Effects of Region Features on the Accuracy of Cross-database Facial Expression Recognition

Yining Yang, Branislav Vuksanovic and Hongjie Ma

School of Energy and Electronic Engineering University of Portsmouth Portsmouth, U.K.

Keywords: Facial Expression Recognition, Cross-database, Facial Region Feature, Local Binary Patterns, Feature Extractions.

Abstract: Facial expression recognition (FER) in the context of machine learning refers to a solution whereby a computer vision system can be trained and used to automatically detect the emotion of a person from a presented facial image. FER presents a difficult image classification problem that has received increasing attention over recent years mainly due to the availability of powerful hardware for system implementation and the greater number of possible applications in everyday life. However, the FER problem has not yet been fully resolved, with the diversity of captured facial images from which the type of expression or emotion is to be detected being one of the main obstacles. Ready-made image databases have been compiled by researchers to train and test the developed FER algorithms. Most of the reported algorithms perform relatively well when trained and tested on a single-database but offer significantly inferior results when trained on one database and then tested using facial images from an entirely different database. This paper deals with the cross-database FER problem by proposing a novel approach which aggregates local region features from the eyes, nose and mouth and selects the optimal classification techniques for this specific aggregation. The conducted experiments show a substantial improvement in the recognition results when compared to similar cross-database tests reported in other works. This paper confirms the idea that, for images originating from different databases, focus should be given to specific regions while less attention is paid to the face in general and other facial sections.

1 INTRODUCTION

Facial expressions and gestures play important roles in human communication as they can express information directly. Facial expression is one of the most powerful and natural ways for human beings to relay their emotions and intentions (Ying-Li, Kanada, & Cohn, 2001), (Li & Deng, 2018). Although facial expression is a universal language that can widely and directly present emotions (Perikos, Paraskevas, & Hatzilygeroudis, 2018), facial expression recognition (FER) poses a challenge in the field of computer vision. One expectation is that a machine will be able to understand and interpret human emotions; however, no machine is capable of utilising emotions for communication. In real life, 55% of human emotions are shown through facial expressions (Chen et al., 2019)(Xie & Hu, 2019) thus FER is important for human–robot interaction. Enabling machines to recognise human behaviour is an area that has been progressively developed in recent years. In previous decades, human–robot interaction has been studied in social and behavioural sciences whereas the techniques for machine-based FER have also become a popular and intensely researched topic in recent years. Thus, finding related applications in psychology, behavioural science and human computer interfacing, such as interpersonal relation prediction by FER, is important (Zhang, Luo, Loy, & Tang, 2018).

Facial expression categorisation can be geometric-based, appearance-based and either local or global (Kumari, Rajesh, & Pooja, 2015). Normally, seven prototypical or basic facial expressions can be detected and identified: neutral, happiness, sadness, fear, disgust, surprise and anger (Ekman & Friesen, 1971), (Ekman, 1994). Emotions are not often presented by prototypic expressions. Instead, emotions are communicated by certain local changes of the face (Ying-Li et al., 2001). Emotions are generated by the muscles of facial action units (AUs) (Liu Yanpeng et al., 2016). The combination of

different AUs can then generate different facial expressions (Liu Yanpeng et al., 2016). However, the capability to influence models based on prototypic emotions to represent the complexity and subtlety of daily emotions is limited (Martinez & Valstar, 2016).

The face can be divided into upper and lower face AUs (Ying-Li et al., 2001). The upper face model trained by the Ekman–Hager (EH) database yielded accuracies of 93.2% and 96.4% when tested using the EH and Cohn–Kanade (CK) databases (Kanade, Cohn, & Tian, 2000), respectively. The lower face model trained by the CK database features an accuracy of 93.4% when tested using the EH database and an accuracy of 96.7% concerning the CK database. The CK database contains 81% Caucasian, 13% African and 6% other groups whereas the EH database includes 24 Caucasian subjects. Thus, this model performs well in Caucasian FER.

Chen proposed a Softmax regression-based deep sparse autoencoder network to recognise facial emotions(Chen et al., 2018). The first steps involve the extraction of a region of interest (ROI) for facial expression image features. The ROI in (Chen et al., 2018) included the eyebrows, eyes and mouth. Extracting these ROI areas can not only reduce the interference in facial information caused by image interference in noncritical parts but also reduce the amount of data and thus improve the computing speed. Their experiments used Japanese female facial expression (JAFFE) (Li & Deng, 2018) and extended CK (CK+) databases(Lucey et al., 2010). The average accuracy results of JAFFE and CK+ databases were 89.12% and 89.03%, respectively. In this experiment. the database was divided into three groups. Therefore, the test dataset differed from the training dataset. However, the results from JAFFE or CK+ only support single databases.

Y. Fan et al studied FER via local regions, including the left eye, nose and mouth, and deep learning algorithms.(Fan, Lam, & Li, 2018). They proposed three points for FER. The first point is a novel multi-region ensemble convolutional neural network (CNN) framework that aims to improve CNN models by using multiple facial regions which include global features and local regions. The second point is that the weighted prediction scores from each sub-network are aggregated to produce a final high-accuracy prediction. The third point is to research the effect of different regions from face images on FER. The network of CNNs could show both low-level profile features and high-level specific features. In the sub-network, each local region (left eye, nose and mouth) and whole face will be inputted into a double-input sub-network. With three particulate regions, three prediction scores are

obtained. The ensemble prediction stage will achieve a final prediction rate based on the weighted sum operation (Fan et al., 2018).

Another paper (Xie & Hu, 2019) presented a method for inputting local facial features and the whole face separately. However, this network features a CNN structure containing two branches and no sub-network. The method, named deep comprehensive multipatch aggregation CNN, consists of two hierarchical features: local and holistic. The local features are extracted from image patches and depict details of expression. The holistic features are extracted from the whole image and provide high-level semantic information. Both features are aggregated before classification. The common method of FER uses only a single feature type, but this method uses two feature types to interpret expressional information. However, in the training step, a novel pooling method that can handle nuisance variations, that is, expressional transformation invariants, is proposed. The evaluation of the method uses CK+ and JAFFE databases and a cross-database evaluation is adopted to test the method (Xie & Hu, 2019).

Liu and Chen proposed the combined CNN–centralized binary pattern (CBP) which consists of CBP and CNN features(Liu & Chen, 2017). The features were then classified using support vector machine (SVM). With the CNN–CBP features, the average recognition accuracies of CK+ and JAFFE databases reached 97.6% and 88.7%, respectively. However, with the same model, the accuracy totalled 34.6% when training with the CK+ database and testing with the JAFFE database.

Other authors (Zavarez, Berriel, & Oliveira-Santos, 2017) have proposed a visual geometry group (VGG)–face deep convolutional network model. When testing with a VGG–fine-tuned model, which was trained with CK+, JAFFE, MMI, RaFD, KDEF, BU3DFE and ARFace databases, the test accuracies of CK+ and JAFFE totalled 88.58% and 44.32%, respectively. In (da Silva & Pedrini, 2015), the model, which also has the poor performance in cross-database, consisted of a histogram of oriented gradient filter and a SVM classifier. The experiment tested four databases, including CK+, JAFFE, MUG and BOSPHORUS. The model accuracy was 42.3% when the model was trained by the CK+ database and tested by JAFFE database. If the training database was JAFFE and the testing database was CK+, the accuracy was 48.2%. The author believed that different cultures could confuse the classifier for recognition(da Silva & Pedrini, 2015). On the basis of local binary patterns (LBPs), (Shan, Gong, & McOwan, 2009) formulates boosted LBP to extract LBP features. The SVM was used for LBP feature

classification. It has a similar result to the above research in a cross-database. When the model was trained by CK+ database, the generalisation performance boosted the LBP-based SVM on different datasets and the accuracy result in approximately 40% on the JAFFE database and 50% on MMI.

Thus, according to reviewed literature, FER accuracy is reasonably high when the training and testing datasets originate from the same, i.e. single, database. However, for cross-database FER experiments, where training and testing datasets originate from different databases, the recognition accuracy severely deteriorates. This paper proposes the use of local areas extracted from facial images to improve the accuracy of the cross-database FER. The rest of the paper is organised in the following way. Section 2 describes the proposed method where details of the face extraction approach are described and the reasons behind the SVM classifier choice in

This approach is justified. Section 3 describes the conducted experiments and discusses the results using this approach. Achieved recognition rates indicate significant improvement compared to similar cross-database FER tests and recognition rates reported in the literature, also reviewed earlier in this section. Section 4 summarises the work and provides certain conclusions.

2 PROPOSED FER APPROACH

To test the cross-database FER performance, a traditional framework-based FER system with LBP and SVM was used in this work. This section provides information about the main stages of the employed FER system illustrated in Fig 1.

2.1 Databases

According to published papers, various ready-made datasets are available to test and assess the performance of different expression recognition algorithms. The proposed algorithm was tested on JAFFE and CK+ databases. Table 1 provides the details for each database. The CK+ database contains facial images from Western (Caucasian) populations. The JAFFE database, on the other hand, consists of the facial images of Asian, specifically Japanese, females. In addition to certain structural differences, cultural differences also exist between the two databases. Thus, their recognition accuracies are relatively poor.
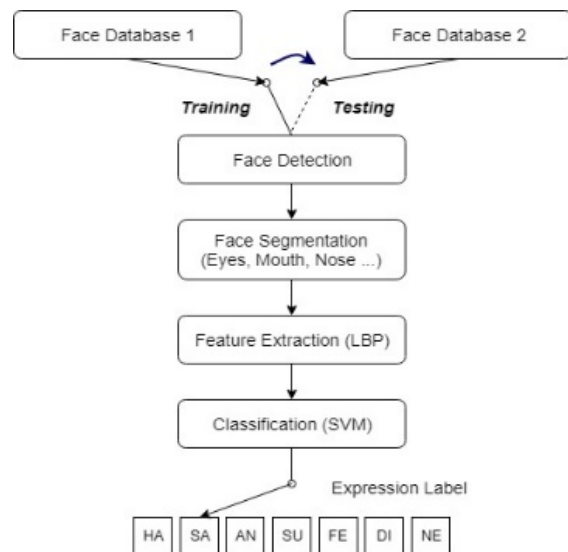

Figure 1: Main stages of the employed FER system (HA: happiness, SA: sadness, AN: anger, SU: surprise, FE: fear, DI: disgust and NE: neutral).

Table 1: Details of two databases used in this paper.

Database	Facial expression	Number of Subjects	Number of images	Number of images used in the experiment
JAFFE	neutral, happiness, sadness, fear, angry, disgust and surprise	10	213 static images	213 static images
CK+		123	327 sequences	700 static images

The JAFFE database (Li & Deng, 2018) contains 3–5 images in each of the seven expressions from each subject.

The CK+ database(Ekman, 1994) (Lucey et al., 2010) consists of 593 expression sequences from 123 subjects, where 327 sequences are labelled with one of the seven expressions (angry, disgust, fear, happy, sad, surprise and contempt). The 123 subjects came from different regions with varying races, ages and genders. Each image sequence contained a set of captured frames when the subject changed from a neutral emotional state and finishes at the peak expression. The neutral frame and four peak frames of each sequences were selected from the 327 labelled sequences. Based on the balance from the eight expressions, each expression will include 100 images. Compared with the

seven facial expressions in the JAFFE database, the same facial expressions (except for contempt) were used for the CK+ database. Hence, the CK+ database included 700 images in total.

2.2 Face Detection

From the database image, the face or feature region should be detected before training or testing the FER model. Face detection could reduce the effect of property information. The face for each input image in the proposed approach is detected by Dlib. The Dlib is an open-source C+ library implementing a variety of machine learning algorithms, including classification, regression, data transformation and structured recognition. The Dlib can be used as a tool for high-quality face recognition (Davis King, 2003).
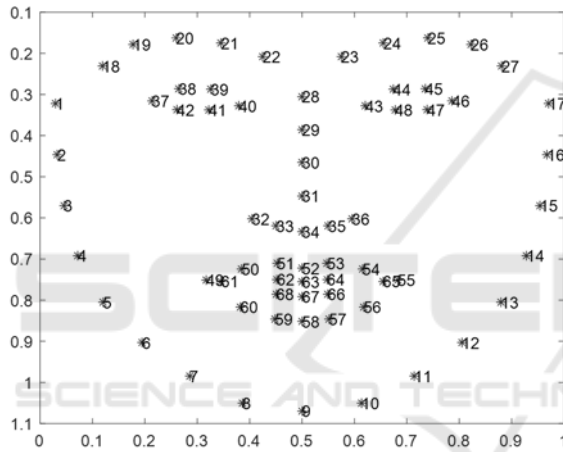

Figure 2: Sixty-eight (68) point face landmarks from Dlib.

The pre-trained facial landmark detector inside the Dlib was used to estimate the location of 68 (x, y) coordinates that were mapped to facial structures. The indexes of the 68 coordinates were visualised in the image scheme above (Fig 2.). The whole face and local region of the face could be detected using the indicated landmarks.(Boyko, Basystiuk, & Shakhovska, 2018).

2.3 Feature Extraction Via LBP

In accordance with the detected image region, features will be extracted by LBP. The original LBP operator was introduced by Ojala et al.(Ojala, Pietikäinen, & Harwood, 1996) and was proven to be a powerful means of texture description. LBP has since been successfully applied to a wide range of other image recognition tasks, such as FER (Sun, Li, Zhou, & He, 2016) (Levi & Hassner, 2015). The operator labels the

pixels of an image by thresholding a 3 x 3 neighbourhood of each pixel with the centre value. The result of thresholding can be considered a binary number (see Fig. 3 for an illustration) and the 256-bin histogram of the LBP labels computed over a region was used as a texture descriptor.
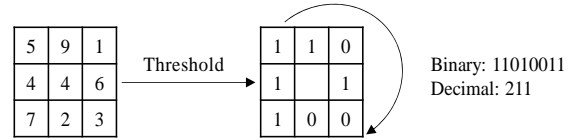

Figure 3: Basic LBP operator.

The LBP operator generates a binary number. The binary number compares the neighbouring pixel values with the centre pixel value. The pattern with eight neighbourhoods is expressed by

$$LBP(x,y) = \sum_{n=0}^{n-1} 2^n \times s(i_n - i_c) \quad (1)$$

i_n is the pixel value at coordinates in the neighbourhood of (x, y) and i_c is the pixel value at coordinate (x, y). The operator $LBP(x, y)$ produces 2^n different outputs, corresponding to 2^n different binary patterns formed by n pixels in the neighborhood.

$$S(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases} \quad (2)$$

The histogram of LBP labels calculated over a region that can be exploited as a feature descriptor is given by

$$H_i = \sum_{x,y} I\{LBP(x,y) = i\} \quad (3)$$

The limitation of the basic LBP operator is its small 3 x 3 neighbourhood which cannot capture dominant features with large-scale structures (Shan et al., 2009). Hence, the operator was later extended to use neighbourhoods of different sizes (Ojala, Pietikäinen, & Mäenpää, 2002).

After labelling an image with the LBP operator, a histogram of the labelled image contains information about the distribution of the local micro-patterns over the whole image. Thus, the histogram can be used to statistically describe image characteristics. Face images can be observed as a composition of micro-patterns that can be effectively described by the LBP histograms. Therefore, LBP features were intuitively used to represent face images (Shan et al., 2009). The LBP histogram computed over the whole face image encoded only the occurrences of the micro-patterns without any indication about their locations.

2.4 Classification Via SVM

A SVM classifier was selected as it is well founded in statistical learning theory and has been successfully applied in various tasks in computer vision (Zhao, 2007). SVM is a technique previously used successfully in facial expression classification. As a powerful machine learning technique for data classification, SVM performs an implicit mapping of data into a higher (maybe infinite) dimensional feature space and then finds a linear separating hyperplane with the maximal margin to separate data in this higher dimensional space (Kanade et al., 2000) (Shan et al., 2009). The quadratic SVM was used for classification in this paper. The multiclass method of the SVM is a one-against-one method. When the model is trained, the SVM will use fivefold cross-validation.

3 EXPERIMENTS AND RESULTS

The performance of the FER system was evaluated using three different sections of facial images from two different databases, where images from the first database were used only for system training and the images from the second database were only used for testing.

3.1 Local Image Detection

Sections of the facial images included a) the original facial image from the database with background, b) narrowed, i.e. extracted facial region image without background and c) local regions (eyes, nose and mouth) used in the experiments (Fig. 4). The eye region includes eyebrows and eyes. The nose region includes nose and sides of the nose. The mouth region includes mouth, jaw and the two sides of the mouth. The local facial regions further increase the proportion of effective information.
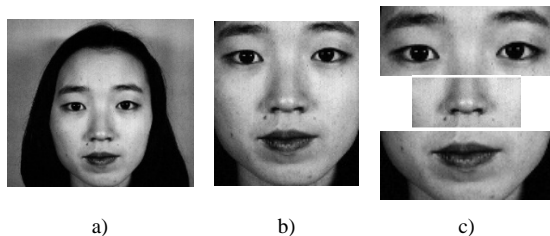

| a) | b) | c) |

Figure 3: Facial regions used in the experiment.

3.2 Classification Selection

Different classifiers have been tested and their

performance compared using features from the original and narrowed facial images in both the JAFFE and CK+ databases in order to select the most appropriate one. The results show that the quadratic SVM offers somewhat superior performance when compared to other available classification algorithms. Those include cubic SVM as well as a k-Nearest Neighbor classifier, although both of those resulted in acceptable recognition rates. Based on the comparison of the three classifiers (Table 2), the quadratic SVM classifier was finally selected for the following set of experiments.

Table 2: Performance of single-database FER for different classifications.

| No | Classification | | Accuracy | | | |
| | | | JAFFE | | CK+ | |
	Type	Name	Wide Image	Narrow Image	Wide Image	Narrow Image
1	SVM	Quadratic SVM	80.3%	81.2%	98.6%	99.6%
2	SVM	Cubic SVM	79.8%	81.2%	98.6%	99.6%
3	KNN	Fine KNN	80.3%	81.6%	97.6%	99.1%

As explained in the previous section, LBP was used to extract relevant facial features which were then classified using the quadratic SVM technique.

3.3 Cross-database FER Via Original and Narrowed Facial Images

Using original and narrowed facial images, the cross-database test was carried out and the following results were obtained:

Table 3: Performance of single- and cross-database FER arrangements for different facial ranges.

| Databases | Single-Database FER | | Cross-Database FER | |
	CK+	JAFFE	CK+ Trained (JAFFE Tested)	JAFFE Trained (CK+ Tested)
Whole Face with Background	99.00%	80.75%	18.78%	15.14%
Narrowed Face (no Background)	98.43%	84.98%	33.33%	32.86%

In the case of original facial images, the classifier performed well with single-database but the recognition rate reached no more than 20% with cross-databases. In

the case of narrowed facial images, the accuracy of a single-database slightly changed, whereas that of a cross-database notably improved. However, both accuracy values were higher than 30%. This is in agreement with (Liu & Chen, 2017) where similarly low recognition rates of 34.6% were reported for cross-database FER recognition. These findings also show that the recognition rate with cross-databases can be improved by reducing background information. This conclusion led to the use of local facial features for FER.

3.4 Cross-database FER Via Local Region Images

To enhance the recognition accuracy by increasing the proportion of useful information, three local facial regions, eyes, nose and mouth, have been extracted and tested for FER.

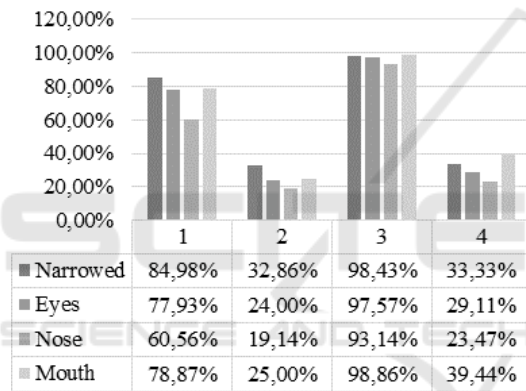|	1	2	3	4
■ Narrowed	84,98%	32,86%	98,43%	33,33%
■ Eyes	77,93%	24,00%	97,57%	29,11%
■ Nose	60,56%	19,14%	93,14%	23,47%
Mouth	78,87%	25,00%	98,86%	39,44%

Figure 4: Accuracy variation for cross-database arrangements compared with single-database system with single features (1: JAFFE single database, 2: JAFFE-trained and CK+-tested, 3: CK+ single database, 4: CK+-trained and JAFFE tested).

After conducting the local facial image test, the accuracy of the mouth region increased by 6.11% when the model was trained by a CK+ database and the expression identified by a JAFFE database. This finding means that the mouth region contains the information most relevant for facial expression recognition. However, the accuracy of models trained using the JAFFE database with cross-databases validation decreased relative to the results obtained with narrowed facial images.

The newly obtained features which comprised pairwise aggregation of local facial features were used for testing. For the model trained with the JAFFE database, when JAFFE and CK+ databases were used for testing, the combination of the eye and mouth regions was more accurate than the single-feature

result and slightly better than the finding obtained using narrowed facial images, respectively. These results suggest that the mouth and eyes contain useful information.

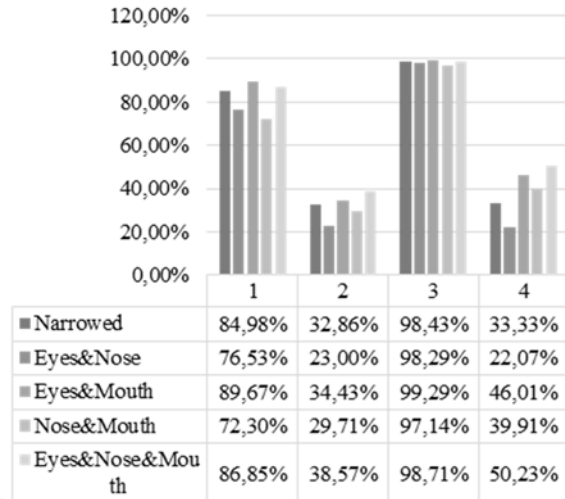|	1	2	3	4
■ Narrowed	84,98%	32,86%	98,43%	33,33%
■ Eyes&Nose	76,53%	23,00%	98,29%	22,07%
■ Eyes&Mouth	89,67%	34,43%	99,29%	46,01%
Nose&Mouth	72,30%	29,71%	97,14%	39,91%
Eyes&Nose&Mouth	86,85%	38,57%	98,71%	50,23%

Figure 5: Accuracy variation for cross-database arrangements compared with single-database system with fusion features (1: JAFFE single database, 2: JAFFE-trained and CK+-tested, 3: CK+ single database, 4: CK+-trained and JAFFE-tested).

For the model trained by the CK+ database, the accuracy of a single database showed no significant improvement. However, when tested by the JAFFE database, the accuracy also significantly improved with the combination features of the eye and mouth regions. Compared to the results achieved using narrowed facial images, the accuracy improved by 12.68%. When compared to the result achieved using only the mouth region, the accuracy improved by 6.57%.

Finally, the results of the conducted experiments can also be compared to the cross-database recognition rates reported in other papers. When using narrowed facial images or single face regions, recognition rates are very similar to those reported elsewhere. However, the recognition rates achieved in this work, when tests are performed using features aggregated by the three local facial regions, are significantly higher compared to other reported results reported. Table 4 provides an overview and comparison of the recognition rates. It is also worth pointing out that, compared to the paper by Zavarez (2017), the proposed method requires less training which reduces the challenges arising from the size of the training set and the computational power required for model training. This indicates that the FER accuracy is improved when using the new feature because it increases the amount of effective information that can support the FER when using cross-databases.

Table 4: Comparison with previous approaches.

	Training Databases (Test: JAFFE)	Accuracy
Liu and Chen (Liu & Chen, 2017)	CK+	34.6%
Shan, Gong, and McOwan (Shan et al., 2009)	CK+	40%
Silva and Pedrini (da Silva & Pedrini, 2015)	CK+	42.3%
Zavarez, Berriel, and Oliveira-Santos (Zavarez et al., 2017)	CK+, JAFFE, MMI, RaFD, KDEF, BU3DFE and ARFace	44.32%
Proposed	CK+	50.23%

Finally, three local features were aggregated to obtain a new feature with the intention of further reducing the proportion of less relevant information contained in the training data set. The new feature was tested by cross-databases. The results of the JAFFE database-trained model tested in a single-database showed a slight improvement over the test results for narrowed facial images. Meanwhile, the accuracy rate of the cross-databases test results increased by 5.71%. The most significant improvement in the recognition accuracy was the recognition results of the model trained by the CK+ database. The accuracy of this cross-databases test reached 50.23% which is a 16.9% improvement compared to the result achieved using narrowed facial images.

This new feature is more sensitive to the expression of "disgust" as indicated by the confusion matrix (Table 5). Although the recognition rate for "disgust" is low, the new feature can accurately recognise this expression. Secondly, this novel feature is relatively sensitive to expressions of both "happiness" and "surprise".

Table 5: Confusion matrices for FER system tests using a fusion feature with three local regions (units: %).

		Prediction Label							
		AN	DI	FE	HA	NE	SA	SU	Acc
	AN	**50.00**	00.00	00.00	03.33	33.33	10.00	03.33	**50.00**
	DI	24.14	**17.24**	24.14	00.00	24.14	06.90	34.5	**17.24**
Real Label	FE	03.13	00.00	**28.13**	06.25	43.75	00.00	18.75	**28.13**
	HA	03.23	00.00	03.23	**80.65**	09.68	00.00	03.23	**80.65**
	NE	03.33	00.00	06.67	00.00	**66.67**	06.67	16.67	**66.67**
	SA	25.81	00.00	03.23	03.23	19.35	**38.71**	09.68	**38.71**
	SU	13.33	00.00	03.33	03.33	10.00	00.00	**70.00**	**70.00**
	Average Accuracy								**50.23**

4 CONCLUSIONS

A large number of different FER systems reported in the literature perform well when training and test samples both originate from the same, precompiled database of facial images. However, the accuracy of results drops drastically when the same system is tested using images from an entirely different database not used in the training phase. This paper investigated the influence of various sections of a facial image on the level of deterioration in cross-database FER system performance. It was found that the drop in system performance is less severe if the background is removed from the image. Comparing the image having been removed the background with important sections of the facial image, the performance of the recognition rate has different improvements when using important sections of the facial image. Encouraging results have been recorded when the mouth region was used in the experiment, a region which was proven to hold a significant and crucial amount of information related to facial expression and emotion of the person in the image. When the new feature aggregated the features from the eyes, nose and mouth, the proportion of effective information further increased. The experiments showed substantial improvement in the recognition results. The recognition accuracy of 50.23% represents a significant improvement when compared to cross-database FER results reported elsewhere in the research literature. It is also worth noting that the result is achieved using a "classical" approach, i.e. without employing deep learning techniques, thus requiring significantly less computing power. Future work will now focus on testing the performance of deep learning algorithms using only the most important sections of facial images in similar cross-database arrangements.

REFERENCES

Boyko, N., Basystiuk, O., & Shakhovska, N. (2018). Performance Evaluation and Comparison of Software for Face Recognition, Based on Dlib and Opencv Library. *Proceedings of the 2018 IEEE 2nd International Conference on Data Stream Mining and Processing, DSMP 2018*, 478–482. https://doi.org/10.1109/DSMP.2018.8478556

Chen, L., Li, M., Su, W., Wu, M., Hirota, K., & Pedrycz, W. (2019). *Adaptive Feature Selection-Based AdaBoost-KNN With Direct Optimization for Dynamic Emotion Recognition in Human–Robot Interaction.pdf*.

Chen, L., Zhou, M., Su, W., Wu, M., She, J., & Hirota, K. (2018). Softmax regression based deep sparse autoencoder network for facial emotion recognition in human-robot interaction. *Information Sciences*, *428*, 49–61. https://doi.org/10.1016/j.ins.2017.10.044

da Silva, F. A. M., & Pedrini, H. (2015). Effects of cultural characteristics on building an emotion classifier through

facial expression analysis. *Journal of Electronic Imaging*, *24*(2), 023015. https://doi.org/10.1117/1.jei. 24.2.023015

Davis King. (2003). D-lib C++ library. Retrieved June 16, 2019, from http://dlib.net/

Ekman, P. (1994). Strong evidence for universals in facial expressions a reply to Russell's mistaken critique. *Psychological Bulletin*, *115*(2), 268–287. https://doi.org/10.1037/0033-2909.115.2.268

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, *17*(2), 124–129. https://doi.org/10.1037/h0030377

Fan, Y., Lam, J. C. K., & Li, V. O. K. (2018). *Multi-region Ensemble Convolutional Neural Network for Facial Expression Recognition*.

Kanade, T., Cohn, J. F., & Tian, Y. (2000). Comprehensive Database for Facial Expression Analysis. *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580)*, 46–53. https://doi.org/10.1109/AFGR.2000. 840611

Kumari, J., Rajesh, R., & Pooja, K. M. (2015). Facial Expression Recognition: A Survey. *Procedia Computer Science*, *58*, 486–491. https://doi.org/10.1016/j.procs. 2015.08.011

Levi, G., & Hassner, T. (2015). Emotion Recognition in the Wild via Convolutional Neural Networks and Mapped Binary Patterns. *ICMI '15 Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, 503–510. https://doi.org/10.1145/2818346.2830587

Li, S., & Deng, W. (2018). *Deep Facial Expression Recognition: A Survey*. 1–25. Retrieved from http://arxiv.org/abs/1804.08348

Liu, Y., & Chen, Y. (2017). Recognition of facial expression based on CNN-CBP features. *Proceedings of the 29th Chinese Control and Decision Conference, CCDC 2017*, 2139–2145. https://doi.org/10.1109/CCDC.2017. 7978869

Liu Yanpeng, Cao Yuwen, Li Yibin, Liu Ming, Song Rui, Wang Yafang, … Ma Xin. (2016). Facial Expression Recognition with PCA and LBP Features Extracting from Active Facial Patches. *Proceedings of The 2016 IEEE International Conference on Real-Time Computing and Robotics*, 368–373. https://doi.org/10. 1109/RCAR.2016.7784056

Lucey, P., Cohn, J. F., 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. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, 94–101. https://doi.org/10. 1109/CVPRW.2010.5543262

Martinez, B., & Valstar, M. F. (2016). Advances, challenges, and opportunities in automatic facial expression recognition. *Advances in Face Detection and Facial Image Analysis*, 63–100. https://doi.org/10. 1007/978-3-319-25958-1_4

Ojala, T., Pietikäinen, M., & Harwood, D. (1996). A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition*, *29*(1), 51–59. https://doi.org/10.1016/ 0031-3203(95)00067-4

Ojala, T., Pietikäinen, M., & Mäenpää, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*(7), 971–987. https://doi.org/10.1109/ TPAMI.2002.1017623

Perikos, I., Paraskevas, M., & Hatzilygeroudis, I. (2018). Facial Expression Recognition Using Adaptive Neuro-fuzzy Inference Systems. *2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS)*, 1–6. https://doi.org/10.1109/ICIS.2018. 8466438

Shan, C., Gong, S., & McOwan, P. W. (2009). Facial expression recognition based on Local Binary Patterns: A comprehensive study. *Image and Vision Computing*, *27*(6), 803–816. https://doi.org/10.1016/j.imavis.2008. 08.005

Sun, B., Li, L., Zhou, G., & He, J. (2016). Facial expression recognition in the wild based on multimodal texture features. *Journal of Electronic Imaging*, *25*(6), 061407. https://doi.org/10.1117/1.jei.25.6.061407

Xie, S., & Hu, H. (2019). Facial Expression Recognition Using Hierarchical Features with Deep Comprehensive Multipatches Aggregation Convolutional Neural Networks. *IEEE Transactions on Multimedia*, *21*(1), 211–220. https://doi.org/10.1109/TMM.2018.2844085

Ying-Li, T., Kanada, T., & Cohn, J. F. (2001). Recognizing upper face action units for facial expression analysis. *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No.PR00662)*, *1*(2), 294–301. https://doi.org/10.1109/CVPR.2000. 855832

Zavarez, M. V., Berriel, R. F., & Oliveira-Santos, T. (2017). Cross-Database Facial Expression Recognition Based on Fine-Tuned Deep Convolutional Network. *Proceedings - 30th Conference on Graphics, Patterns and Images, SIBGRAPI 2017*, 405–412. https://doi.org/ 10.1109/SIBGRAPI.2017.60

Zhang, Z., Luo, P., Loy, C. C., & Tang, X. (2018). From Facial Expression Recognition to Interpersonal Relation Prediction. *International Journal of Computer Vision*, *126*(5), 550–569. https://doi.org/10.1007/s11263-017-1055-1

Zhao, G. (2007). Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions. *Pattern Analysis and Machine Intelligence*, *29*(6), 915–928. https://doi.org/10.1109/ TPAMI.2007.1110