Expert Systems with Applications 40 (2013) 5395–5401

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Theory of fractional covariance matrix and its applications in PCA and 2D-PCA

Chaobang Gao [a,b,*], Jiliu Zhou [b,c], Qiang Pu [a,b]

[a] College of Information Science and Technology, Chengdu University, Chengdu 610106, China
[b] Key Laboratory of Pattern Recognition and Intelligent Information Processing, Sichuan Province (Chengdu University), Chengdu 610106, China
[c] School of Computer Science, Sichuan University, Chengdu 610064, China

ARTICLE INFO

Keywords:
Fractional variance
Fractional covariance
Fractional covariance matrix
FPCA
2D-FPCA

ABSTRACT

In this paper, according to the definition and applications of fractional moments, we give new definitions of the fractional variance and fractional covariance. Furthermore, we give the definition of fractional covariance matrix. Based on fractional covariance matrix, principal component analysis (PCA) and two-dimensional principal component analysis (2D-PCA), we propose two new techniques, called fractional principal component analysis (FPCA) and two-dimensional fractional principal component analysis (2D-FPCA), which extends PCA and 2D-PCA to fractional order form, and extends the transition recognition ranges of PCA and 2D-PCA. To evaluate the performances of FPCA and 2D-FPCA, a series of experiments are performed on two face image databases: ORL and Yale. Experiments show that two new techniques are superior to the standard PCA and 2D-PCA if choosing different order between 0 and 1.

1. Introduction

Dimensionality reduction is one of the fundamental problems in computer vision, machine learning and biometric recognition (Jenssen, 2010; Qiu and Fang, 2008; Saul et al., 2005). The goal of dimensionality reduction is to capture the meaningful low-dimensional structures embedded in high-dimensional data and obtain more useful representations of the data for subsequent analysis such as classification, visualization, clustering, or outlier detection (Belkin and Niyogi, 2003; Lu et al., 2003). Dimensionality reduction methods strive to present high-dimensional data in a low-dimensional space, in a way of faithfully capturing the meaningful structures and unexpected relationships embedded in images. Principal component analysis (PCA) is a technique for classical feature extraction and data representation in low-dimensional space, which is widely used in pattern recognition and computer vision (Härdle and Simar, 2007). Sirovich and Kirby first used PCA to efficiently represent pictures of human faces. They argued that any face image could be reconstructed approximately as a weighted sum of a small collection of images that define the facial basis (eigenimages) and the mean image of face. Two-dimensional PCA (2D-PCA) was proposed by Yang and Yang (2002), Wang et al. (2005), Yang and Zhang (2004) to cut the computational cost of the standard PCA by presenting high-dimensional data in a low-dimensional space. In addition, linear discriminant analysis is also a popular feature extraction method in pattern recognition. It searches for a set of projection vectors onto which the data points in a same class are close to each other while data points in different classes are far from each other (Belhumeur et al., 1997; Li and Yuan, 2005; Li et al., 2009; Martinez and Kak, 2001; Swets and Weng, 1996). The important difference between PCA and 2D-PCA is that PCA treats an image as a vector and 2D-PCA views an image as a matrix. In general, 2D-PCA has much lower dimensionality than PCA. Although there are differences between 2D-PCA and PCA in image representation, the methods of dimensionality reduction of image are the same, which chooses the corresponding eigenvectors to form the projection matrix according to the magnitude of eigenvalues of the covariance matrix. It is well known that the covariance matrix is a key tool and plays an important role in the dimensionality reduction (Härdle and Simar, 2007; Kim et al., 2003; Martinez and Kak, 2001; Min et al., 2004; Swets and Weng, 1996). Based on this, in this paper, we define a new covariance matrix: fractional covariance matrix (FCM), and develop two main straightforward dimensionality reduction techniques, called fractional principal component analysis (FPCA) and two-dimensional fractional principal component analysis (2D-FPCA) for image feature extraction and dimensionality reduction by applying FCM to PCA and 2D-PCA.

2. Fractional covariance matrix

2.1. Fractional variance

In statistics, the higher integer moments help to reduce the sequence analyzed to a finite set of k statistically stable parameters,

* Corresponding author at: College of Information Science and Technology, Chengdu University, Chengdu 610106, China. Tel./fax: +86 28 84616783.

E-mail addresses: kobren427@163.com (C. Gao), zhoujl@scu.edu.cn (J. Zhou), 349599683@qq.com (Q. Pu).

0957-4174/$ - see front matter
http://dx.doi.org/10.1016/j.eswa.2013.03.048

keeping invariant the values of the first kth moments. In recent decades, the higher integer moments are generalized for the fractional and even complex moments (Kozubowski, 2001; Nigmatullin, 2006; Tallis and Light, 1968). And the applications of fractional moments are currently done in several scientific fields. Especially in signal processing, new statistical methods of signal detection have been developed based on the statistics of fractional moments that is turned out very effective in detection of superweak signal completely hidden in the random sequence analyzed, but nevertheless distorting the statistical behavior of the random sequence (Kozubowski, 2001; Nigmatullin, 2006; Tallis and Light, 1968). Variance is a special second-order moments and is also a crucial quantity in probability and statistics theory (Härdle and Simar, 2007). It is therefore of interest to investigate in pattern recognition for various variances as well. Based on the the definition and applications of fractional moments, in this paper, we will give new definitions of the fractional variances and fractional covariance, then apply them to dimensionality reduction.

Although we can start with a more general setting, for simplicity, we choose to consider the variance of continuous function f with respect to a weight function p (including probability densities) defined on a closed and bounded domain Ω in $\mathbb{R}^m$ instead of a distribution function. More precisely, unless stated otherwise, in the following discussions, let Ω be a fixed, nonempty, closed and bounded domain in $\mathbb{R}^m$ and $p:\Omega\to[0,\infty)$ be a fixed function which satisfies $\int_\Omega p d\omega = 1$. For any continuous function $f:\Omega\to\mathbb{R}$, we write

$$E(f,p)=\int_\Omega pfd\omega \tag{1}$$

which may be regarded as the weighted mean of the function f with respect to the weight function p. Recall that the variance of a random variable f with respect to a density function p, is as follows:

$$D(f,p)=E[(f-E(f,p))^2,p]. \tag{2}$$

According to the definition of fractional moments (Tallis and Light, 1968; Kozubowski, 2001; Nigmatullin, 2006), we may give a new definition of fractional (r-order) variance of the function $f:\Omega\to(0,\infty)$ as follows:

$$D_r(f,p)=E[(f^r-E^r(f,p))^2,p], \tag{3}$$

where r is a real number. Clearly, when $r = 1$, (3) is the general variance. Since the order r may be any fraction, (3) is called fractional variance. Thus the variance is extended to fractional order form. Such a definition is compatible with the generalized integral means. Indeed, according to the power mean inequality, we may see that

$$D_r(f,p)\geqslant 0,\quad \forall r\in\mathbb{R}. \tag{4}$$

As a statistical application, for any function $f:\Omega\to(0,\infty)$, the mean of the random variable $f(X)$ is

$$E[f(X)]=\int_\Omega pfd\omega=E(f,p), \tag{5}$$

and its variance is

$$D[f(X)]=E[(f(X)-E(f(X)))^2]=\int_\Omega p\{f-E[f(X)]\}^2d\omega=D(f,p). \tag{6}$$

Therefore,

$$D_r[f(X)]=D_r(f,p),\quad r\in\mathbb{R} \tag{7}$$

may be regarded as a fractional (r-order) variance of the random variable $f(X)$. Furthermore, the fractional variance and general variance have the following relation:

Theorem 1. *Let M and m be the maximum and minimum of the random variable $f(X)$. For $r\leqslant 1$, we have*

$$r^2M^{r-1}D[f(X)]\leqslant D_r[f(X)]\leqslant r^2m^{r-1}D[f(X)]. \tag{8}$$

Proof. For $r\leqslant 1$, we have

$$\begin{aligned}D_r[f(X)]&=\int_\Omega p\{f^r(X)-E^r[f(X)]\}^2d\omega\\&=\int_\Omega p\left[\frac{r\int_{E[f(X)]}^{f(X)}t^{r-1}dt}{f(X)-E[f(X)]}\right]^2\{f(X)-E[f(X)]\}^2d\omega\\&\geqslant\int_\Omega p[r^2\max\{f^r(X),E^r[f(X)]\}]\{f(X)-E[f(X)]\}^2d\omega\\&\geqslant r^2M^{r-1}\int_\Omega p\{f(X)-E[f(X)]\}^2d\omega=r^2M^{r-1}D[f(X)],\end{aligned} \tag{9}$$

where M denotes the maximum of $f(X)$. Similarly, we can prove that

$$D_r[f(X)]\leqslant r^2m^{r-1}D[f(X)], \tag{10}$$

where m denotes the minimum of $f(X)$. Hence, for $r\leqslant 1$, by (9) and (10), we get

$$r^2M^{r-1}D[f(X)]\leqslant D_r[f(X)]\leqslant r^2m^{r-1}D[f(X)].$$

When $r = 1$, the equalities in (8) hold. This ends the proof of Theorem 1. □

It is well known that the variance $D[f(X)]$ of the random variable $f(X)$ denotes the deviation from the mean $E[f(X)]$, hence the fractional variance $D_r[f(X)]$ denotes the fractional order deviation from the mean $E[f(X)]$. From Theorem 1, we know that the fractional variance $D_r[f(X)]$ is between $r^2M^{r-1}D[f(X)]$ and $r^2m^{r-1}D[f(X)]$, which can make the deviation measurement diverse or selective.

2.2. Fractional covariance matrix

Furthermore, we may define the fractional (r-order) covariance of the random variables $f(X)$ and $g(Y)$ as follows

$$C_r(f(X),g(Y))=E[(f(X)^r-E^r(f(X)))(g(Y)^r-E^r(g(Y)))], \tag{11}$$

where r is a real number. Similar to the fractional variance, (3) is called fractional covariance. Clearly, fractional covariance (11) is general covariance when $r = 1$ and (11) is general variance when $f(X) = g(Y)$ and $r = 1$. Similar to the general covariance, fractional covariance is always measured between 2 dimensions. If we have a data set over 2 dimensions, there is more than one covariance measurement that can be calculated. For example, from a 3 dimensional data set (dimensions $f(X)$, $g(Y)$, $h(Z)$) you could calculate $C_r[f(X),g(Y)]$, $C_r[f(X),h(Z)]$, and $C_r[g(Y),h(Z)]$. In general, for an n-dimensional data set, we can calculate $\frac{n!}{2(n-2)!}$ different fractional covariance values. Thus the fractional covariance matrix (FCM) for a set of data with n-dimensions is:

$$C_r^{n\times n}=(c_{i,j},c_{i,j}=C_r(Dim_i,Dim_j)),\quad i=1,2,\ldots,n,\ j=1,2,\ldots,n, \tag{12}$$

where $C_r^{n\times n}$ is a matrix with n rows and n columns, and Dim_i is the ith dimension. So, for an n-dimensions data set, the matrix has n rows and n columns, each element in the matrix is a result of calculating the fractional covariance between two separate dimensions.

3. FPCA and 2D-FPCA

3.1. Theory of FPCA

PCA technique first finds the covariance matrix of data with n-dimensions, then solves its eigenvalues by which the corresponding eigenvectors are ordered from high to low (Belhumeur et al., 1997; Wang et al., 2005; Yang and Zhang, 2004). This gives us the components in significance order and ignores the components of lesser significance without losing much information if its corresponding eigenvalues are small. After doing that, the final data set is an approximate to the original one, but with lesser dimensions. To be precise, if the data originally have n dimensions, we calculate n eigenvectors and eigenvalues, then we choose only the top p ($p < n$) eigenvectors, the final data set has only p dimensions.

If we apply fractional covariance matrix (12) to PCA instead of the general covariance matrix, we will obtain a new technique, fractional principal component analysis (FPCA). A m by n image can be expressed as a d-dimensional vector ($d = mn$), where the rows of pixels in the image are placed one after the other to form one vector. Let N be the number of the training images. We can then put N images together into one big image-matrix like this:

$$X = (X_1, X_2, \ldots, X_N),$$

where let $X_i = (X_{i1}, X_{i2}, \ldots, X_{id})^T$, $i = 1,2,\ldots,N$. Then the FCM (12) of the N images can be expressed as:

$$C_r^{d\times d} = \sum_{i=1}^{N}\left(X_i^{(r)} - \left(\frac{1}{N}\sum_{i=1}^{N}X_i\right)^{(r)}\right)\left(X_i^{(r)} - \left(\frac{1}{N}\sum_{i=1}^{N}X_i\right)^{(r)}\right)^T, \tag{13}$$

where $X_i^{(r)} = (X_{i1}^r, X_{i2}^r, \ldots, X_{id}^r)^T$. Let $C_r^{d\times d}$ have the eigenvalues-eigenvector pairs (λ_1, e_1), $(\lambda_2, e_2), \ldots, (\lambda_d, e_d)$, and $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_d$. Then the ith principal component is given by

$$Y_i = e_i' X = e_{i1}X_1 + e_{i2}X_2 + \cdots + e_{id}X_d, \quad i = 1,2,\ldots,d$$

with those choices

$$D(Y_i) = D_1(Y_i) = e_i' C_r^{d\times d} e_i = \lambda_i. \tag{14}$$

By (8), we know that the fractional variance of Y_i is not equal to λ_i. This is the difference between the fractional variance and general variance. To compress the data, we can then choose to transform the data only using the top p eigenvectors. This gives us a final data set with only p dimensions, which is a dimension reduced space. Here, we can choose p eigenvectors as the projection subspace according to the following contribution rate of eigenvalues (CRE),

$$\mathrm{CRE} = \frac{\sum_{i=0}^{p}\lambda_i}{\sum_{i=0}^{d}\lambda_i}. \tag{15}$$

Once we have chosen the components (eigenvectors) that we wish to keep in our data and formed a feature vector, we simply take the transpose of the vector and left multiply to the original data set as follows:

$$Y = AX, \tag{16}$$

where $A = (e_1, e_2, \ldots, e_p)$, the projection matrix is formed by the top p eigenvectors. The matrix Y is a feature matrix of the image sample X. We call this kind of data compression technique: fractional principal component analysis (FPCA). After a transformation by FPCA, two feature matrixes are obtained for training and testing images, respectively. Then, in low-dimensional feature space, a nearest neighbor classifier is used for recognition.

Here, we need to discuss the reconstruction error. From Cheng et al. (2009), Masashi and Mitsuyo (1968), we know that the mean reconstruction error need be minimized, which means that the contribution rate of eigenvalues is maximized. Conversely, the larger the contribution rate of eigenvalues is, the smaller the mean reconstruction error is. Therefore, in this paper, we only consider the variable of the contribution rate of eigenvalues.

3.2. Theory of 2D-FPCA

Similarly, if applying FCM (12) to 2D-PCA (Li and Yuan, 2005; Wang et al., 2005; Yang and Yang, 2002; Yang and Zhang, 2004), we will obtain a new technique, two-dimensional fractional principal component analysis (2D-FPCA). Let N be the number of training images and A_j be an m by n image matrix. Then the FCM of all training samples $A_1, A_2, \ldots, A_N$ can be expressed as:

$$C_r^{n\times n} = \frac{1}{N}\sum_{i=1}^{N}\left(A_i^{(r)} - \left(\frac{1}{N}\sum_{i=1}^{N}A_i\right)^{(r)}\right)^T\left(A_i^{(r)} - \left(\frac{1}{N}\sum_{i=1}^{N}A_i\right)^{(r)}\right), \tag{17}$$

where let

$$A_j = \left(a_j^{kl}\right)_{m\times n}, k = 1,2,\ldots,m,\ l = 1,2,\ldots,n,$$

then

$$A_j^{(r)} = \left(\left(a_j^{kl}\right)^r\right)_{m\times n}, k = 1,2,\ldots,m, l = 1,2,\ldots,n.$$

The optimal projection axis is the eigenvector of $C_r^{n\times n}$ corresponding to the largest eigenvalue. In general, it is not enough to have only one optimal projection axis. We usually need to select a set of projection axes, $X_1, X_2, \ldots, X_p$, are the orthogonal eigenvectors of $C_r^{n\times n}$ corresponding to the first p largest eigenvalues. For a given image sample A, let

$$Y_k = AX_k, \quad k = 1,2,\ldots,p. \tag{18}$$

Then, we obtain a family of projected feature vectors, $Y_1, Y_2, \ldots, Y_p$, called the principal components (vectors) of the sample image A. These principal components are then used to form an $m \times p$ matrix $B = (Y_1, Y_2, \ldots, Y_p)$. B is also called the feature matrix or feature image of the image sample A. We call this kind of data compression technique: two-dimensional fractional principal component analysis (2D-FPCA). After a transformation by 2D-FPCA, a feature matrix is obtained for each image. Then, a nearest neighbor classifier is used for classification.

4. Experiments and results

4.1. Comparison of PCA with FPCA

The performance of FPCA is evaluated on the ORL face database, and compared with the classical unsupervised method PCA. The ORL database (http://www.cam-orl.co.uk) contains images from 40 individuals, each providing 10 different images. For some subjects, the images are taken at different times. The facial expressions (open or closed eyes, smiling or nonsmiling) and facial details (glasses or no glasses) also vary. First, an experiment is performed on the first five image samples in each class for training, the remaining images for test. Thus, the total numbers of training and testing are 200 respectively. All images are grayscale and normalized to a resolution of 23×28 pixels. The fractional covariance matrix is obtained by (13). Here, the size of the fractional covariance matrix is 644×644. And it is very easy to calculate its eigenvalues, corresponding eigenvectors, and the contribution rate of eigenvalues by (15). The results of CRE are shown in Table 1. According to the contribution rate of eigenvalues, we can choose the eigenvectors corresponding to the top p largest eigenvalues, $e_1, e_2, \ldots, e_p$, as projection axes. After the projection of the image sample onto these axes by (16), we obtain constructed subimages in a low-dimensional space. The nearest neighbor classifier with Euclidean distance is finally employed for classification.

Table 1
Contribution rate of eigenvalues.

v/dim	5	10	15	20	25	30	35	40	50	60
PCA	0.5607	0.6995	0.7681	0.8133	0.8450	0.8693	0.8885	0.9041	0.9281	0.9448
0.01	0.5749	0.7044	0.7728	0.8181	0.8502	0.8735	0.892	0.9070	0.9290	0.9452
0.1	0.5723	0.7029	0.7716	0.8171	0.8492	0.8728	0.8914	0.9065	0.9287	0.9449
0.2	0.5697	0.7015	0.7705	0.8161	0.8482	0.8721	0.8907	0.9060	0.9284	0.9447
0.3	0.5674	0.7003	0.7695	0.8153	0.8473	0.8714	0.8902	0.9055	0.9281	0.9446
0.4	0.5654	0.6995	0.7687	0.8145	0.8465	0.8708	0.8896	0.9050	0.9279	0.9445
0.5	0.5637	0.6989	0.7680	0.8139	0.8459	0.8703	0.8892	0.9046	0.9278	0.9444
0.6	0.5624	0.6986	0.7677	0.8135	0.8455	0.8699	0.8888	0.9043	0.9277	0.9444
0.7	0.5615	0.6985	0.7675	0.8132	0.8452	0.8696	0.8886	0.9042	0.9278	0.9445
0.8	0.5609	0.6986	0.7675	0.8131	0.8450	0.8694	0.8885	0.9041	0.9278	0.9446
0.9	0.5606	0.6990	0.7677	0.8132	0.8450	0.8693	0.8884	0.9041	0.9279	0.9447
1	0.5607	0.6995	0.7681	0.8133	0.8450	0.8693	0.8885	0.9041	0.9281	0.9448

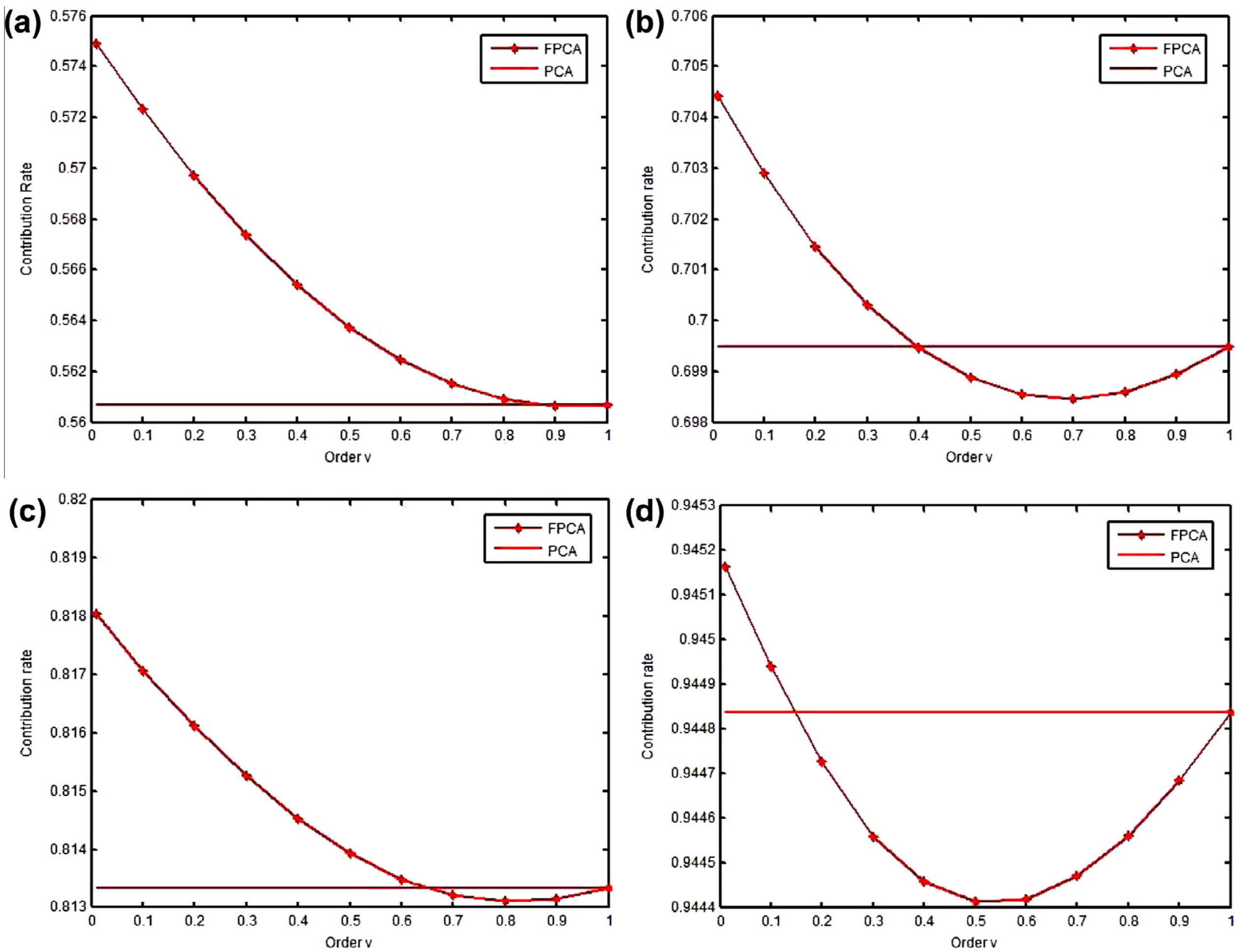

Fig. 1. Comparison of CRE between PCA and FPCA on ORL database in different dimensionality: (a) dimensionality is 5, (b) dimensionality is 10, (c) dimensionality is 20, (d) dimensionality is 60.

To visually analyze the data in Table 1 and discuss the relationship between the order v and dimensionality, we integrate nonlinear curves by the data when the dimensionality is 5, 10, 20, and 60, respectively. The results are shown in Fig. 1.

Fig. 1 shows the CRE is first down and then up with the increasing of the order r for FPCA. Specially, the lower the order v is, the greater the CRE difference between FPCA and PCA is. In general, the recognition accuracy is higher when the CRE is greater. Hence, when the order is smaller, the recognition accuracy of FPCA is greater than that of PCA. When the order $r = 1$, both the contribution rate of eigenvalues and the recognition accuracies of FPCA and PCA are equal. The top recognition accuracies of FPCA and PCA are showed in Fig. 2.

Fig. 2(a) shows the curve of the top recognition accuracy of FPCA is always above that of PCA in the same dimensionality, that is to say, the performance of FPCA is better than that of PCA. More specifically, when the dimensionality is 5 and the order $r = 0.01$, the top recognition accuracy of FPCA is 0.76 and the recognition accuracy of PCA is 0.71. From Fig. 2(b), it is clear that the top recognition accuracy of FPCA is always more than that of PCA in the same dimension. Overall, the top recognition accuracy of FPCA is about 5% up greater than that of PCA.

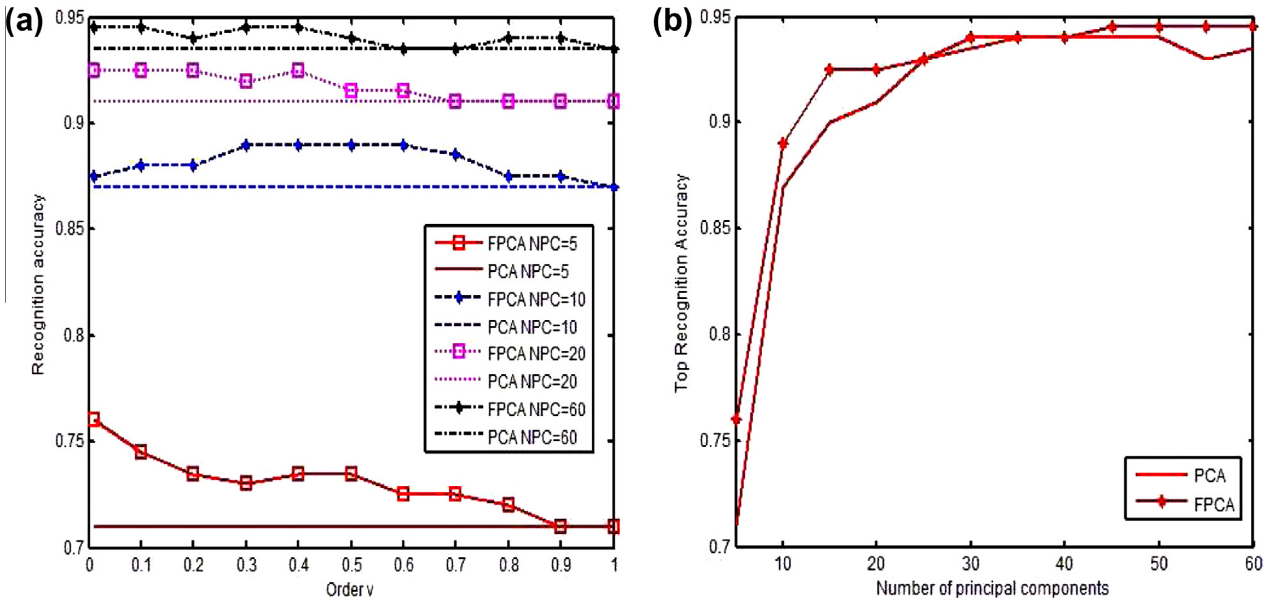

Fig. 2. Comparison of the top recognition accuracy between PCA and FPCA on ORL database, where NPC denotes the number of principal components (i.e., dimensionality): (a) comparison in different numbers of principal components, (b) comparison of the top recognition accuracy.

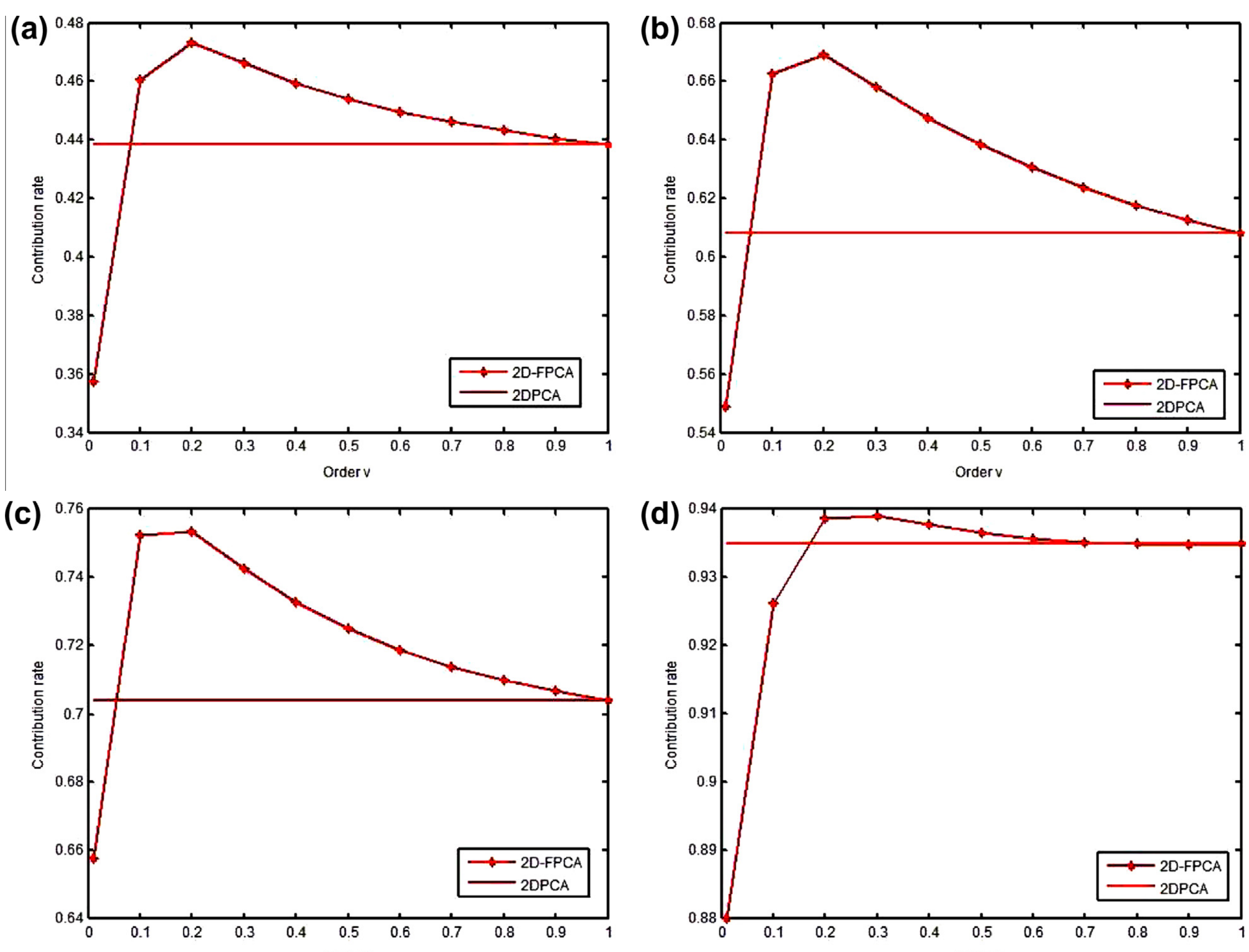

Fig. 3. Comparison of CRE between 2D-PCA and 2D-FPCA on Yale database in different dimensionality: (a) dimensionality is 1, (b) dimensionality is 2, (c) dimensionality is 3, (d) dimensionality is 10.

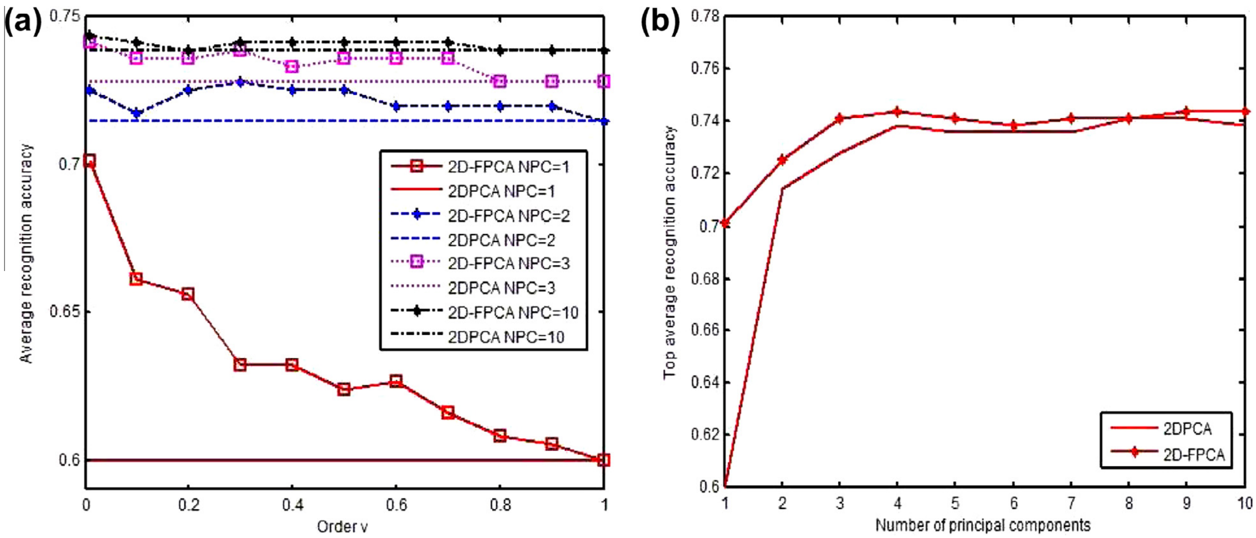

Fig. 4. Comparison of average recognition accuracy between 2D-PCA and 2D-FPCA on Yale Database, where NPC denotes the number of principal components (i.e., dimensionality): (a) comparison of the average recognition accuracy in different numbers of principal components, (b) comparison of the top average recognition accuracy.

4.2. Comparison of 2DPCA with 2D-FPCA

The performance of 2D-FPCA is evaluated on the Yale face database, and compared with the classical unsupervised method, 2DPCA. Yale face database contains 165 images of 15 individuals (each person has 11 different images) under various facial expressions and lighting conditions. Each image was manually cropped and resized to 40×40 pixels in this experiment. First, an experiment is performed on six randomly selected images from each class for training, while the remaining is kept for testing. The FCM is calculated by (17). After calculating the eigenvectors of FCM, it is easy to obtain the CRE of 2D-FPCA and 2D-PCA by (15). The results are showed in Fig. 3. Then, the nearest neighbor classifier with Euclidean distance is finally utilized for classification. Recognition is similar to the 2D-PCA method, refer to Li and Yuan (2005), Wang et al. (2005), Yang and Yang (2002), Yang and Zhang (2004), for details.

From Fig. 3, we can see that the contribution rate of eigenvalues of 2D-FPCA is smaller than that of 2D-PCA when the order $0 < r < 0.1$ and it is greater than that of 2D-PCA when $0.1 \leqslant r < 1$. When $r = 1$, both the contribution rate of eigenvalues and recognition accuracies of 2D-FPCA are equal to those of 2D-PCA. Since the training images are chosen randomly and the recognition accuracy depends on the training images, we take average recognition accuracy from five experimental results as the recognition accuracy for each dimensionality and each order. The results of recognition accuracy are showed in Fig. 4.

Fig. 4(a) shows the curve of the average recognition accuracy of 2D-FPCA is always above that of 2D-PCA in the same dimensionality, that is, the performance of 2D-FPCA is better than that of 2D-PCA. More specifically, when the dimensionality is 1, the average recognition accuracy of 2D-FPCA is obviously higher than that of 2D-PCA, where the average recognition accuracy of 2D-FPCA and 2D-PCA achieves 0.7013 and 0.6, respectively. In general, the average recognition accuracy of 2D-FPCA is about 5% greater than that of 2D-PCA, and it is 10% up larger than that of 2D-PCA. Here note that CRE of 2D-FPCA is smaller than that of 2D-PCA when the order $0 < r < 0.1$, but the average recognition accuracy of 2D-FPCA is still higher than that of 2D-PCA, which is similar to FPCA and PCA and just the speciality of 2D-FPCA contrasted to 2D-PCA, that is to say, the mean reconstruction errors of 2D-PCA and PCA are same in theory, but they depend on the accuracy of estimated sample covariance matrix in practice (Cheng et al., 2009), and their recognition accuracies are different. Therefore, although the CRE of 2D-FPCA and 2D-PCA is different at the some time, their recognition accuracies also depend on the accuracy of estimated sample covariance matrix in practice. From Fig. 4(b), it is clear that the top average recognition accuracy of 2D-FPCA is superior to 2D-PCA in the same dimension.

5. Conclusions

In this paper, according to the definition and applications of fractional moments, we give new definitions of the fractional variance and fractional covariance. Furthermore, we give the definition of fractional covariance matrix. Based on this, we propose two new dimensionality reduction techniques: FPCA and 2D-FPCA. The characteristics of FPCA and 2D-FPCA algorithms say that the order between 0 and 1 favours recognition selectivity, which expands the transition recognition ranges of PCA and 2D-PCA. Their advantage is that the CRE is greater and the recognition accuracies are higher than PCA and 2D-PCA when the dimensionality is smaller and the order is also smaller. We also demonstrate that FPCA and 2D-FPCA achieve higher recognition accuracy than typical PCA and 2D-PCA on ORL face database and Yale face database, respectively. In general, the top recognition accuracy of FPCA is about 5% up greater than that of PCA when the order is smaller, and the average recognition accuracy of 2D-FPCA is about 5% greater than that of 2D-PCA and 10% up larger than that of 2D-PCA when the dimensionality is smaller and the order is smaller. Overall, the performances of FPCA and 2D-FPCA are better than those of PCA and 2D-PCA. In the future, we will think that fractional variance and fractional covariance can be applied in the image processing and give valuable information for image analysis.

Acknowledgments

This work is supported by the project of Science and Technology Department of Sichuan Province (No. 2011JY0077), and the project of Chengdu City Economic and Information Technology Commission (No. 201201014).

References

Belhumeur, P. N., Hespanha, J. P., & Kriegman, D. J. (1997). Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 19*(7), 711–720.

Belkin, M., & Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation, 15*(6), 1373–1396.

Cheng, Z.-D., Zhang, Y.-J., & Fan, X. (2009). Criteria for 2DPCA superior to PCA in image feature extraction. *Chinese Journal of Engineering Mathematics, 26*(6), 951–961. in Chinese.

Härdle, W., & Simar, L. (2007). *Applied multivariate statistical analysis* (2nd ed.). New York: Springer.

Jenssen, R. (2010). Kernel entropy component analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 32*(5), 847–860.

Kim, M., Kim, D., & Lee, S. (2003). Face recognition using the embedded HMM with second-order block-specific observations. *Pattern Recognition, 36*(11), 2723–2735.

Kozubowski, T. J. (2001). Fractional moment estimation of Linnik and Mittag–Leffler parameters. *Mathematical and Computer Modelling, 34*, 1023–1035.

Li, Z., Lin, D., & Tang, X. (2009). Nonparametric discriminant analysis for face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 31*(4), 755–761.

Li, M., & Yuan, B. (2005). 2D-LDA: A statistical linear discriminant analysis for image matrix. *Pattern Recognition Letters, 26*(5), 527–532.

Lu, J., Plataniotis, K., & Venetsanopoulos, A. (2003). Face recognition using kernel direct discriminant analysis algorithms. *IEEE Transactions on Neural Networks, 14*(1), 117–126.

Martinez, A. M., & Kak, A. C. (2001). PCA versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 23*(2), 228–233.

Masashi, O., & Mitsuyo, K. (1968). Minimization of eigenvalues of a matrix and optimality of principal components. *The Annals of Mathematical Statistics, 39*(3), 859–863.

Min, W. L., Lu, K., & He, X. F. (2004). Locality pursuit embedding. *Pattern Recognition, 37*(4), 781–788.

Nigmatullin, R. R. (2006). The statistics of the fractional moments: Is there any chance to 'read quantitatively' any randomness? *Signal Processing, 86*, 2529–2547.

Qiu, G., & Fang, J. (2008). Classification in an informative sample subspace. *Pattern Recognition, 41*(3), 949–960.

Saul, L. K., Weinberger, K. Q., Ham, J. H., Sha, F., & Lee, D. D. (2005). *Spectral methods for dimensionality reduction*. MIT Press.

Swets, D. L., & Weng, J. (1996). Using discriminative eigenfeatures for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 18*(8), 831–836.

Tallis, G. M., & Light, R. (1968). The use of fractional moments for estimating the parameters of a mixed exponential distribution. *Technometrics, 10*(1), 161–175.

Wang, L., Wang, X., Zhang, X., & Feng, J. (2005). The equivalence of two-dimensional PCA to line-based PCA. *Pattern Recognition Letters, 26*(1), 57–60.

Yang, J., & Yang, J. Y. (2002). From image vector to matrix: a straightforward image projection technique – IMPCA vs. PCA. *Pattern Recognition, 35*(9), 1997–1999.

Yang, J., & Zhang, D. (2004). Two-dimensional PCA: A new approach to appearance-based face representation and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 26*(1), 131–137.