the effects of the outliers, $\alpha = 3$ is selected to introduce the maximum difference to the inter-class images and minimum difference to the intra-class images.

After Min-Max classification, a column vector z of $trainlen \times 1$ size containing the weights obtained after comparing the test image with each of the *trainlen* number of train images is calculated using Eq. 7.

$$z(i) = \sum_{j=1}^{m \times n} s(i,j) \tag{7}$$

The classification output *out* shown in Eq. 8 is the maximum value of z corresponding to the train image that shows the maximum match. The recognized emotion class is the class of the matched train image.

$$out = max(z(i)) \tag{8}$$

RESULTS AND COMPARISON

To benchmark the performance of the proposed algorithm, leave-one-out cross-validation method has been used. In this method, one image of each expression of each person is applied for testing and the remaining images are used for training Poursaberi et al. (2012). The cross-validation is repeated 30 times to obtain a statistically stable performance of the recognition system and to ensure that all the images in

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2794v1 | CC BY 4.0 Open Access | rec: 8 Feb 2017, publ: 8 Feb 2017

Algorithm 1 Emotion Recognition using Min-Max classifier

Require: Test image Y,Train images X_t,trainlen,window size N and M
Crop the images to a dimension of $m \times n$
for $t = 1$ to $trainlen$ **do**
  $C(i,j) = \frac{X_t(i,j)-\mu(i,j)}{6\sigma(i,j)}$
  $W(i,j) = \sqrt{\frac{1}{M^2}\sum_{k=-b}^{b}\sum_{h=-b}^{b}[C(k+i,h+j)-\mu(i,j)]^2}$
  Store the value of W to an array $train$ of dimension $trainlen \times m \times n$
end for
for $t = 1$ to $trainlen$ **do**
  $V(i,j) = \frac{Y(i,j)-\mu(i,j)}{6\sigma(i,j)}$
  $test(i,j) = \sqrt{\frac{1}{M^2}\sum_{k=-b}^{b}\sum_{h=-b}^{b}[V(k+i,h+j)-\mu(i,j)]^2}$
  $s(t,j) = [\frac{min[train(t,j),test(1,j)]}{max[train(t,j),test(1,j)]}]^3$
  $z(t) = \sum_{j=1}^{m\times n} s(t,j)$
  $out = max(z(t))$
end for

JAFFE database are used for testing at least once. The overall emotion recognition accuracy of the system is obtained by averaging the results of the cross-validation trials. Fig. 3 shows the different accuracy rates obtained for each trial by varying feature detection window size M from 3 to 21 and keeping normalization window size at $N = 11$. It is shown that the maximum emotion recognition accuracy this normalization window size can be achieved with the detection window size of 11.

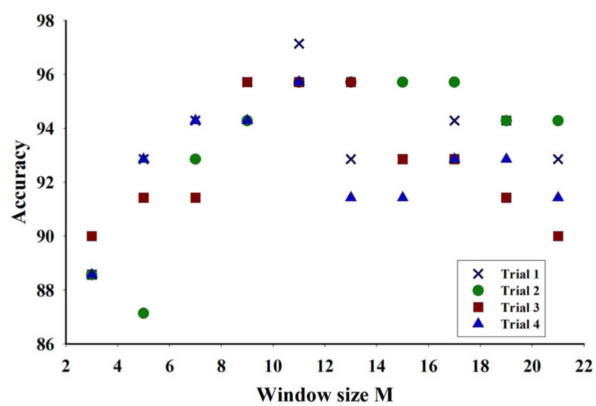


Figure 3. The accuracy rates obtained for four trials of leave-one-out cross-validation method for different feature detection window size M ranging from 3 to 21.

Fig. 4 shows the recognition accuracy rates obtained for different normalization and feature detection window sizes ranging from 3 to 21 for a single trial.

To evaluate the performance of the proposed Min-Max classifier, it has been compared with the other classifiers, such as Nearest NeighborSaha and Wu (2010) and Random ForestLiaw and Wiener (2002), after normalization and feature detection. The obtained accuracy values are shown in Table 1.

Table 1. Testing of feature detected images on other classifiers

Classifier	*Accuracy*(%)
Nearest Neighbor	92.85
Random Forest	88.57
Proposed Min-Max classifier	**98.57**

The proposed method achieves a maximum accuracy of 98.57% for a window size of $M = N = 11$

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2794v1 | CC BY 4.0 Open Access | rec: 8 Feb 2017, publ: 8 Feb 2017

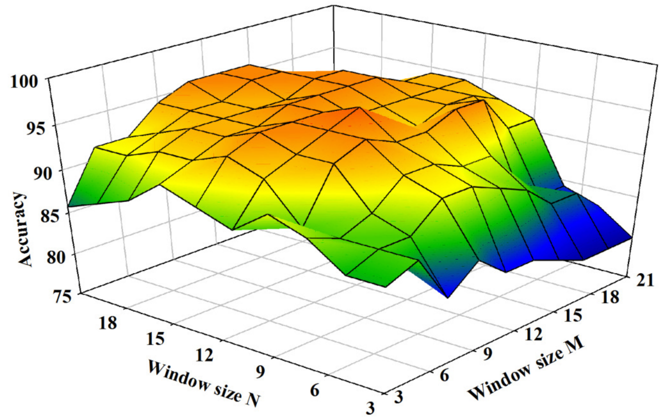


Figure 4. Graph shows the accuracy rates obtained for different normalization and feature detection window sizes ranging from M=N=3 to 21 for one trial of leave-one-out cross-validation method.

which outperforms the other existing methods in the literature for leave-one-out cross-validation method. Table 2 shows the performance comparison of the proposed method with the other existing systems on the same database using leave-one-out cross-validation method. Applying the proposed method, we achieved emotion recognition accuracy of 98.57% proving the significance of emotion detection method with Min-Max similarity classifier.

Table 2. Comparison of proposed emotion recognition system with other existing system based on leave-one-out cross-validation method

Existing systems	*Method used*	*Accuracy(%)*
Cheng et. al Cheng et al. (2010)	Gaussian Process	93.43
Hamester et. al Hamester et al. (2015)	Convolutional Neural Network	95.80
Frank et. al Shih et al. (2008)	DWT + 2D-LDA +SVM	95.71
Poursaberi et. al Poursaberi et al. (2012)	Gauss Laguerre wavelet+ KNN	96.71
Hegde et. al Hegde et al. (2016)	Gabor and geometry based features	97.14
Proposed method	**Min-Max classifier**	**98.57**

In addition, comparing to the proposed emotion recognition system, the other existing methods require specialized feature extraction and dimensionality reduction techniques before classification stage. The main advantages of the proposed emotion recognition system are its simplicity and straightforwardness.

1 DISCUSSION

The main advantages of the proposed facial motion recognition approach are high recognition accuracy and low computational complexity. To achieve high recognition accuracy, the effective feature selection is required. In the existing methods, the complex algorithms for feature selection are applied without normalizing the image. The normalization stage is important and has a considerable effect on the accuracy of the feature selection process. In the proposed algorithm, we apply simple pre-processing methods to normalize the images and eliminate intensity offsets that effects the accuracy of the feature selection process and leads to the increase of emotion recognition accuracy, in comparison to the existing methods. The effect of the proposed Min-Max classification method on recognition accuracy is also important. Table 1 shows the application of the other classification method for the same approach where proposed Min-Max classifier illustrates the performance improvement. In comparison to the existing method, the proposed Min-Max classifier has an ability to suppress the outliers that significantly impacts overall performance of this approach.

The simplicity and straightforwardness of the proposed approach are also important due to the resultant low computational complexity. Most of the existing methods use complicated feature extraction and classification approaches that double the complexity of the facial recognition process and require the device with large computational capacity to process the images. We address this problem applying direct local mean and standard deviation based feature detection methods and simple Min-Max classification method. In comparison to the existing feature detection methods, such as PCA Rajendran et al. (2014) and LDA Shih et al. (2008), the proposed method is straightforward and does not require dimensionality reduction. Moreover, simple Min-Max classification method also reduce the computational time, comparing to the traditional classification approaches, such as SVM Basu et al. (2015), KNN Kamal et al. (2016b), Gaussian process Cheng et al. (2010) and Neural Network Dailey et al. (2002). Therefore, the algorithm can be run on the device with low computational capacity.

CONCLUSION

In this paper, we have represented the approach to improve the performance of emotion recognition task using template matching method. We have demonstrated that the pixel normalization and feature extraction based on local mean and standard deviation followed up by the Mix-Max similarity classification can result in the improvement of overall classification rates. We achieved emotion recognition accuracy of 98.57% that exceeds the performance of the existing methods for the JAFFE database for leave-one-out cross-validation method. The capability of the algorithm to suppress feature outliers and remove intensity offsets results in the increase of emotion recognition accuracy. Moreover, the proposed method is simple and direct, in comparison to the other existing methods requiring the application of dimensionality reduction techniques and complex classification methods for computation and analysis. Low computational complexity is a noticeable benefit of the proposed algorithm that implies the reduction of computational time and required memory space. This method can be extended to the other template matching problems, such as face recognition and biometric matching. The drawback of the proposed method, as in any other template matching method, is the metric learning requiring to create the templates for each class that, in turn, consumes additional memory space to store the templates.

REFERENCES

Basu, A., Routray, A., Shit, S., and Deb, A. K. (2015). Human emotion recognition from facial thermal image based on fused statistical feature and multi-class svm. In *2015 Annual IEEE India Conference (INDICON)*, pages 1–5.

Brunelli, R. and Poggio, T. (1993). Face recognition: features versus templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(10):1042–1052.

Chao, W.-L., Ding, J.-J., and Liu, J.-Z. (2015). Facial expression recognition based on improved local binary pattern and class-regularized locality preserving projection. *Signal Processing*, 117:1 – 10.

Cheng, F., Yu, J., and Xiong, H. (2010). Facial expression recognition in jaffe dataset based on gaussian process classification. *IEEE Transactions on Neural Networks*, 21(10):1685–1690.

Chiranjeevi, P., Gopalakrishnan, V., and Moogi, P. (2015). Neutral face classification using personalized appearance models for fast and robust emotion detection. *IEEE Transactions on Image Processing*, 24(9):2701–2711.

Cruz, A. C., Bhanu, B., and Thakoor, N. S. (2014). Vision and attention theory based sampling for continuous facial emotion recognition. *IEEE Transactions on Affective Computing*, 5(4):418–431.

Dailey, M. N., Cottrell, G. W., Padgett, C., and Adolphs, R. (2002). Empath: A neural network that categorizes facial expressions. *Journal of cognitive neuroscience*, 14(8):1158–1173.

Danisman, T., Bilasco, I. M., Martinet, J., and Djeraba, C. (2013). Intelligent pixels of interest selection with application to facial expression recognition using multilayer perceptron. *Signal Processing*, 93(6):1547 – 1556. Special issue on Machine Learning in Intelligent Image Processing.

Fasel, B. and Luettin, J. (2003). Automatic facial expression analysis: a survey. *Pattern recognition*, 36(1):259–275.

Ghimire, D. and Lee, J. (2013). Geometric feature-based facial expression recognition in image sequences using multi-class adaboost and support vector machines. *Sensors*, 13(6):7714–7734.

Gupta, S., Mehra, A., et al. (2015). Speech emotion recognition using svm with thresholding fusion. In *Signal Processing and Integrated Networks (SPIN), 2015 2nd International Conference on*, pages 570–574. IEEE.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2794v1 | CC BY 4.0 Open Access | rec: 8 Feb 2017, publ: 8 Feb 2017

Hamester, D., Barros, P., and Wermter, S. (2015). Face expression recognition with a 2-channel convolutional neural network. In *2015 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8.
Hariharan, A. and Adam, M. T. P. (2015). Blended emotion detection for decision support. *IEEE Transactions on Human-Machine Systems*, 45(4):510–517.
Hegde, G., Seetha, M., and Hegde, N. (2016). Kernel locality preserving symmetrical weighted fisher discriminant analysis based subspace approach for expression recognition. *Engineering Science and Technology, an International Journal*, 19(3):1321 – 1333.
Jia, J., Xu, Y., Zhang, S., and Xue, X. (2016). The facial expression recognition method of random forest based on improved pca extracting feature. In *2016 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, pages 1–5.
Kamal, S., Sayeed, F., and Rafeeq, M. (2016a). Facial emotion recognition for human-computer interactions using hybrid feature extraction technique. In *Data Mining and Advanced Computing (SAPIENCE), International Conference on*, pages 180–184. IEEE.
Kamal, S., Sayeed, F., Rafeeq, M., and Zakir, M. (2016b). Facial emotion recognition for human-machine interaction using hybrid dwt-sfet feature extraction technique. In *Cognitive Computing and Information Processing (CCIP), 2016 Second International Conference on*, pages 1–5. IEEE.
Kamarol, S. K. A., Jaward, M. H., Parkkinen, J., and Parthiban, R. (2016). Spatiotemporal feature extraction for facial expression recognition. *IET Image Processing*, 10(7):534–541.
Latif, M. H. A., Yusof, H. M., Sidek, S. N., and Rusli, N. (2015). Thermal imaging based affective state recognition. In *2015 IEEE International Symposium on Robotics and Intelligent Sensors (IRIS)*, pages 214–219.
Li, X., Ruan, Q., Jin, Y., An, G., and Zhao, R. (2015). Fully automatic 3d facial expression recognition using polytypic multi-block local binary patterns. *Signal Processing*, 108:297 – 308.
Liaw, A. and Wiener, M. (2002). Classification and regression by randomforest. *R news*, 2(3):18–22.
Lyons, M. J., Akamatsu, S., Kamachi, M., Gyoba, J., and Budynek, J. (1998). The japanese female facial expression (jaffe) database. In *Proceedings of third international conference on automatic face and gesture recognition*, pages 14–16.
Marvadi, D., Paunwala, C., Joshi, M., and Vora, A. (2015). Comparative analysis of 3d face recognition using 2d-pca and 2d-lda approaches. In *Engineering (NUiCONE), 2015 5th Nirma University International Conference on*, pages 1–5. IEEE.
P., S., D., K., and Tripathi, S. (2015). Pose invariant method for emotion recognition from 3d images. In *2015 Annual IEEE India Conference (INDICON)*, pages 1–5.
Poursaberi, A., Noubari, H. A., Gavrilova, M., and Yanushkevich, S. N. (2012). Gauss–laguerre wavelet textural feature fusion with geometrical information for facial expression identification. *EURASIP Journal on Image and Video Processing*, 2012(1):1–13.
Poursaberi, A., Yanushkevich, S., and Gavrilova, M. (2013). An efficient facial expression recognition system in infrared images. In *Emerging Security Technologies (EST), 2013 Fourth International Conference on*, pages 25–28. IEEE.
Rajendran, S., Kaul, A., Nath, R., Arora, A., and Chauhan, S. (2014). Comparison of pca and 2d-pca on indian faces. In *Signal Propagation and Computer Technology (ICSPCT), 2014 International Conference on*, pages 561–566. IEEE.
Saha, A. and Wu, Q. J. (2010). Curvelet entropy for facial expression recognition. In *Pacific-Rim Conference on Multimedia*, pages 617–628. Springer.
Shih, F. Y., Chuang, C.-F., and Wang, P. S. (2008). Performance comparisons of facial expression recognition in jaffe database. *International Journal of Pattern Recognition and Artificial Intelligence*, 22(03):445–459.
Sudha, V., Viswanath, G., Balasubramanian, A., Chiranjeevi, P., Basant, K., and Pratibha, M. (2015). A fast and robust emotion recognition system for real-world mobile phone data. In *2015 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, pages 1–6.
Sun, Y. and An, Y. (2010). Research on the embedded system of facial expression recognition based on hmm. In *2010 2nd IEEE International Conference on Information Management and Engineering*, pages 727–731.
Tang, H., Yin, B., Sun, Y., and Hu, Y. (2013). 3d face recognition using local binary patterns. *Signal Processing*, 93(8):2190 – 2198. Indexing of Large-Scale Multimedia Signals.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2794v1 | CC BY 4.0 Open Access | rec: 8 Feb 2017, publ: 8 Feb 2017

Ververidis, D. and Kotropoulos, C. (2008). Fast and accurate sequential floating forward feature selection with the bayes classifier applied to speech emotion recognition. *Signal Processing*, 88(12):2956 – 2970.

Wang, X., Liu, X., Lu, L., and Shen, Z. (2014). A new facial expression recognition method based on geometric alignment and lbp features. In *2014 IEEE 17th International Conference on Computational Science and Engineering*, pages 1734–1737.

Wei, W., Jia, Q., and Chen, G. (2016). Real-time facial expression recognition for affective computing based on kinect. In *2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA)*, pages 161–165.

Zhang, L., Tjondronegoro, D., and Chandran, V. (2011). Toward a more robust facial expression recognition in occluded images using randomly sampled gabor based templates. In *2011 IEEE International Conference on Multimedia and Expo*, pages 1–6.

Zhang, L., Tjondronegoro, D., and Chandran, V. (2014). Random gabor based templates for facial expression recognition in images with facial occlusion. *Neurocomputing*, 145:451 – 464.

Zhang, Y. D., Yang, Z. J., Lu, H. M., Zhou, X. X., Phillips, P., Liu, Q. M., and Wang, S. H. (2016). Facial emotion recognition based on biorthogonal wavelet entropy, fuzzy support vector machine, and stratified cross validation. *IEEE Access*, 4:8375–8385.

Zhao, S., Rudzicz, F., Carvalho, L. G., Marquez-Chin, C., and Livingstone, S. (2014). Automatic detection of expressed emotion in parkinson's disease. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4813–4817.

Zhao, X., Zou, J., Li, H., Dellandréa, E., Kakadiaris, I. A., and Chen, L. (2016). Automatic 2.5-d facial landmarking and emotion annotation for social interaction assistance. *IEEE Transactions on Cybernetics*, 46(9):2042–2055.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2794v1 | CC BY 4.0 Open Access | rec: 8 Feb 2017, publ: 8 Feb 2017