TELKOMNIKA, Vol.10, No.1, March 2012, pp. 113~118
ISSN: 1693-6930
accredited by DGHE (DIKTI), Decree No: 51/Dikti/Kep/2010

Efficient Kernel-based Two-Dimensional Principal Component Analysis for Smile Stages Recognition

Rima Tri Wahyuningrum*, Fitri Damayanti
Informatics Engineering Department, University of Trunojoyo Madura
Raya Telang PO BOX 2 Kamal, Bangkalan, Indonesia
e-mail: rima_tw@yahoo.co.id, fitri2708@yahoo.com

Abstrak

Akhir-akhir ini, pendekatan analisis komponen utama dua dimensi (2DPCA) telah diusulkan untuk representasi dan pengenalan taraf senyuman. Esensi dari 2DPCA adalah menghitung vector eigen atau yang disebut matrik kovarian tanpa mengkonversi matrik citra ke bentuk vektor sehingga ukurannya lebih kecil, lebih mudah mengevaluasi kovarian matriknya, komputasinya berkurang dan unjuk kerjanya juga meningkat dibandingkan PCA tradisional. Usaha untuk meningkatkan dan menyempurnakan unjuk kerja pengenalan taraf senyuman. Pada makalah ini diusulkan konsep kernel 2DPCA yang efisien untuk mengembangkan struktur non linier pada data masukan. Makalah ini mendiskusikan perbandingan algoritma 2DPCA berbasis Kernel standar dan 2DPCA berbasis Kernel efisien untuk pengenalan taraf senyuman. Dari hasil ujicoba didapatkan 2DPCA berbasis Kernel memiliki akurasi yang lebih bagus daripada pendekatan lainnya. Sedangkan penggunaan 2DPCA berbasis Kernel efisien mampu mempercepat prosedur pelatihan 2DPCA berbasis Kernel standar, sehingga algoritma ini menghasilkan waktu komputasi lebih efisien dan menghemat penggunaan memori dibandingkan dengan 2DPCA berbasis Kernel standar.

Kata kunci: *2DPCA, ekstraksi cirri, Kernel efisien, Kernel standar, pengenalan taraf senyuman*

Abstract

Recently, an approach called two-dimensional principal component analysis (2DPCA) has been proposed for smile stages representation and recognition. The essence of 2DPCA is that it computes the eigenvectors of the so-called image covariance matrix without matrix-to-vector conversion so the size of the image covariance matrix are much smaller, easier to evaluate covariance matrix, computation cost is reduced and the performance is also improved than traditional PCA. In an effort to improve and perfect the performance of smile stages recognition, in this paper, we propose efficient Kernel based 2DPCA concepts. The Kernelization of 2DPCA can be benefit to develop the nonlinear structures in the input data. This paper discusses comparison of standard Kernel based 2DPCA and efficient Kernel based 2DPCA for smile stages recognition. The results of experiments show that Kernel based 2DPCA achieve better performance in comparison with the other approaches. While the use of efficient Kernel based 2DPCA can speed up the training procedure of standard Kernel based 2DPCA thus the algorithm can achieve much more computational efficiency and remarkably save the memory consuming compared to the standard Kernel based 2DPCA.

Keywords*: 2DPCA, efficient Kernel, feature extraction, smile stages recognition, standard Kernel*

1. Introduction

Feature extraction is the key problem to the pattern recognition task. The aim of feature extraction is to project the high dimensional sample data onto the optimal projection matrix and yield the low dimensional feature data as representative as possible. Principal component analysis (PCA) is a classical feature extraction and data representation technique widely used in the areas of pattern recognition and computer vision [1],[2]. In PCA, the technique maps the sample vectors considered as a point in high dimensional space into low dimensional subspace. While in 2DPCA, an image covariance matrix can be constructed directly using the original image matrices and this overcomes the weaknesses of PCA in which 2D face image matrices must be previously transformed into 1D image vectors. As a result, 2DPCA has three important advantages over PCA. First, it is easier to evaluate the covariance matrix accurately. Second, less time is required to compute the corresponding eigenvectors. The last one is that 2DPCA

Received August 10^{th}, 2011; Revised January 4^{th}, 2012; Accepted January 9^{th}, 2012

can effectively avoid the small sample size (SSS) problem, which will achieve good recognition accuracy when only one sample is contained in each class. Further, the performance of 2DPCA is usually better than PCA as demonstrated in Yang [3].

In previous studies many research have discussed about face recognition, but only a few research have discussed about smile stages recognition. Smile stages recognition is a part of Aesthetic Dentistry on Orthodontic Rehabilitation. This is also a part of face expression recognition. Most well known appearance-based smile stages recognition methods are based on feature extraction techniques such as principal component analysis (PCA) and linear discriminant analysis (LDA) [4], 2DPCA [5], 2DLDA [6], kernel Laplacian-lips [7]. Meanwhile, many serious studies have been conducting about the smile expression and/or recognition or detection. Philips presented the classification of smile patterns to identify various smile patterns in dentistry [8]. Whitehill *et al* pointed out that the smile expression of a face, in the form of an image and a video can be detected [9]. Furthermore, Wojdel and Rothkrantz proposed the combination of fuzzy system and artificial neural network to recognize the expression of oral features [10].

Kernel method is one of the nonlinear data classifier methods that have been successfully implemented on machine learning algorithm, like Kernel Principal Component Analysis (KPCA) [11-13], Kernel Fisher Discriminant (KFD) [14], and Kernel Independent Component Analysis (KICA) [15]. KPCA is the non-linear generalization of the conventional PCA via the kernel trick, which is a powerful method for classification and regression application. Likewise, the Kernelization of 2DPCA can be benefit to develop the nonlinear structures in the input data [16], [17]. Because the 2DPCA is working on the row of the sample image, the non linear generation of 2DPCA need map each row of samples onto feature space via kernel trick. As a result, the dimension of kernel matrix of training data is very high, and the procedure of diagonalizing the kernel matrix is quite time consuming. To overcome the suffering of computation cost, in this research, we present an efficient kernel based 2DPCA for smile stages recognition.

This research classifies smile stages pattern based on the smile scenarios to distinguish the smiling stage faces at pattern I, pattern III and pattern IV. We employ 30 face images for each smile stage pattern. The result of research can be used to support medical fields, such as complicated oral facial surgeries, periodontal fractions, and traumatic tooth fractures. This research will also be very useful for patients before and after surgery.

2. The Algorithm Research

2.1. Two Dimensional PCA (2DPCA)

In 2D approach, the image matrix does not need to be previously transformed into a vector, so a set of M sample images is represented as $\{ P_1, P_2, \ldots, P_M \}$ with $P_i \in R^{kxs}$, which is a matrix space of size kxs. The total scatter matrix is redefined as

$$G_t = \sum_{i=1}^{M} (P_i - \mu_P)^T \, (P_i - \mu_P) \tag{1}$$

with $\mu_P = \frac{1}{M} \sum_{i=1}^{M} P_i \in R^{kxs}$ is the mean image of all samples. $G_t \in R^{sxs}$ is also called image covariance (scatter) matrix. A linear transformation mapping the original kxs image space into a kxn feature space, where $n < s$. The new feature matrices $Q_i \in R^{kxn}$ are defined by the following linear transformation:

$$Q_i = (P_i - \mu_P) \, W \in R^{kxn} \tag{2}$$

where i = 1, 2, …, M and $W \in R^{sxn}$ is a matrix with orthonormal columns. In 2DPCA, the projection W_{opt} is chosen to maximize $tr(W^T G_t W)$. The optimal projection $W_{opt} = [w_1 \, w_2 \, \ldots \, w_n]$ with $\{w_i \mid i = 1, 2, \ldots, n\}$ is the set of s-dimensional eigenvectors of G_t corresponding to the n largest eigenvalues. After a transformation by 2DPCA, a feature matrix is obtained for each image. Then, a nearest neighbor classifier is used for classification. Here, the distance between two arbitrary feature matrices Q_i and Q_j is defined by using Euclidean distance as follows:

$$d(Q_i, Q_j) = \sqrt{\sum_{u=1}^{k} \sum_{v=1}^{s} (Q_i(u,v) - Q_j(u,v))^2} \tag{3}$$

Given a test sample Q_t, if $d(Q_t, Q_c)$ = min j $d(Q_i, Q_j)$, then the resulting decision is Q_t belongs to the same class as Q_c.

2.2. Standard Kernel Based 2DPCA

Kernel based 2DPCA is a feature extraction method closely related to 2DPCA [16]. The algorithm can be seen as follow:

Idea: Consider each row of all training image matrices as a column vector sample and apply KPCA.

Input: A set of M sample images is represented as { P_1, P_2, …, P_M } with $P_i \in R^{kxs}$. Let $r_l = (P_i^{j*})^T \in R^s$, where $i = 1..M, j = 1..k$ and $l = k(i-1) + j$, be a column vector which is the transpose of the row j^{th} of image matrix i^{th}.

Algorithm:

1. Centering projected samples $\psi(r_l)$
2. Define kernel matrix $K \in R^{kMxkM}$ by

$$K_{i,j} = k(r_i, r_j) = \psi(r_j) \tag{4}$$

where $i, j = 1..kM$ and $\psi : R^s \rightarrow R^f$

3. Solve the Eigen problem $K\alpha = \lambda\alpha$, $\alpha \in R^{kN}$ and $u^\psi = \sum_{i=1}^{kM} \alpha_i \psi(r_i) \in R^f$
4. Projecting the image in R^{kxf} to a lower dimensional space spanned by the eigenvectors u^ψ. Let $P \in R^{kxs}$ be a sample whose projection is $\psi(P)$ in R^{kxf}, then the projection of $\psi(P)$ onto the eigenvectors u^ψ is :

$$\psi(P)u^\psi = \begin{bmatrix} \psi((P_i^{1*})^T)^T \\ \cdots \\ \psi\left(\left(P_i^{k*}\right)^T\right)^T \end{bmatrix} u^\psi = \begin{bmatrix} \psi((P_i^{1*})^T)^T u^\psi \\ \cdots \\ \psi\left(\left(P_i^{k*}\right)^T\right)^T u^\psi \end{bmatrix} = \begin{bmatrix} \sum_{t=1}^{nM}(\alpha_t((P_i^{1*})^T)^T \psi(r_{t)}) \\ \cdots \\ \sum_{t=1}^{nM}(\alpha_t\left(\left(P_i^{k*}\right)^T\right)^T \psi(r_{t)}) \end{bmatrix} \in R^k \tag{5}$$

The above algorithm is called standard Kernel based 2DPCA for two reasons. Firstly, it explicitly develops the model of Kernel based 2DPCA with analytic solution. Secondly, as concern as the computation complexity, the model is primary. For the size of kernel matrix K is the rM x rM, the complexity of diagonalizing the kernel matrix K is $O((rN)^3)$. Because rM is always large in the real pattern recognition tasks, the standard Kernel based 2DPCA suffers from computational problem.

2.3. Efficient Kernel Based 2DPCA

The efficient Kernel based 2DPCA turns to solve the eigenvalue problem of the new kernel matrix $\acute{K}$ and the diagonal sub matrix K *(j, j) (j=1,…,r).* Thus the complexity of the proposed method is max $\left(O\left((\sum_{j=1}^r h_j)^3\right), O(rN^3)\right)$. The proposed method divided the kernel matrix into several sub matrices and then performed dimension reduction on each sub matrix to obtain the approximate new kernel matrix with smaller size. In addition, the proposed approach just needs to store new kernel matrix $\acute{K}$ with lower dimension rather than the standard kernel matrix K. Finally, we diagonalized the new kernel matrix to extract the eigenvectors. The procedure of efficient algorithm for Kernel based 2DPCA is:

1. Divide the kernel matrix K into r x r block matrix $K(j,l)_{j,l=1,\ldots,r}$, where $K(j,l)$ is M x M sub matrix.
2. Obtain the principal components of the sub matrix $K(j,j)_{j=1,\ldots,r}$.
3. Select the leading principal components of $K(j,l)$ to compute the matrix $\acute{K}$ *(j, l)* and then construct the $\bar{G}_t$(instead of diagonalizing G_t since $\bar{G}_t$ is approximate to G_t.

$$\bar{G}_t = \sum_{j=1}^{r} \sum_{m=1}^{h_j} \bar{Y}^m(j)\, \bar{Y}^m(j)^T \tag{16}$$

where $\bar{Y}^m(j) = \sum_{i=1}^{M} \delta_i^m(j)\,(P_j - \mu_P)$ with $\delta^m(j) = (\delta_1^m(j), \ldots, \delta_M^m(j))^{\mathrm{T}}$ is the eigenvector of eigenequation.

4. Diagonalize $\bar{G}_t$ to get eigenvalue λ′ and eigenvectors α' based on

$$\bar{G}_t\,\alpha' = \lambda'\,\alpha' \qquad (17)$$

where $\alpha' = \sum_{j=1}^{r}(P_i - \mu_P)\Delta(j)\partial^j$ with ∂^j are the weights of leading eigenvalue $(\partial^j = (\partial_1^j, \dots, \partial_{h_j}^j)^T)$ and $\Delta(j)$is the coefficients matrix

5. Compute projections of test points onto the eigenvector α' using

$$w_s^{\varphi} = \varphi(b)\alpha'^{s} = \begin{pmatrix}\varphi(b^1)\\ \dots \\ \varphi(b^r)\end{pmatrix}^T \sum_{j=1}^{r}\left(P_j - \mu_P\right)\Delta(j)\partial^{j(s)} = \sum_{j=1}^{r} K(j)\Delta(j)\,\partial^{j(s)} \qquad (18)$$

where $s = 1, \dots, A$. Let $\lambda'_1, \dots, \lambda'_A$ A are the leading eigenvalues of $\acute{K}$, the corresponding eigenvectors are $\alpha'^1, \dots, \alpha'^A$.

Hence, the nonlinear principal component matrix of b corresponding to ϑ is $B^{\vartheta} = (w_1^{\varphi}, \dots, w_A^{\varphi})$.

3. Research Method

In this research, we decide two main processes, which are training process and testing process. The proposed system is shown in Figure 1, that are consists of three modules: image preprocessing, feature extraction, and classification. The entire system flows are briefly described as follows. The first module employs smiling faces data. Every smiling face image was the size 50x50 pixels. The data is manually cropped against a face data at oral area and produces spatial coordinate [5.90816 34.0714 39.3877 15.1020]. This process causes the face data size reduction into 40x16 pixels. The coordinate is being employed as a reference for the automatically cropping process against all other face data.

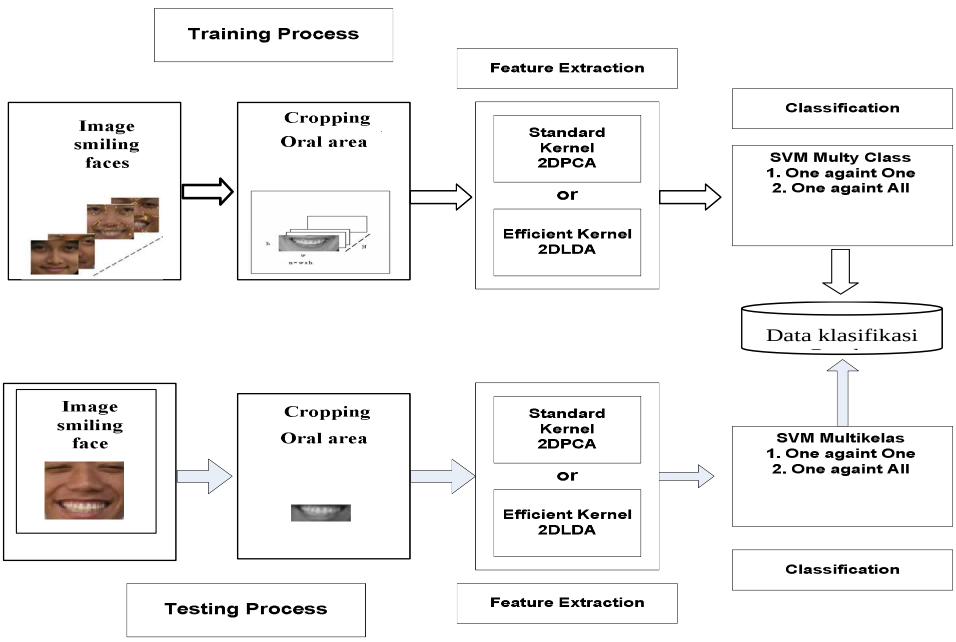

Figure 1. The proposed system

Next, first experiment the feature extraction performs standard Kernel based 2PCA to compute the projection matrices from the input space to a high dimensional feature space by a nonlinear mapping function. Second experiment, we use efficient Kernel based 2DPCA from the input matrix space to a high dimensional feature space by a nonlinear mapping function defined in a similar way as KPCA and standard Kernel based 2DPCA. Finally, to determine the classification results, it is necessary to be conducted similarity measurement based on standard Kernel based 2DPCA and efficient Kernel based 2DPCA by using Euclidean distance.

4. Results and Analysis

This section evaluates the performance of PCA algorithm [1], 2DPCA [2], KPCA [11], standard Kernel based 2DPCA [16] and our new approach efficient Kernel based 2DPCA based on smile stages database. For the experiment, standard Kernel based 2DPCA and efficient Kernel based 2DPCA method uses the 90 smiling faces data, i.e., 30 data for Stage I, 30 data for Stage III, and 30 data for Stage IV (see Figure 2). The data itself has been validated by dentist specializing in tooth conservation. Initially, every face data has the size of 50x50 pixels at each stage. It is manually cropped against a face data at oral area. This process causes the face data size reduction into 40x16 pixels. All experiments are carried out on laptop with CPU : Intel Core 2 Duo T6500 @ 2.10 GHz, RAM : 2GB and MATLAB 7.6 software platform.

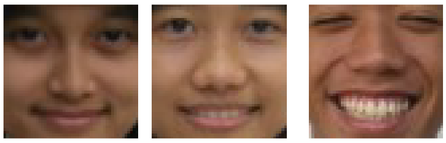

Figure 2. Visualization of smile at stage I, stage III, and stage IV

As the method of three-fold cross validation is being applied, data at each stage is divided into 3 groups. The first 2/3 data (20 data) becomes the training data, while the last 1/3 data (10 data) work as the testing data for each group. Those groups are being rotated with no overlap, thus all of them have the experience of becoming testing data. The training data were used to learn the subspace, while the testing data were then projected into the higher-dimensional representation subspace. The total number of training data and testing data are 60 and 30 respectively. From Section 2.3, the size of kernel matrix K of the standard Kernel based 2DPCA is $rM \times rM$ = 720 x 720 in our experiment, we firstly divide K into an $r \times r$ = 12 x 12 block matrix $K(j, l)$, where $K(j, l)$ is $M \times M$ = 60 x 60 sub matrix. To each sub matrix $K(j, l)$, we select the 95% leading components to compute the $\Delta(j)$and construct $\bar{G}_t$. Table 1 based on previous research shows comparison of recognition accuracy on smile stages recognition with PCA, 2DPCA and KPCA algorithm.

Table 1. Comparison of recognition accuracy on smile stages recognition

Method	**Accuracy (%)**
PCA	74.05
KPCA	86.50
2DPCA	91.44

Table 2. Comparison of training time, space needed and recognition accuracy with polynomial kernel (d = 5)

Method	Dimension	Time (s)	Space (Mb)	Accuracy (%)
Standard Kernel based 2DPCA	720x84	467	74	93.5
Efficient Kernel based 2DPCA	175x26	186	19	93.5

Table 3. Comparison of training time, space needed and recognition accuracy with gaussian kernel ($2\sigma^2 = 3 \, x \, 10^5$)

Method	Dimension	Time (s)	Space (Mb)	Accuracy (%)
Standard Kernel based 2DPCA	720x245	598	154	95.67
Efficient Kernel based 2DPCA	190x58	261	47	95.67

From Tables 1, 2, and 3, we can find that the efficient Kernel based 2DPCA method achieves the same best results as the standard Kernel based 2DPCA method than other methods. We find that the efficient approach can significantly reduce the training time and space compared to the standard Kernel based 2DPCA. While, use of Gaussian kernel can recognize more test data accurately, but they need more time and space. In addition, we only show results of 60 training data because the highest number of training data may cause the out of memory problem.

4. Conclusion

In this paper, we propose efficient Kernel based 2DPCA for solving the large computation complexity of standard Kernel based 2DPCA method. The results of smile stages recognition experiments are shown that Kernel based 2DPCA obtains better accuracy than KPCA method as well as the efficient Kernel based 2DPCA can effectively speed up the training procedure and remarkably save the memory consuming compared to the standard Kernel based 2DPCA. We also proved that Kernel based 2DPCA could be implemented by using KPCA technique.

References

[1] Kirby M, Sirovich L. Application of the Karhunen-loeve Procedure for The Classification of Human Faces. *IEEE Transactions Pattern Analysis of Machine Intelligent*. 1990; 12(1): 103-108.
[2] Turk MA, Pentland AP. Eigenfaces for Recognition. *Journal of Cognitive Neuroscience.* 1991: 3(1) 71-86.
[3] Yang J, Zhang D, Frangi AF, Yang JY. Two Dimensional PCA: A New Approach to Appearance-based Face Representation and Recognition. *IEEE Transactions Pattern Analysis of Machine Intelligent*. 2004; 26(1): 131-137.
[4] Cahyono GR, Purnomo MH, Haryadi M. *Smile Stages Classification Based on Aesthetic Dentistry Using Eigenfaces, Fisherfaces and Multiclass SVM.* 4th International Conference on Biomedical Engineering. 2008: 45-50.
[5] Wahyuningrum RT, Purnomo MH, Purnama IKE. *Smile Stages Recognition in Orthodontic Rehabilitation Using 2DPCA Feature Extraction.* First International Conference on Green Computing and The Second AUN/SEED-NET Regional Conference on ICT. 2010: 214-216.
[6] Wahyuningrum RT, Purnama IKE, Purnomo MH. *Smile Stages Classification by Using Feature Extraction Based on 2DPCA and 2DLDA in Orthodontic Rehabilitation.* 6th International Conference on Biomedical Engineering. 2010: 120– 125.
[7] Purnomo MH, Sarjono TA, Muntasa A. *Smile Stages Classification Based on Kernel Laplacian-lips Using Selection of Non Linear Function Maximum Value.* IEEE International Conference on Virtual Environments Human-Computer Interfaces and Measurement Systems (VECIMS). 2010: 151-156.
[8] Philips E, The Classification of Smile Pattern. *Journal Can Dentist Association*, 1999, 65:252-254
[9] Whitehill J, Littlewort G, Fasel I, Bartlett M, Movellan J. Developing a Practical Smile Detector. www.mplab.ucs.edu/~jake/pami_paper.pdf
[10] Wojdel JC, Rothkrantz LJM. Mixed Fuzzy-system and Artificial Neural Network Approach to Automated Recognition of Mouth Expression. *Knowledge based Systems*, Faculty of Information Technology and Systems, Delft University of Technology Delft, The.Netherlands. www.citeseer.ist.psu.edu/220537.html
[11] Schölkopf B, Smola A, Muller KR. Nonlinear Component Analysis as a Kernel Eigenvalue Problem. *Neural Computation.* 1998; 10(5): 1299-1319.
[12] Schölkopf B, Mika S, Burger CJC, Knirsch P, Muller KR, Raetsch G, Smola A. Input Space vs. Feature Space in Kernel Based Methods. *IEEE Transaction on Neural Network.* 1999; 10(5): 1000-1017.
[13] Rosipal R, Girolami M, Trejo L, Cichocki A. An Expectation-maximization Approach to Nonlinear Component Analysis. *Neural Computation.* 2001; 13: 505-510.
[14] Mika S, Rätsch G, Weston J, Schölkopf B, Muller KR. *Fisher Discriminant Analysis with Kernels.* IEEE Workshop on Neural Network for Signal Processing IX. 1999: 41-48.
[15] Francis RB, Michael IJ. Kernel Independent Component Analysis. *Journal of Machine Learning Research*. 2002; 3: 1-48.
[16] Zheng WM, Zou CR, Zhao L. An Improved Algorithm for Kernel Principal Component Analysis. *Neural Processing Letters.* 2005; 22: 49-56.
[17] Muntasa A, Indah AS, Mauridhi HP. Appearance global and local structure fusion for face image recognition. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2011; 9(1): 125-132.