ISSN 1816-6075 (Print), 1818-0523 (Online)
Journal of System and Management Sciences
Vol. 12 (2022) No. 5, pp. 57-72
DOI:10.33168/JSMS.2022.0504

Sparse Representation with Principal Component Analysis in Face Recognition

Ming-Chun Yo, Siew-Chin Chong+, Kuok-Kwee Wee, Lee-Ying Chong

Faculty of Information Science & Technology, Multimedia University, Melaka, Malaysia

chong.siew.chin@mmu.edu.my (corresponding author)

Abstract. Face recognition has been one of the most reliable biometric technologies due to its easy and non-intrusive method during acquisition procedure. Multiple algorithms and methods have been developed and invented by the researchers and computer scientists in order to increase and improve the performance of face recognition. Sparse representation method has attracted a lot of attention in the fields of machine learning recently and it boosts the research of sparsity-based pattern recognition among the researchers. In this research paper, we aim to investigate the impact of sparse representation in face recognition. The proposed method utilizes the fusion of Principal Component Analysis and Sparse Representation Classification to enhance the accuracy of the face recognition. Experimental results demonstrate that the proposed method can achieved almost 99% accuracy using the FERET dataset.

Keywords: sparse representation, face recognition, FERET dataset, classification, feature extraction.

1. Introduction

Since today's improvements in information technology necessitate the requirement for high security, the use of biometric technology as an identity and recognition method has gained a lot of awareness in this globalized period. The biometric authentication systems are gaining importance gradually. However, face recognition is regarded as one of the dependable biometric technologies when it comes to security concerns and it is commonly used for identification due to its simple and non-intrusive collection approach.

Face recognition can be done in a variety of ways. However, there are three main phases which are pre-processing, feature extraction and classification can be broken down in the most common process of face recognition. In pre-processing phase, the captured face image has to be segmented from the background which means just crop for the region on interest so that the face image can be analyzed more precisely on later process. After that, the feature extraction step extracts the first stage's significant features and information. It converts the segmented picture into an array of quantitative data in a computer-readable format and constructs a feature vector depending on the architecture of the obtained data. Then, the classifier detects a person's identification based on the detected traits on their face.

In the recent years, sparse representation has been widely used in face recognition and it promotes sparsity-based pattern recognition research among the researchers. The testing data is expressed as sparse and linear combination of the training data and l_1-norm or l_2-norm of forming residual is able to determine the fidelity in sparse representation. The literature works have concluded that the sparse representation-based classification (SRC) method is able to handle some of the challenges posed by lighting variations, random pixel corruption and huge block occlusion or disguise (Wright, J., et al., 2009). Therefore, the major goal of this study is to determine how sparse representation-based classification combined with feature extraction technique which is principal component analysis affects facial recognition.

2. Literature Review

2.1. Face recognition

Face recognition is one of the famous and commonly used biometric technologies recently. It is because face recognition is distinct comparing to the other biometric techniques. In face recognition, the individual in the image is not required to engage with the procedure. In contrast, it is necessitated people's active involvement in the procedure while using other techniques. However, the image is able to be captured without the individual even being aware of it happened in face recognition.

Face recognition has considered as pattern recognition's subset in visual. Humans use their eyes to receive and perceive the visual information. After that, the brain can convert the information that received to the valuable concepts. However, no matter the information is a picture or video, it will be always considered as a matrix of many pixels for a computer. The computer will have to think what each piece of data in the data represents. It is a crude classification challenge for visual model identification. For all the computers, it is important to differentiate the face in the face recognition. Face recognition technology has gained popularity in a lot of applications and it has become the future development direction.

2.2. Sparse representation

Sparse representation gains much attention among the researchers in recent years (Zhang, Z., et al., 2015). A signal can be represented by a sparse and dictionary atoms' linear combination, according to sparse representation concept (Li Y., 2013). It intends to represent signals with as few as possible significant coefficients (Giron-Sierra J.M., 2017).

Since the use of sparse representation in performing robust face recognition was first suggested by John Wright in 2008, there are many academics have tried to use the sparse representation method to the domain of pattern recognition particularly in image classification (Wright, J., et al., 2009). The main objective is to categorize the provided test image into multiple predetermined classes. From the view of the visual neurons' characteristics, it has been proved that the natural images are able to be sparsely represented.

2.3. Classification

In face recognition, the classification used the features that given by feature extractor to allocate the item for a class. It is needed to evaluate the features presented and make decision about which category that the data belongs to. Data can be categorized in two ways which are organized and unstructured. It is an approach for categorizing data into a set of subjects. Its goal is to decide the new data will belong to which class or category. It can be roughly categorized by 4 steps. First of all, initialize the classifier that is going to be used. Then, train the classifier and predict the target. Lastly, evaluate the classifier model.

2.4. Types of classification method

2.4.1. Sparse representation based classification (SRC)

Sparse representation-based classification (SRC) method supposes that samples which are used for training from the identical class can be used to appropriately reflect the testing sample. To be more precisely, SRC represents the testing data using a linear combination of training data and then calculates the linear representation system's sparse representation coefficients. After that, it uses the coefficients and training data to compute each subject's reconstruction residuals. Lastly, the testing data will be then categorized into a class that resulting in the minimum reconstruction residual. It has proved that while dealing with the problem of image classification on disguised or distorted images, the SRC method has enormous strength exist (Wright, J., et al., 2009). For this kind of scenarios, the natural image is able to be represented sparsely and the image classification job can be accomplished by using the sparse representation theory.

2.4.2. Euler sparse representation based classification (Euler SRC)

Euler Sparse Representation based Classification (Euler SRC) method is essentially the SRC method with Euler sparse representation. To be more specifically, it first transforms the images into the complex space using the Euler representation which has no influence on outliers or lighting. After that, it performs complex SRC with Euler representation. Euler SRC method comes from the same dimensionality reduction family tree as the kernel sparse representation, but it is clearly different in the way of organizing mapping. It has been demonstrated that Euler representation is explicit with no increase of the image space dimensionality, enabling the fast and simple computation (Liu, Y., et al., 2018).

2.4.3. Support vector machine (SVM)

Vapnik and Cortes developed support vector machine (SVM) in 1995. It is an algorithm that dedicated to the challenge of facial recognition problem that is tiny in sample but with high dimensions (L. Li., et al., 2020). Normally, the feature extractor provides the facial characteristics in face recognition and the hyperplane for identifying various faces is generally found by SVM. Classification that conducted by SVM is discovering the hyperplane which contains the largest margin between the two subjects by the aid of the nearest data point which is called as support vector. A hyperplane can be considered as that divides a group of items with distinct class memberships. This method is able to distinguish data in two-dimensional plane, three-dimensional space or even higher, only the hyperplane or plane had become from the discovered decision boundary.

3. Proposed Implementation

3.1. Proposed method

In this research paper, the Face Recognition Technology (FERET) had been chosen as the dataset. After pre-processing the dataset, the dataset had been split and conducted the dataset arrangement. Then, an h5 file format dataset had been generated and the proposed feature extraction technique which is Principal Component Analysis (PCA) had been used to reduce the redundant data from the dataset and remain the distinguishable features. After that, the proposed classification method which is Sparse Representation based Classification (SRC) had been used and compared the performance with other classification methods such as Euler Sparse Representation based Classification (Euler SRC) and Support Vector Machine (SVM). Three of this classification method had been combined with the feature extraction technique in order to see which one performs better for face recognition. Fig 1 shows the proposed method flowchart.

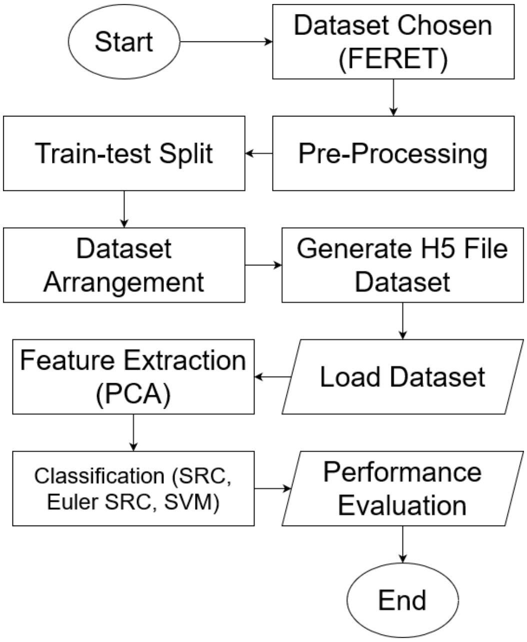

Fig. 1: Flowchart of the proposed method.

3.2. Feature extraction: principal component analysis (PCA)

For this face recognition process, implementing feature extraction technique is an important step. The technique of extracting descriptors or distinctive traits based on a raw image is known as feature extraction. However, the feature extraction technique that will be used in this research paper is Principal Component Analysis (PCA) (Pearson, K., 1901). It is one of the most successful and classical feature extraction techniques for face recognition experiment to apply.

PCA's goal is to decrease the enormous dimensionality of the data space to the lower inherent dimensionality of the feature space that is required to economically characterize the data. The core concept behind utilizing PCA is to convert a vast 1D vector of pixels generated from a 2D image into dense feature space's principal components for face recognition. It can be known as eigenspace projection. The covariance matrix's eigenvectors obtained from a group of images are recognized and used to compute eigenspace (Kim, K., 2012).

In mathematical formulation of PCA, it consists of three main procedures. Firstly, the training samples are used to generate the data matrix. Then, the samples for training are projected to the columns of matrix. Eventually, make a comparison between testing samples and training samples by projecting testing samples into the subspace, the samples of testing can be recognized successfully.

Let a face image *X* (*x*, *y*) as the intensity values' two-dimensional *m* * *n* array. An image is also able to be considered as the dimension *m* * *n* vector. Let the training set of images $\{X_1, X_2, \ldots, X_n\}$. The set's mean image is defined by Equation (1):

$$\mu = \frac{1}{n}\sum_{k=1}^{n} X_k \tag{1}$$

Then, it is able to represent all feature vectors' scatter degree that is related to the average vector by computing Covariance Matrix. The Covariance Matrix *C* is computed by referring to Equation (2):

$$C = \frac{1}{n}\sum_{k=1}^{n}(X_k - \mu)(X_k - \mu)^T \tag{2}$$

Next, the eigenvalues and their related eigenspace's eigenvectors are calculated by using Equation (3):

$$CV = \lambda \mathrm{V} \tag{3}$$

where *V* is the eigenvectors' set associated with its eigenvalue λ. After calculating them, it is able to use the eigenvalues to sort the eigenvectors in descending order. The eigenvectors associated with the *k* biggest eigenvalues are *k* principal components.

After that, every average centered image will be projected in eigenspace using Equation (4):

$$W_k = V_k^T (X_k - \mu) \tag{4}$$

During the testing stage, every testing image must be average centered and projected onto the same eigenspace as determined in the training stage. In eigenspace, there is a comparison will be made between this projected image and projected training image. Similarity measurements are used to compare images. Finally, it is able to be used for recognizing by the image of training which is most similar to the image of testing.

3.3. Classification: Sparse representation based classification (SRC)

Each classification method has different characteristics which fit various types of data experiments. Face recognition performance also influences by classification method that has been chosen. For this research, classification method that will be used for is Sparse Representation based Classification (SRC) (Wright, J., et al., 2009).

SRC method uses sparse representation for the set of images to correctly classify the testing image through handling a linear equations' system. The sparse representation-based classification method's overall approach can be summarized as four steps. SRC represents the testing data using the training data's linear combination and then calculates the linear representation system's sparse representation coefficients. After that, it uses the coefficients and training data to compute each subject's reconstruction residuals. Lastly, the testing data is able to be then categorized into a class that resulting in the minimum reconstruction residual.

Assume that there are *n* training images, $X = \{x_1, x_2, \ldots, x_n\}$ from *c* classes. Let X_i represent the images from the *i*-th class and the testing image is *y*. Firstly, every image needs to be normalized in order to ensure that have unit l_2-norm. After that, the testing image can be represented by exploiting all the training images' linear combination and the following l_1-norm minimization problem is able to be solved by referring to Equation (5):

$$\alpha^* = \arg\min ||\alpha||_1 \ \ s.t. ||y - X\alpha||_2^2 \leq \varepsilon \tag{5}$$

Then, calculate each class's representation residual by using Equation (6):

$$r_i = ||y - X_i \alpha_i^*||_2^2 \tag{6}$$

where α_i^* represents the representation coefficients vector associated with the *i*-th class.

Lastly, the test image's identity *y* can be produced by determining with the Equation (7):

$$label(y) = \arg\min_i(r_i) \tag{7}$$

3.4. Performance evaluation metric

Performance evaluation metric is also one of the necessary parts for every face recognition process pipeline. It is used to monitor and measure the performance of a model during training and testing. It evaluates the result and tells how accurate is that the face recognition experiment performed. Every face recognition experiment needs an evaluation metric to judge the performance. By analyzing the performance, one can identify the factors that contribute to improve the accuracy of face recognition.

For this research paper, the classification report had been included to monitor and measure the performance of the classification models after fitting. There are several main classification metrics are displayed in classification report including accuracy, precision, recall, f1-score, macro average and weighted average. It will be clearly to see how accurate is that the face recognition experiment performed.

3.5. Precision, recall and F1-score

Precision refers to the classifier's capacity to avoid labeling a sample as positive which is actually negative whereas recall refers to the classifier's ability to locate all actual samples that is positive. To evaluate model performance comprehensively, both precision and recall should be examined. Thus, f1-score takes both of them into account to serves as a useful metric. F1-score is harmonic mean of precision and recall for a more balanced summarization of model performance. It can also call as fmeasure which is a hybrid metric for unbalanced classes. In the case of multi-class classification, averaging methods had been adopted for f1-score calculation, resulting in a set of different average scores such as macro average and weighted average in the classification report. The precision, recall and f1-score is able to calculate by Equation (8), (9), (10) below:

$$Precision = TP / (TP + FP) \tag{8}$$

$$Recall = TP / (TP + FN) \tag{9}$$

$$F1\text{-}Score = 2 * ((Precision * Recall) / (Precision + Recall)) \tag{10}$$

3.6. Macro average and weighted average

Macro averaging is the most straightforward among the numerous averaging methods. The macro average is computed by taking the arithmetic mean also known as unweighted mean of all the per-class. This method treats all classes equally regardless of the support values. However, the weighted average can be computed through averaging all per-class responses with taking into account every class's support. In dataset, the class’s amount of actual instances is called as support. ‘Weight’ essentially refers to the proportion of each class’s support relative to the sum of all support values. With weighted averaging, the output average would have accounted for the contribution of each class as weighted by the number of examples of that given class.

3.7. Accuracy

A metric for assessing classification models is accuracy. Accuracy refers to the percentage of correct predictions made by the model. To be precisely, the face recognition accuracy can be evaluated according to the correctly classified face samples per the total number of testing face samples of the same person. The formula can be written as the following Equation (11):

$$Accuracy = Number\ of\ correct\ predictions / Total\ number\ of\ predictions \tag{11}$$

Besides that, the accuracy of binary classification is also able to be computed in terms of positives and negatives by using Equation (12):

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \tag{12}$$

3.8. Experiment settings

Systematic experiments are executed using benchmark facial dataset which is Face Recognition Technology (FERET) dataset to examine the proposed method’s performance (Phillips, P. J., et al., 2000). The dataset was collected in a controlled condition and the images were taken in uniform context and illumination. The condition of the environment is normally to appear consistent during the process of collecting the face images. This kind of dataset will provide the exact overall classes quantity, class label and quantity of images per class. There are 11338 images of 1199 individuals were collected in this dataset which in different positions and at different times. Fig 2 shows some FERET dataset's clipped sample images. Every column consists of images in identical subject under three different illuminations and expression conditions.

Fig. 2: Sample Images from FERET dataset.

The FERET dataset's face photos were taken in a semi-controlled setting, with the identical technical configuration used in every shooting session to ensure uniformity throughout the data collection. The face images were captured under different face expressions, lightnings, poses and angles. Hence it would be more challenging compared to other standard benchmark datasets such as ORL and AR dataset. The latter did not consider various environment conditions and would not reflect the challenges that might be arise in real applications. In addition, those face images of a subject were usually taken separately on different days. However, there was some minor variation in images collected on different dates due to the device needed to be rebuilt in every session. For the purpose of studying the appearance changes, facial expression and controlled pose variation of the individuals, some subjects being photographed multiple times and their first and last photography could be over two years elapse.

For the experimental analysis, a subset of original FERET dataset will be taken instead of using the whole dataset to perform the face recognition experiment. There are 2,000 images in total are chosen from the original FERET dataset at random in experimental setup of the proposed work. Those images will be divided into 200 classes which consists of 10 images in each class. The experiment was performed following 80% and 20% proportion for the train-test split.

In this experiment setup, the standard dataset is trained under the limited protocol in which no external data is used. The experiments were executed to assess the performance as well as superiority of the proposed method according to the benchmark dataset. Table 1 summarizes the experimental settings of FERET dataset.

Table 1. Summary of dataset experimental settings.

Dataset	FERET
Total Images	2000

Number of Subjects	200
Total Images per Subject	10
Number of Training Images per Subject	8
Number of Testing Images per Subject	2
Total Training Images	1600
Total Testing Images	400

In the experiment, the raw images will be cropped and resized with the dimension of 61 * 73 pixels. Fig 3 shows a sample image after pre-processing. In addition, Difference of Gaussians (DoG) filter is used to improve each facial image's quality while also suppressing Gaussian noise simultaneously. To be objective in assessing the proposed method's performance and each realization, no external training sample is used in the proposed method. Moreover, there is no any commercialized tool or pre-processing is used in this experiment.

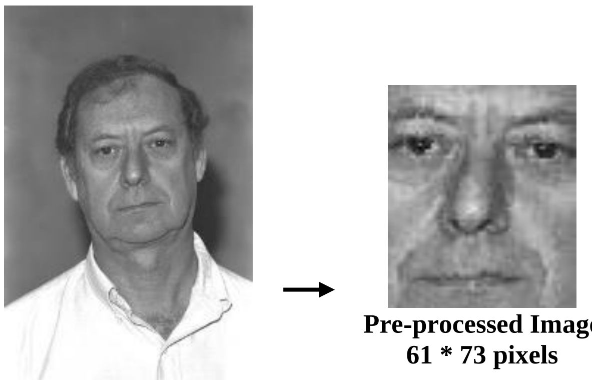

Fig. 3: Sample of pre-processed image.

4. Performance Evaluation and Findings

4.1. Experiment results

In this section, the experiment for this research paper had already been conducted successfully and the evaluation results of applying three different classification methods which are SRC, Euler SRC and SVM combined with feature extractor which is PCA for recognition based on the FERET Dataset had also been analysed and reported.

4.2. Sparse representation based classification (SRC)

After SRC model had been built and fitted, the prediction can be done by using the testing set. Hence, the performance of the SRC model is able to evaluate after doing the prediction. Fig 4 shows the SRC classification report regarding object 187 to 199 and its total accuracy, macro average and weighted average.

	precision	recall	f1-score	support
187	1.00	1.00	1.00	2
188	1.00	1.00	1.00	2
189	1.00	1.00	1.00	2
190	1.00	1.00	1.00	2
191	1.00	1.00	1.00	2
192	1.00	1.00	1.00	2
193	1.00	1.00	1.00	2
194	1.00	1.00	1.00	2
195	0.67	1.00	0.80	2
196	1.00	1.00	1.00	2
197	1.00	1.00	1.00	2
198	1.00	1.00	1.00	2
199	1.00	1.00	1.00	2
accuracy			0.99	400
macro avg	0.99	0.99	0.99	400
weighted avg	0.99	0.99	0.99	400

Fig. 4: SRC Classification report.

4.3. Euler sparse representation based classification (Euler SRC)

For Euler SRC model had been built and fitted, the parameters including lamb and alpha were set to 0.5 and 1.9 respectively. Fig 5 shows the Euler SRC classification report regarding object 187 to 199 and its total accuracy, macro average and weighted average.

	precision	recall	f1-score	support
187	1.00	0.50	0.67	2
188	1.00	1.00	1.00	2
189	0.67	1.00	0.80	2
190	0.67	1.00	0.80	2
191	0.67	1.00	0.80	2
192	1.00	1.00	1.00	2
193	1.00	1.00	1.00	2
194	1.00	1.00	1.00	2
195	1.00	1.00	1.00	2
196	1.00	1.00	1.00	2
197	1.00	1.00	1.00	2
198	0.67	1.00	0.80	2
199	0.67	1.00	0.80	2
accuracy			0.80	400
macro avg	0.78	0.80	0.77	400
weighted avg	0.78	0.80	0.77	400

Fig. 5: Euler SRC classification report.

4.4. Support vector machine (SVM)

Similarly, the testing set had been done the prediction can evaluate the performance. However, there are different hyperparameter values had been tuned to train the SVM model. Table 2 summarizes the accuracy with different gamma value of SVM model.

Table 2. Accuracy with different gamma values

Gamma Value	Accuracy
1	0.79
0.01	0.91
0.001	0.97

Based on the table 2, the accuracy is keep increasing when decrease the gamma value of the model. The highest accuracy is 97% which gamma value is equal to 0.001. Thus, the performance had been evaluated using the highest accuracy model. Fig 6 shows the SVM classification report of object 187 to 199 and its total accuracy, macro average and weighted average.

	precision	recall	f1-score	support
187	1.00	1.00	1.00	2
188	1.00	1.00	1.00	2
189	1.00	1.00	1.00	2
190	1.00	1.00	1.00	2
191	1.00	1.00	1.00	2
192	1.00	1.00	1.00	2
193	1.00	1.00	1.00	2
194	1.00	1.00	1.00	2
195	1.00	1.00	1.00	2
196	1.00	1.00	1.00	2
197	1.00	1.00	1.00	2
198	1.00	1.00	1.00	2
199	1.00	1.00	1.00	2
accuracy			0.97	400
macro avg	0.99	0.97	0.97	400
weighted avg	0.99	0.97	0.97	400

Fig. 6: SVM classification report.

4.5. Experiment Results Comparison

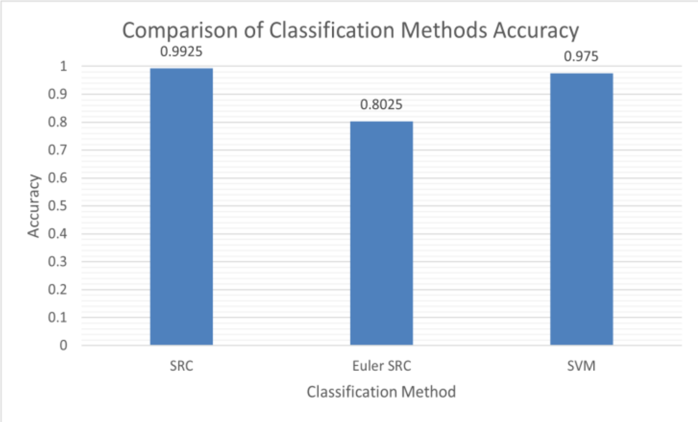

Fig. 7: Comparison of classification methods accuracy.

The three classification methods which are SRC, Euler SRC and SVM had been compared for the accuracy in this research project. The summary result shows that SRC method has the best performance among these three classification methods because it gets the highest accuracy which is almost 99.25% in the experiment. SVM for this experiment had been tuned to achieve 97.5% accuracy. However, Euler SRC method only achieves 80.25% although it also has been considered as a high accuracy, but it is still the lowest if compared to the other two classification methods. It implicates that Euler SRC method may not be suitable on large dataset and it also requires the highest training time when it is compared with the other two classification methods.

4.6. Existing method results comparison

Refer Table 3, the comparison between various methods by using the same FERET dataset is presented. Enhanced Sparse Representation based Classification achieves 82.17% accuracy (Peng, Y., et al., 2018), Sparse Representation and Transfer Learning Method achieves around 95% accuracy (Z. Liu, et al., 2019) while Sparse Representation based Classification on K-Nearest Subspace achieves around 74% accuracy (Mi, J.X., et al., 2013). In addition, Adaptive Boosting (Hao, Z., et al., 2018), Support Vector Machine (Wei, J., et al., 2011) and Convolutional Neural Network (Guo, S., et al., 2016) which is a deep learning method achieves 90.7%, 92% and 99.66% respectively. In comparison with all these methods, the proposed method can achieve 99.25% accuracy.

Table 3. Comparison various method results with FERET dataset.

Authors	Year	Methods/Techniques	Accuracy
Yali Peng, et al.	2018	Enhanced Sparse Representation based Classification	82.17%

Zhi Liu, et al.	2019	Sparse Representation Method and Transfer Learning Method	95%
Jian-Xun Mi, et al.	2013	Sparse Representation based Classification on K-Nearest Subspace	74%
Zeng Hao, et al.	2018	Adaptive Boosting	90.7%
Jin Wei, et al.	2011	Support Vector Machine	92%
Shanshan Guo, et al.	2016	Convolutional Neural Network	99.66%
(Proposed Method)	2022	Sparse Representation based Classification with Principal Component Analysis	99.25%

This encouraging result shows that the proposed work which includes SRC classification method with PCA is steadily performed well in this large dataset. Although it is a hand-crafted method, it still achieves a competitive result comparing with the deep learning methods such as CNN, in this face recognition experiment. This proposed method has been considered as hand-crafted method because it is a traditional machine learning method which performs the face recognition using feature extraction technique and trainable classifier algorithm only without involving any deep learning process. In contrast, deep learning method is a complex algorithm which contains many layers and needs a long time to train. It means that it is complicated and requires huge training data in the execution process. However, the hand-crafted method just only run within 30 minutes and it still can get the 99.25% accuracy result which has the performance that is comparable to the deep learning method.

5. Conclusion

In this research paper, experiments had been conducted to perform the face recognition task in order to analyse the impact of sparse representation. The proposed work comprises the classification method which is Sparse Representation based Classification (SRC) with the most popular feature extraction technique, Principal Component Analysis (PCA) by using a benchmark facial dataset, FERET dataset. Besides that, the classification method such as Euler Sparse Representation based Classification (Euler SRC) and Support Vector Machine (SVM) had also been combined with PCA for comparing with SRC method to measure their performance in face recognition. Based on the experiment results and findings, it proves that through implementing the feature extractor PCA and classifier SRC on facial images can achieve the best overall results which is 99.25% comparing with other classification methods and also able to compete with the state-of-the-art methods in the similar field of study. In future works, the proposed method will be tested under unconstrained datasets such as Labeled Faces in the Wild (LFW) to observe its stability and effectiveness in a more challenging real-world scenario. The execution time will be another performance metric to be analysed so that the complexity can be compared with the state-of-the-art methods.

Acknowledgements

This work was supported by the Internal Research Fund (IR Fund) MMUI/220025, 2022 from Multimedia University.

References

Giron-Sierra J.M. (2017). Sparse representations. *Digital Signal Processing with Matlab Examples,* Volume 3. *Signals and Communication Technology*. *Springer*, *Singapore*.

Guo, S., Chen, S., & Li, Y. (2016). Face recognition based on convolutional neural network and support vector machine. *IEEE International Conference on Information and Automation (ICIA).*

Hao, Z., Feng, Q., & Kaidong, L. (2018). An optimized face detection based on Adaboost algorithm. *2018 International Conference on Information Systems and Computer Aided Education (ICISCAE)*.

Kim, K. (2012). Face recognition using principle component analysis.

L. Li, X. Mu, S. Li & H. Peng. (2020). A Review of Face Recognition Technology. *IEEE Access*, vol. 8, (pp. 139110-139120).

Li Y. (2013). Sparse Representation for Machine Learning. *Advances in Artificial Intelligence. Canadian AI 2013. Lecture Notes in Computer Science*, vol 7884. *Springer, Berlin, Heidelberg.*

Liu, Y., Gao, Q., Han, J., & Wang, S. (2018). Euler Sparse Representation for Image Classification. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1).

Mi, J.X., & Liu, J.X. (2013). Face Recognition Using Sparse Representation-based Classification on K-nearest subspace. *PLoS ONE*, 8(3): e59430.

Pearson, K. (1901). LIII. on lines and planes of closest fit to systems of points in space. *The London*, *Edinburgh*, *and Dublin Philosophical Magazine and Journal of Science*, 2(11), 559–572.

Peng, Y., Li, L., Liu, S., Li, J., & Cao, H. (2018). Virtual samples and sparse representation-based classification algorithm for face recognition. *IET Computer Vision*, 13(2), 172–177.

Phillips, P. J., Moon, H., Rivzi, S. A. & Rauss, P. J. (2000). The FERET Evaluation Methodology for Face-Recognition Algorithms.

Wei, J., Jian-qi, Z., & Xiang, Z. (2011). Face recognition method based on support vector machine and particle swarm optimization. *Expert Systems with Applications*, 38(4), 4390–4393.

Wright, J., Yang, A. Y., Ganesh, A., Sastry, S. S., & Ma, Y. (2009). Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2), 210–227.

Z. Liu, D. Jiang, Y. Li, Y. Cao, M. Wang & Y. Xu. (2019). Automatic Face Recognition Based on Sparse Representation and Extended Transfer Learning. *IEEE Access*, 7, 2387-2395.

Zhang, Z., Xu, Y., Yang, J., Li, X., & Zhang, D. (2015). A Survey of Sparse Representation: Algorithms and Applications. *IEEE Access*, *IEEE*, 3.