3D Face Recognition using Kernel-based PCA Approach

Marcella Peter, Jacey-Lynn Minoi and Irwandi Hipni Mohamad Hipiny

Faculty of Computer Science and Information Technology,
Universiti Malaysia Sarawak, Malaysia.
marcellapeter@yahoo.com

Abstract. Face recognition is commonly used for biometric security purposes in video surveillance and user authentications. The nature of face exhibits non-linear shapes due to appearance deformations, and face variations presented by facial expressions. Recognizing faces reliably across changes in facial expression has proved to be a more difficult problem leading to low recognition rates in many face recognition experiments. This is mainly due to the tens degree-of-freedom in a non-linear space. Recently, non-linear PCA has been revived as it posed a significant advantage for data representation in high dimensionality space. In this paper, we experimented the use of non-linear kernel approach in 3D face recognition and the results of the recognition rates have shown that the kernel method outperformed the standard PCA.

Keywords: 3D face, Facial recognition, Kernel PCA.

1 Introduction

Face recognition is a biometric system aim to identify a person in a digital image by analyzing and comparing facial patterns. Face recognition process may be a simple task for human but it is a challenge and difficult task for the computer as it requires the involvement of a combination of statistical approaches, Artificial Intelligent, computer vision and machine learning methods. As mentioned in [1], face recognition rates dropped when facial expression is included in a system. In addition, other facial variants and factors such as pose, facial expression, hairstyle, makeup, mustache, beard and wearing glasses would also contribute to the low recognition rate [2].

For the past two decades, various face recognition techniques were developed by researchers with the aim to improve the performance of face recognition. The general pipeline for an automated face recognition system involves four steps: facial image acquisition and detection, face feature extraction, and face classification and recognition [3, 4]. Facial image acquisition and detection are usually done in the pre-processing phase. This is followed by facial feature extraction process, where features such as the eyes, the nose, and the mouth are extracted and classified. Facial features are important in many face-related applications, as in face alignment and feature extraction. The conventional method in feature extraction commonly used linear transformation approach, i.e. the Principal Component Analysis (PCA). PCA is a common

© Springer Nature Singapore Pte Ltd. 2019

R. Alfred et al. (eds.), *Computational Science and Technology*, Lecture Notes in Electrical Engineering 481, https://doi.org/10.1007/978-981-13-2622-6_8

and yet powerful dimensionality reduction technique for extracting and projecting data structures into lower subspaces. It is readily performed by solving eigenvalues and eigenvectors to extract discriminant principal components [6].

1.1 From PCA to Kernel-based PCA

PCA is an effective technique commonly used in face recognition study for discriminant analysis. PCA is widely implemented in 2D and 3D face recognition systems as the base method [1, 17, 18]. Depth information in 3D data makes face recognition system to become more robust because it does not depend on illumination and pose [3]. Recent work by [1] implements 3D model-based face recognition system that also utilizes PCA method to recognize person under any facial expressions. However, the limitation is that PCA could not deal with large range of face variants. In other words, to simplify object with complex structures in a linear subspace, PCA might not be sufficient enough [6]. 3D face data has a complicated structure in a high dimensional space. The nature of non-linear shape exists when the face images are projected under different conditions with different lighting, expression variant, identity variant, and poses that lead to shape deformations [7, 8]. Due to this matter, a kernel approach is proposed to overcome the non-linearity. The idea of kernel is to map the input face images into a higher dimensional space in which the manifold of the face linearity is simplified.

In comparison to other non-linear techniques of feature extraction, kernel method does not require non-linear optimization, and only the solution from eigenvalue measurements [9]. Kernel learns non-linear functions while working with several training examples. Kernel can make the data linearly separable by projecting the data onto a higher dimensional feature space and we use a kernel function to do this. By choosing different form of kernel function, it can handle different non-linear problems. Kernel functions that most researchers adopt are the Polynomial kernel and Gaussian kernel. Kernels have been successfully used in Support Vector Machine (SVM) for classification purpose. One of the reason for SVM [10] success is the 'kernel trick' implicitly reduce computations of data in feature space of each input vector and it directly gives result into the feature space [9, 13].

Schölkopf et al. [11] extended classical linear PCA to Kernel PCA method. Kernel PCA projects the input space into feature space, by mapping the principal component vectors using kernel function. Recent work by Wang [12] has applied Kernel PCA to explore the complicated structure of 2D image for face recognition. Pre-image is reconstructed using Gaussian kernel PCA and then use it to design Kernel-based Active Shape Model [12]. Their results have shown that the error rate is lower with 11.54% for Kernel PCA compared to linear PCA with 23.06%.

2 Related Works

Researches in 3D face recognition are mostly using linear approaches and techniques [1, 17, 18]. There also exist non-linear approaches and they are being used and improved for face applications. Recent works by [19] used discriminative depth estimation approach that adopts convolutional neural network, which retains the subjects' discriminant information and multilayer convolutional network, to estimate the depth information from a 2D face image. Their approach investigates the manifold of network layers from a single color image to extract new space, where in this case the depth space, for discriminant features to be extracted using the network to improve the 2D face recognition.

A similar non-linear approach by [20] proposed a 3D morphable model that used three deep neural networks aim to reconstruct the 3D face data input. The proposed fitting algorithm works by estimating network encoder from decoded two parameters shape and texture, with the assistance of a geometry-based rendering layer. The proposed works are directed towards applications in 3D facial recognition and facial synthesis. However, these two approaches require a large number of parameters which may lead to overfitting of training data when implemented into a small number of data. Therefore optimization such as using non-linear regression approximation [21] is needed to resolve the problem. The idea to extract hidden layers without high computation and taking the advantage of kernel methods as indicated in [9] motivates this research to explore Kernel PCA method to extract non-linear principle components from a 3D data, which is more practical to work on with a small number of dataset.

This project will use existing 3D face dataset. The project hence move on to the analysis of a statistical model that can fit 3D face data points as input data to define a shape space by selecting basis direction that holds the greatest variance of a face data where covariance matrix is then computed for the face data to be projected into the shape space. The paper will present the non-linear Kernel PCA approach on 3D face datasets and its experimental analysis by comparing the recognition rates with the baseline.

3 Kernel-based PCA

The algorithm of Kernel-based PCA is adopted and can be found in [14]. Given a set of m centered or with zero mean samples

$$x_k = [x_k, \dots, x_{kn}]^T \in R^n, \tag{1}$$

The purpose of PCA is to find the directions of projection that get the most out of the variance C, which is corresponding to finding the eigenvalues from the covariance matrix

$$\lambda w = Cw \tag{2}$$

for eigenvalues $\lambda \geq 0$ and eigenvectors $w \in R^n$. In Kernel PCA, each vector x is projected from the input space, R^n, to a high dimensional feature space, R^f, by a nonlinear mapping function:

$$\Phi: R^n \to R^f, f \gg n . \tag{3}$$

Note that the dimensionality of the feature space can be huge. In R^f, the corresponding eigenvalue problem is $\lambda w^{\Phi} = C^{\Phi} w^{\Phi}$ where $C^{\Phi}$is a covariance matrix. All solution w^{Φ} with $\lambda \neq 0$ lie in the span of $\Phi, \dots, \Phi(x_m)$ and there exist coefficients $a_i$such that

$$w^{\Phi} = \sum_{i=1}^{m} a_i \Phi(x_i) \tag{4}$$

Denote an $m \times m$ matrix K by

$$K_{ij} = K\left(\Phi_i, \Phi_j\right) = \left(\Phi_i\right).\left(\Phi_j\right), \tag{5}$$

the Kernel PCA problem becomes

$$m\lambda K a = K^2 a \tag{6}$$

$$m\lambda a = K a \tag{7}$$

Where a denotes a column vector with entries $a_1, \dots, a_m$. The above derivation assumes that all the projected samples $\Phi(\mathrm{x})$ are centered in R^f. As to centralize the vectors $\Phi(\mathrm{x})$ in R^fcan be found in [9]. We can now project the vectors in R^f to a lower dimensional space spanned by the eigenvectors w^{Φ}. Let x be a test sample whose projection is $\Phi(\mathrm{x})$ onto w^{Φ}, is the nonlinear principal components corresponding to

$$\Phi: w^{\Phi}.\Phi(\mathrm{x}) = \sum_{i=1}^{m} a_i\left(\Phi(x_i)\,.\,\Phi(x_i)\right) = \sum_{i=1}^{m} a_i\,(x_i, x) \tag{8}$$

The first q $(1 \leq q \leq m)$ nonlinear principle components or the eigenvectors w^{Φ} are extracted using the kernel function without expensive operation that explicitly projects samples to high dimensional space R^f. The first q components correspond to recognition where each x encodes a face image.

According to [15], first order polynomial kernel of Kernel PCA is a special case of traditional PCA. The polynomial kernel can be expressed as

$$K(x, y) = (x^T y)^d \tag{9}$$

where, the power of d is specified or formed beforehand by the user.

4 Experiment and Results

An experiment is conducted using Kernel PCA on 3D face dataset to demonstrate its effectiveness and compared results obtained with linear PCA. The Kernel PCA is used as feature extraction through matrix decomposition to extract nonlinear principal components (eigenvectors) and K^{th} Nearest Neighbor classifier is used to measure the Euclidean distance as classification mechanism.

4.1 Datasets

The experiment is carried on Imperial College London 3D face database that contains 240 face surface models of 60 subjects with 4 females and 56 males. The subjects were also classified in terms of their ethnicity as 8 South Asians, 6 East Asians, 1 Afro-Caribbean and 45 Caucasians. These subjects were mostly students within an age range of 18-35 years. The facial, and was acquired in several different head positions and three facial expression poses. The facial expressions were smiling, frowning, and neutral. The 3D face dataset has already been preprocessed, thus preprocessing stage is omitted. Details of the preprocessing can be found in [22]. Fig. 1 shows the example images adopted from Imperial College London in 2D (left) and 3D (right) environment.

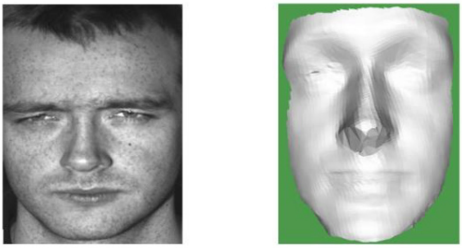

Fig. 1. Sample from Imperial College London face database [15]

4.2 Relationship between distance measurement to face recognition result

The PCA face recognition system is normally implemented along with K^{th} Nearest Neighbor (KNN) algorithm. The algorithm finds the closest K neighbors with minimum distance from a subspace to classify a testing set. Different parameters can be used with KNN such as different value of K and distance model. The K value in KNN during the recognition process affects the overall face recognition rate. As the recognition process is based on the shortest distance, therefore the face recognition rate varies based on different type of distance classifier such as Euclidean distance and Manhattan distance which are used widely in evaluating the rate of facial recognition. In this project, Euclidean distance is chosen simply because of the idea to measure the shortest distance, for example the length between the straight lines of two points. In general, the distance between two points of x and y, in Euclidean space, R^n is given by:

$$d(x,y) = \|x - y\| = \sqrt{(|x_i - y_i|)^2} \quad (10)$$

4.3 Experimental Plan

The 3D face recognition experiment started with training and followed by recognition process. Herewith, a proposed experiment procedure by using cross validation approach. The experiment begins with two of the face class generated from 3D face dataset are selected and used as a training set and testing set for face recognition. As shown in Table 1, there are four sets of 3D face database that are classified according to different expressions. The four classes are named as neutral 1, neutral 2, frowning and smiling, with 60 subjects for each set. As mentioned in [16], the use of different K value for different instances may give positive outcome for classification accuracy. Thus, we followed the approach proposed in [16] which allows selection of a local value of K to find the best classification. In this experiment, the K value = 3 is used, which yields the lowest number of error during classification. Equation (11) is used to evaluate the rate of recognized face for each testing set.

$$Face\ recogntion\ rate = \frac{Correct\ Match\ Count}{Total\ face\ test\ count} \times 100\ \% \tag{11}$$

Table 1. Number of samples of each face dataset.

Face class	Number of sample
Neutral 1	60
Neutral 2	60
Frowning	60
Smiling	60

Training. The face recognition system read each subject from the training dataset selected and extract 3D points based on their referred directory path. Next, each training subject with their label using Kernel PCA is projected into the training subspace, which acts as a knowledge base for the recognition process. This process is iterated until all 60 subjects are projected into the training subspace.

Recognition. The face recognition system read each subject from the 3D testing dataset selected and extract 3D points based on their referred directory path. Next, each testing subject is projected into the training subspace using Kernel PCA. From the feature vectors obtained from the subspace projection, determine Euclidean distance for each test subject and the training subjects based on the selected K value (3 NN). After that, the distance between the testing subjects with the training subjects are compared. Then, the testing subject is classified as the same label as the training subject that has the shortest distance. The process is iterated until all 60 subjects in testing dataset are tested. To validate the newly classified subjects, cross validation is performed and the rate of recognition is computed. Then, the result is displayed with the recognition rate computed earlier. The test is repeated by using different testing sets.

The experiments are arranged as Test 1: Frowning, Test 2: Neutral 1, Test 3: Neutral 2, and Test 4: Smiling. The following Table 2 presents the experiment result for Test 1. The results of face recognition using Kernel PCA are compared with PCA results, which has been set as the baseline approach in this research.

Table 2. Face recognition rate for Test 1 (%).

Test 1: Frowning (3-NN)		
Training Set	**Kernel PCA**	**PCA**
Frowning	98.33	100.00
Neutral 1	90.00	85.00
Neutral 2	83.33	76.67
Smiling	36.67	20.00
Total average rate:	**77.08**	**70.42**

4.4 Results and Analysis

Table 3 presents the average rate of 3D face recognition based on Kernel PCA and PCA, respectively. Based on the results, we found that by using Neutral 1 and Neutral 2 as the training set, both results would give the highest recognition as compared to Frowning and Smiling. Meanwhile for comparison purposed between Kernel PCA and PCA method, the average recognition rate shows that Kernel PCA achieved higher recognition rate than PCA.

Using frowning test set in Kernel PCA earns a recognition rate of 77.08%, while 70.42% of recognition rate was achieved when using PCA. Besides that, Neutral 1 test set in Kernel PCA could achieved a recognition rate of 82.50% meanwhile with PCA only earns 78.33% of recognition rate. Using Neutral 2 test set, Kernel PCA achieved a recognition rate of 85%, which is much higher than PCA that earns 77.92% of recognition rate. In the last experiment of the test set, smiling test set have shown 64.59% of recognition rate using Kernel PCA and PCA achieved 47.08% of the recognition rate. Fig. 2 illustrates the comparison of both methods. The overall recognition rate for Kernel PCA is 77.29%, which is higher than PCA with 52.69%.

Table 3. Average face recognition rate of two methods (%)

	Frowning	Neutral 1	Neutral 2	Smiling
Kernel PCA	77.08	82.50	85.00	64.59
PCA	70.42	78.33	77.92	47.08

Fig. 2. Recognition rate of PCA compared to Kernel PCA (%).

4.5 Discussion

The comparison between Kernel PCA and PCA methods is as presented in Fig. 2 shows that Kernel PCA could achieved a higher recognition rate than PCA technique alone. We have also found out that training the neutral face will always provide higher face recognition rate if compared to using frowning and smiling faces. This is because neutral faces have lesser face variant compared to frowning and smiling. Smiling has more facial variances at the lower part which involve cheek and mouth. Meanwhile, frowning has more variances over top part of a face namely, the eyes and eyebrows. Therefore, in most of literature, neutral faces are preferably used for its lesser face variants.

From the experiment, we have also identified two misclassified faces using PCA. The two subjects (see Fig. 3) are incorrectly matched when Neutral 1 training sample is used on Neutral 2 testing sample. Based on the observation made in Fig. 3, the 2D face on the top left and 3D face at the bottom left is the same person but the recognition system has wrongly matched the face to the one on the right side. This could be due to the similarity of facial structures between the two subjects. However, when tested using Kernel PCA, the two faces were correctly matched. This shows that Kernel PCA could correctly extract facial features. However, in shown Table 2, when Frowning test set is tested with Frowning training set using Kernel PCA method, the recognition rate recorded at 98.33%. Based on the analysis, the test subject is correctly classified but with false rejection. This limitation problem will be further investigated in the future work.

The experiment has proved that Kernel PCA as a non-linear approach, able to extract the non-linearity property within face dataset to yield a higher face recognition rate compared to PCA. Noticed that Kernel PCA had increased the face recognition rate to 24.6% compared to PCA (see Fig. 2).

Fig. 3. Incorrect match of two subjects.

5 Conclusion and Future Works

This paper described the application of Kernel-based PCA approach in face recognition domain. Kernel is an inherently multi-disciplinary domain. It is vital to look at it from all fields and perspectives to have an insight on how to develop an efficient automatic 3D face recognition system. The result of the overall recognition rate for Kernel PCA is 77.29%, while PCA has 52.69%. Experiment results proved that Kernel-based PCA approach outperforms linear PCA in 3D face recognition. The future work will focus on analyzing facial recognition method using other 3D face datasets, testing out other kernel methods and further investigate factors such as modifying the number of principal components and number of training samples.

Acknowledgement. The authors would like to thank Suriani Ab Rahman, Phoon Jai Hui for contributing to the research in this paper, and Newton Fund for the financial support for the publication.

References

1. Agianpuye, A. S., & Minoi, J. L.: Synthesizing neutral facial expression on 3D faces using Active Shape Models. In *Region 10 Symposium, 2014 IEEE* (pp. 600-605). IEEE. (2014).
2. Wen, Y., Lu, Y., Shi, P., & Wang, P. S.: Common Vector Based Face Recognition Algorithm. In *Pattern Recognition, Machine Intelligence and Biometrics* (pp. 335-360). Springer Berlin Heidelberg. (2011).

3. Hassabalah, M. & Aly, S.: Face Recognition: Challenges, Achievements and Future Directions. In IET Computer Vision Journals, Vol. 9, Iss. 4, pp. 614-626. (2015).
4. Chen, W., Yuen, P. C., Fang, B., & Wang, P. S.: Linear and Nonlinear Feature Extraction Approaches for Face Recognition. In *Pattern Recognition, Machine Intelligence and Biometrics* (pp. 485-514). Springer Berlin Heidelberg. (2011).
5. M. Kirby & L. Sirovich.: Application of the Karhunen-Lòeve procedure for the characterization of human faces. IEEE Transactions on Pattern Analysis and Machine Intelligence 12(1):103-108. (1990).
6. Devi, R., B., Laishram, R., & Singh, Y., J.: Modelling Objects Using Kernel Principal Component Analysis. ADBU Journal of Engineering Technology 2.1. (2015).
7. Shah, J., H., et al.: A Survey: Linear and Nonlinear PCA Based Face Recognition Techniques. Int. Arab J. Inf. Technol. 10.6: 536-545. (2013).
8. Lee, C, S. & Elgammal, A.: Non-linear Factorized Dynamic Shape and Appearance Model for Facial Expression Analysis and Tracking. In IET Computer Vision, Vol. 6, Iss. 6, pp. 567-580. (2012).
9. Schölkopf, B., Smola, A., & Müller, K. R.: Nonlinear component analysis as a kernel eigenvalue problem. *Neural computation*, *10*(5), 1299-1319. (1998).
10. Schölkopf, B., Sung, K., Burges, C., Girosi, F., Niyogi, P., Poggio, T. & Vapnik. V.: Comparing support vector machines with gaussian kernels to radial basis function classifiers. IEEE Trans. Sign. Processing, 5:2758 –2765. (1997).
11. Schölkopf, B., Smola, A., & Müller, K. R.: Kernel principal component analysis. In *International Conference on Artificial Neural Networks* (pp. 583-588). Springer Berlin Heidelberg. (1997).
12. Wang, Q.: Kernel principal component analysis and its applications in face recognition and active shape models. *arXiv preprint arXiv:1207.3538*. (2012).
13. Alaíz, C. M., Fanuel, M., & Suykens, J. A.: Convex Formulation for Kernel PCA and Its Use in Semisupervised Learning. *IEEE Transactions on Neural Networks and Learning Systems.* (2017).
14. Yang M. H.: Face Recognition Using Kernel Methods. Advances in Neural Information Processing Systems. MIT Press, 13: 960 – 966. (2001).
15. Imperial College London 3D face database. (n.d).
16. García-Pedrajas, N., del Castillo, J. A. R., & Cerruela-García, G.: A Proposal for Local k Values for k-Nearest Neighbor Rule. *IEEE transactions on neural networks and learning systems*, *28*(2), 470-475. (2017).
17. Okuwobi, I. P., Chen, Q., Niu, S., & Bekalo, L.: Three-dimensional (3D) facial recognition and prediction. Signal, Image and Video Processing, 10(6), 1151-1158. (2016).
18. Ouamane, A., Chouchane, A., Boutellaa, E., Belahcene, M., Bourennane, S., & Hadid, A.: Efficient tensor-based 2d+ 3d face verification. *IEEE Transactions on Information Forensics and Security*, 12(11), 2751-2762. (2017).
19. Cui, J., Zhang, H., Han, H., Shan, S., & Chen, X.: Improving 2D face recognition via discriminative face depth estimation. *Proc. ICB*, 1-8. (2018).
20. Tran, L., & Liu, X.: Nonlinear 3D Face Morphable Model. *arXiv preprint arXiv:1804.03786.* (2018).
21. Villarrubia, G., De Paz, J. F., Chamoso, P., & De la Prieta, F.: Artificial neural networks used in optimization problems. *Neurocomputing*, *272*, 10-16. (2018).
22. Papatheodorou, T.: 3d face recognition using rigid and non-rigid registration. PhD Thesis, Imperial College. (2006).