

A Novel Pattern Classification Scheme: Classwise Non-Principal Component Analysis (CNPCA)

Guorong Xuan, Peiqi Chai, Xiuming Zhu,
Qiuming Yao, Cong Huang
Dept. of Computer Science,
Tongji University, Shanghai, China
grxuan@public1.sta.net.cn

Yun Q. Shi, Dongdong Fu
Dept. of Electrical and Computer Engineering
New Jersey Institute of Technology
Newark, New Jersey, USA
shi@njit.edu

Abstract

This paper¹ presents a novel pattern classification scheme: Class-wise Non-Principal Component Analysis (CNPCA), which utilizes the distribution characteristics of the samples in each class. The Euclidean distance in the subspace spanned by the eigenvectors associated with smallest eigenvalues in each class, named CNPCA distance, is adopted as the classification criterion. The number of the smallest eigenvalues is selected in such a way that the classification error in a given database is minimized. It is a constant for the database and can be determined by experiment. The CNPCA classification scheme usually outperforms other classification schemes under the situations of high computational complexity (associated with high dimensionality of features and/or calculation of inverse variance matrix) or high classification error rate (e.g., owing to the scattering of between-class being less than that of within-class). The experiments have demonstrated that this method is promising in practical applications.

1. Introduction

The existing classification methods are facing many challenges in the high dimensionality of features and/or calculation of inverse variance matrix. Another is that the classification error rate is sometimes unsatisfactory. One of such examples is when the scattering of between-class is less than that of within-class. This situation exists in the applications of Steganalysis or Computer Forensics.

One way to solve the high dimensionality problem is to map the original high-dimension feature space

onto a lower-dimension feature space. PCA (Principal Component Analysis) is often used for this purpose. However, the PCA sometimes does not provide a satisfactory performance in classification, because the PCA averages not only the characteristics of between-class but also the characteristics of within-class. The MCA (Minor Component Analysis) has also been studied in literatures [1] recently. Both PCA and MCA use the global variance instead of class-wise variance. The PCNSA (Principal Component Null Space Analysis) [2] adopts the null covariance of the within class as a sub-space, but it is hard to define the null covariance to form the null sub-space in PCNSA.

Considering the above limitations of PCA, MCA, and PCNSA, a new scheme for classification, class-wise non-principal component analysis (named CNPCA for short in this paper), is proposed in this paper. CNPCA makes the use of characteristics of each class distribution such as class mean and class variance.

The rest of the paper is organized as follows. The basic concept of CNPCA is introduced in Section 2. The relationships between CNPCA distance and other conventional distances are analyzed in Section 3. The experimental results of CNPCA classification are presented in Section 4 and the conclusion is drawn in Section 5.

2. The basic concept of CNPCA

The CNPCA (Class-wise Non-Principal Component Analysis) is suitable for solving the aforementioned high dimensionality and the scattering of between-class being less than that of within-class problems. The class-wise statistics is used by CNPCA to show the characteristics of each class distribution such as class variance instead of the global variance used in PCA. The CNPCA performs well for classification under certain conditions. Furthermore, there is no need to do

¹ This research is supported partly by National Natural Science Foundation of China (NSFC) on the project (90304017), and by New Jersey Commission of Science and Technology via New Jersey Center of Wireless Networking and Internet Security (NJWINS).

the dimension reduction so that the computation is simple.

Let \mathbf{x} denote the n -dimensional random vectors in the k^{th} class, and assume that there are in total K different classes. When the eigenvalues of the covariance matrix generated from all of \mathbf{x} are ranked from the largest to the smallest in non-increasing order, the corresponding eigenvector matrix can be expressed as:

$$\Phi_k = (\Phi_k)_{n \times n} = [\Phi_{rk}, \Psi_{rk}]_{n \times n},$$

where the $\Phi_k = (\Phi_k)_{n \times n}$ is the eigenvector matrix with all eigenvectors of the k^{th} class; the $\Phi_{rk} = (\Phi_{rk})_{n \times r}$ is the principal components matrix with all r eigenvectors of the k^{th} class; the $\Psi_{rk} = (\Psi_{rk})_{n \times (n-r)}$ is the non-principal components matrix with all $(n-r)$ eigenvectors of the k^{th} class; the k^{th} class's non-principal components Ψ_{rk} and principal components Φ_{rk} are complementary to each other; the n is the dimensionality; the r , ($r \leq n$), is defined as the number of eigenvectors associated with the largest eigenvalues; or the $(n-r)$ is the number of eigenvectors associated with the smallest eigenvalues.

2.1 CNPCA distance

In CNPCA classification, for the k^{th} class n -dimensional random vector samples, the Euclidean distance in the subspace spanned by the $(n-r)$ class non-principal components is adopted as the classification criterion. For a sample \mathbf{x} , the CNPCA distance of the k^{th} class is defined as:

$$D_{rk} = \{\Psi'_{rk} (x - M_k)\}' \{\Psi'_{rk} (x - M_k)\} \quad (1)$$

where D_{rk} stands for the summation of Euclidean distances between the sample \mathbf{x} and the mean of the k^{th} class M_k in the $(n-r)$ dimensional CNPCA space, D_{rk} can be represented by the classwise non-principal components matrix Ψ_{rk} .

CNPCA distance D_{rk} can also be represented by Φ_{rk} :

$$D_{rk} = (x - M_k)(x - M_k) \quad (2)$$

$$- \{\Phi'_{rk} (x - M_k)\}' \{\Phi'_{rk} (x - M_k)\}$$

Obviously, there are two special cases. When $r=0$, CNPCA distance is reduced to Euclidean distance while when $r=n$, CNPCA distance equals to 0. Hence the case of $r>0$ and $r<n$ is usually used.

2.2 CNPCA classification

In CNPCA classification, a sample \mathbf{x} is firstly mapped into the $(n-r)$ non-principal components subspace of each class. The distances in these subspaces between \mathbf{x} and the mean of each class are then calculated respectively. Finally the \mathbf{x} is classified to class k to which the distance is the minimum.

$$\hat{k} = \arg \min_k \{ D_{rk} \}.$$

The number of r is an important parameter for CNPCA. It can be estimated by minimizing the classification error rate \mathcal{E} :

$$\hat{r} = \arg \min_r \{ \mathcal{E}(D_{rk}) \}$$

3. Discussion of CNPCA distance

In this section, the relations between the CNPCA and a few frequently used distances are discussed.

3.1 CNPCA distance vs. Euclidean distance

Euclidean distance is represented in the eigenvector space as follows.

$$D_{Ek} = (x - M_k)' \Phi_k \Phi_k' (x - M_k) = (x - M_k)' \times$$

$$\Phi_k \begin{bmatrix} (I)_{r \times r} & (0)_{r \times (n-r)} \\ (0)_{(n-r) \times r} & (I)_{(n-r) \times (n-r)} \end{bmatrix} \times \Phi_k' (x - M_k)$$

Substitute the block matrix $(0)_{r \times r}$ for $(I)_{r \times r}$, we have D_{rk} as shown in Formula (1):

$$D_{rk} = (x - M_k)' \Psi_{rk} \Psi_{rk}' (x - M_k) = (x - M_k)' \Phi_k \times$$

$$\begin{bmatrix} (0)_{r \times r} & (0)_{r \times (n-r)} \\ (0)_{(n-r) \times r} & (I)_{(n-r) \times (n-r)} \end{bmatrix} \times \Phi_k' (x - M_k)$$

The CNPCA classifier works with an r value which is derived for minimizing classification error rate \mathcal{E} . The r can be larger than 0 or equal to 0. When $r=0$, the CNPCA and the Euclidean distance classifier are the same and hence perform equally well. When $r>0$, the CNPCA outperforms the Euclidean distance classifier.

3.2 CNPCA distance vs. Mahalanobis distance

The Mahalanobis distance is shown as:

$$D_{Mk} = (x - M_k)' \Sigma_k^{-1} (x - M_k)$$

$$= (x - M_k)' \Phi_k \Lambda_k^{-1} \Phi_k' (x - M_k)$$

$$= (x - M_k)' \Phi_k \times$$

$$\begin{bmatrix} (\lambda^{-1})_{r \times r} & (0)_{r \times (n-r)} \\ (0)_{(n-r) \times r} & (\lambda^{-1})_{(n-r) \times (n-r)} \end{bmatrix} \Phi_k' (x - M_k)$$

where λ^{-1} is diagonal component of the inverse eigenvalue matrix. The substitution is as follows:

$$(\lambda^{-1})_{r \times r} \rightarrow (0)_{r \times r}, \quad (\lambda^{-1})_{(n-r) \times (n-r)} \rightarrow (I)_{(n-r) \times (n-r)},$$

we have D_{rk} as defined in Formula (1).

One of the advantages of CNPCA classification is that it does not need to calculate the inverse variance matrix, thus avoids a potential computational problem. In practice, e.g., the calculation of inverse matrix encountered in the situation of high dimensionality of feature vectors often causes problem in computation. Furthermore when the assumption of Gaussian distribution is not valid, the Mahalanobis distance classifier is no longer optimal any more.

3.3 CNPCA distance vs. Tangent distance

The CNPCA distance D_{rk} is similar to the tangent distance (ref. [3] p.191). CNPCA classification is shown in Formula (3) by using eigenvector matrix as tangent matrix and mean vector M_k as prototype.

$$D_{rk} = \min_{a_{rk}} \left[\|(x - M_k) - \Phi_{rk} a_{rk}\| \right] \quad (3)$$

CNPCA distance can be viewed as an optimal adjustment. If

$$\frac{d(\|(x - M_k) - \Phi_{rk} a_{rk}\|)}{da_{rk}} = 0,$$

we have $a_{rk} = \Phi_{rk}'(x - M_k)$. Substituting a_{rk} into formula (3), we get the same D_{rk} as shown in formula (2).

One of the characteristics of CNPCA classification is the easiness in using eigenvector matrix instead of tangent matrix. Hence the CNPCA classifier by optimal adjustment usually outperforms tangent distance classifier.

3.4 CNPCA vs. PCNSA

As mentioned in Section 1, the PCNSA (Principal Component Null Space Analysis) [2] adopts the null covariance of the within class as a sub-space. However, when the eigenvalues are not exactly equal to zero, the PCNSA finds separate directions for each class along the null covariance of the within class sub-space according to a pre-defined criterion, $\lambda \leq \lambda_0$, where $\lambda_0 = 10^{-4} \lambda_{\max}$. Differently, the proposed CNPCA finds separate directions for each class along the smaller covariance of the within class sub-space via minimizing the classification error under a given database to determine an optimal number, (n-r), of eigenvalues, instead of using a fixed λ_0 to define the null space as in PCNSA.

4. Experimental results

4.1 Experiment 1 --two-class 8D data

The 8-dimensional two-class data used in the experiment 1 can be found in [4, p.46]. The mean vectors and the covariance matrices are as follows:

$$M_1 = 0 = [0, 0, \dots, 0]^t, \quad \Sigma_1 = I = \text{diag}(1, 1, \dots, 1)$$

$$M_2 = M = [m_1, \dots, m_8]^t, \quad \Sigma_2 = \Lambda = \text{diag}(\lambda_1, \dots, \lambda_8)$$

where m_i and λ_i are detailed in Table 1 :

Table 1. Data of experiment 1

i	1	2	3	4	5	6	7	8
m_i	3.86	3.10	0.84	0.84	1.64	1.08	0.26	0.01
λ_i	8.41	12.06	0.12	0.22	1.49	1.77	0.35	2.73

We introduce a factor F to adjust the mean vectors $M \Rightarrow (F \cdot M)$. In Figure 1, error rate is shown as a function of r for different F . The r is the number of eigenvectors associated with largest eigenvalues. The $r=4$ is selected finally in CNPCA, because the minimum of error rate appears at $r=4$ under either $F=0.1$ or $F=1$. At $r=0$ under $F=2$, CNPCA will degenerate into Euclidean distance classifier.

The MAP (maximum a posteriori classifier) is denoted as Bayesian classifier in this paper (ref. [3] p.111). In Figure 2, the error of CNPCA will be smaller than that of Bayesian classifier pre-processed by PCA (PCABayes), when the dimensionality is less than 5, 3 and 1 under $F=0.1$, $F=1$ and $F=2$ respectively. When F is small, the CNPCA outperforms the PCA, which is followed by a Bayesian classifier.

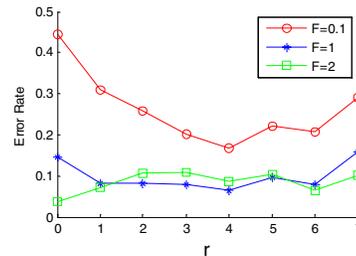


Figure 1. Error rate as a function of r for different F

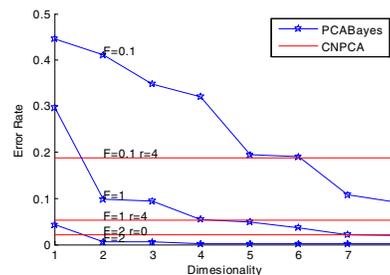
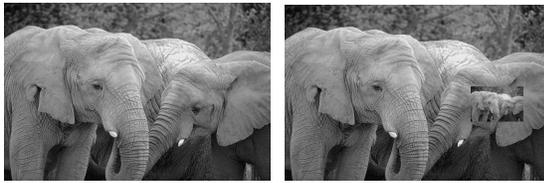


Figure 2. Comparison of CNPCA and PCABayes

4.2 Experiment 2 -- tampering detection



(a) Original (b) tampered
Figure 3. Original and tampered image

We use the 1096 images of image database CorelDRAW (www.corel.com) for our tampered (forensic) experiment. The tampered image is generated by embedding a sub-image, which is of 1/25 of the original image into each image. In Figure 3 an example of the original image and its corresponding tampered image is shown.

There are only two classes in this experiment: the original image and tampered image, as shown in Figure 3. The difference between these two images is very small while the images in each class (1096) are very diverse. Therefore, it is a typical example in which the between-class clustering is larger than the within-class scattering.

We extract the four diagonal lines' coefficients $n = 4 \times 256 - 6 = 1018$ in the image gray level co-occurrence matrix (256x256) shown in Figure 4 as the feature vector [5]. In this experiment, $r = 496$ is chosen as optimal value for the r , see Figure 5. In experiment 2, 896 images are randomly selected for training, the rest 200 images are used for testing. We do the test 10 times to average the results.

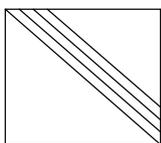


Figure 4.
 co-occurrence matrix

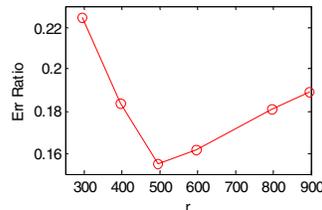


Figure 5. Error rate vs. r

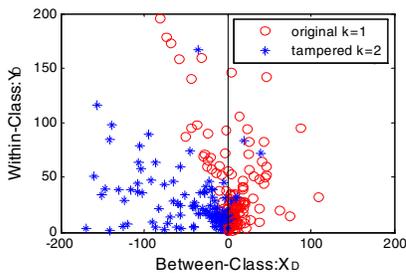


Figure 6. Distribution of Samples

The distribution of samples is shown in Figure 6. X_D shows the difference of the CNPCA distances of

between-class of tampered and non-tampered (original) images:

$$X_D = D_{k=2,r} - D_{k=1,r}$$

Y_D represents, respectively, the CNPCA within-class distances for two different classes.

$$Y_D = \begin{cases} D_{k=1,r}, k=1 & \text{original image} \\ D_{k=2,r}, k=2 & \text{tampered image} \end{cases}$$

The false positive rate $FP = 37/200$, i.e., among 200 test images there are 37 original images are mistakenly classified as tampered images (circles in red as $(X_D < 0)$). The false negative rate $FN = 26/200$, i.e., among 200 test images, there are 26 tampered images are wrongly classified as non-tampered images (stars in blue as $(X_D > 0)$). The average detection error rate is $63/400$ (15.75%).

5. Conclusions

(1) In the proposed CNPCA classification scheme, the distance between the samples and the mean of each class in the corresponding class-wise non-principal components subspace is considered as the classification criterion.

(2) The CNPCA classification overcomes the problem of non-existing inverse matrix for covariance matrix in some conventional classification methods. It is simple in that there is no need to do the dimension reduction.

(3) The CNPCA keeps classification information in that it does not average the covariance matrix of all classes. On the contrary, the covariance matrix of each class plays an important role in the computation of CNPCA distance.

(4) The CNPCA classification method is particularly suitable to solve high dimensionality problem and/or between-class clustering with within-class scattering problem.

6. References

- [1] B. Li, K. Wang, and D. Zhang, "On-Line Signature Verification Based on PCA (Principal Component Analysis) and MCA (Minor Component Analysis)", *ICBA 2004*, Springer-Verlag Berlin Heidelberg, LNCS 3072, pp. 540-546
- [2] N. Vaswani, R. Chellappa, Classification probability analysis of principal component null space analysis, *ICPR2004*,
- [3] R. O. Duda, P. E. Hart, D. G. Stork, *Pattern Classification*, 2nd Edition, John Wiley & Sons, 2001.
- [4] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd Edition, Academic Press, Inc. Boston, 1990.
- [5] G. Xuan, C. Huang, X. Zhu, P. Chai, Y. Q. Shi and D. Fu, "Steganalysis with high-dimensional features by using co-occurrence matrix and class-wise non-principal component analysis (CNPCA), May 2006 (to appear).