

Enhanced Incremental Bi-Directional Principal Component Analysis with Forgetting Factors

Chien Shing Ooi, Kah Phooi Seng
Faculty of Science and Technology
Sunway University,
46150 Selangor, Malaysia

Li-Minn Ang
School of Engineering
Edith Cowan University,
Joondalup,
WA 6027, Australia.

Abstract— Feature extraction plays an important role in face recognition system as it can reduce dimensions and reserve the most significant features which need to be classified and recognized. Principal Component Analysis (PCA) has been one of the popular techniques that used in pattern recognition related research areas. Researches have been also carried out to improve the performance of this technique, mainly based on tensor type and incremental type. Incremental Bi-Directional Principal Component Analysis (IBDPCA) is one of the latest improved versions of PCA which combined the merits from tensor and incremental type. However, IBDPCA lacks of the moderations between the latest and previous data when updating the means. This can leads to difficulty in evaluating the data accurately due to larger size of previous data, and also more memory waste. This paper proposed a technique which overcomes the limitations by adopting the IBDPCA with forgetting factors, in order to down-weight the previous overloaded data with relevant factors. To evaluate the proposed technique, two experiments were carried out to compare it with IBDPCA on two different databases: FERET and CMU PIE. The experiment results indicate the better performance in recognition rate by using the proposed algorithm.

Keywords—Principal Component Analysis, incremental, forgetting factors, PCA, IBDPCA

I. Introduction

Principal Component Analysis (PCA) [1] as a popular and efficient feature extraction method, has been widely used in various pattern recognition related areas to reduce the dimensionality of training data. However, the original technique requires the transformation from image matrix (2-dimensions) to column vector (1 dimension) which caused more computational time used in performing data training. Besides, if the training sample is in large size, the technique may face the difficulty in evaluating covariance matrix accurately. Recent related researches were carried out to solve the limitations and improve the performance in accuracy and computational cost, and the efforts mainly can be divided into two types: tensor based and incremental based.

Tensor type versions PCA is initially based on the technique called Two-Dimensional Principal Component Analysis (2DPCA) [2] which amended the way of processing the input data (mostly images) from forming one-dimension vector into forming two-dimension image matrix. Instead, the input training data are initially is remained in 2D form and all

training data are formed together into tensor structure. This not only saves the time for reconstructing the pixels into 1-D vectors, but able to improve the performance by using different calculation method on the matrixes. The shortcoming of this method in dealing with one direction of the matrix (either row or column only) was then solved by Bi-Directional Principal Component Analysis (BDPCA) [3] which considered the computation on both row and column direction. Two-Directional two-dimensional Principal Component ((2D)²PCA) which works identical to BDPCA was also reported by [4] that the time consumed in training samples and accuracy has surpassed the PCA and 2DPCA. Incremental based PCA (IPCA) [5] in the other hand, was developed to solve the batch mode training limitation. Batch approach requires all the training samples are ready to be process once the computation process started. In other words, whole training samples are required to train again if a new training sample is added into the training set. With the incremental learning algorithm, the training computational cost is highly reduced especially when training large scale of data. IPCA's main disadvantage is using the original PCA based method which required transforming input matrix data to 1D vector before the subspace computation.

One of the recent efforts of combining these two types is the Incremental Bi-Directional Principal Component Analysis (IBDPCA) [6] It used the merits from BDPCA which provides the two directions of computational process on input matrix and also adopted the advantages of incremental learning process from IPCA to replace the batch algorithm with partitioned SVD subspace updating method. The analysis on the technique was reported to have more competitive recognition rate and much shorter computational time compared to BDPCA. However the incremental learning process of IBDPCA took account of both old and new data without any moderations between them. This would cause inaccuracy in evaluating the covariance matrix when training set is increasing and in large scale.

This paper presents the IBDPCA with forgetting factors that improves the accuracy and computational time by moderating the new and previous data. The idea of using forgetting factor in SVD based incremental learning was introduced by Levy and Lindenbaum [7] to provide better estimation on the more recent images. The method was adopted and analyzed on IPCA-based visual tracking research because of its advantages in updating the subspace with new

images in the application. This paper is organized as follows. In Section II, the theories that used in IBDPCA-FF are briefly presented. In Section III, the algorithm of FF-IBDPCA is presented and simulation results are shown in Section IV and V to compare the performance of the proposed method with the previous methods and conclude the paper.

II. Previous Works

A. Singular Value Decomposition (SVD)

As one of the most powerful approaches in linear subspace analysis, *SVD* is a basic and useful mathematical tool in face recognition. Muller et al. [8] describes the way *SVD* can be applied to solve image processing related problems, as well as the use of *SVD* to obtain eigen-faces or the result of PCA, which preserved the important features of facial images for face recognition.

The *SVD* of an $m \times n$ matrix A can be represented as $A = USV^T$, where $U \in \mathbb{R}_{m \times m}$, $V \in \mathbb{R}_{n \times n}$ are orthogonal square matrices, and $\Sigma \in \mathbb{R}_{m \times n}$ is diagonal storing the singular values. For the use of face recognition, the matrix U is representing the subspace which identical to the eigenvectors with sorted highest values of eigenvalues.

B. Incremental Bi-Directional Principal Components Analysis (IBDPCA)

To eliminate the shortcoming of the batch mode training process in BDPCA, IBDPCA [6] was developed by using the SVD-based incremental learning process. Firstly, the input images, $\{A_1, A_2 \dots A_K\}$ can be formed into the tensor structure, A_{tensor} whereby A is the image matrix for image from first to K number of samples of the tensor set. For initialization, it computes the eigenvalues and eigenvectors of matrices S_{col} and S_{row} via BDPCA method in both column and row directions. Alternatively, it can also unfold sample tensor A along the first two modes to get matrices $A^{(1)}$ and $A^{(2)}$, then apply *SVD* to obtain the corresponding singular systems [Note: $A^{(1)}$ is the original $m \times n$ matrix image which represents the row vectors, $A^{(2)}$ is the transposed $n \times m$ image which represents the column vectors.]:

$$[U_{k_{col}} \ D^{(1)} \ V] = SVD(A^{(1)}), \quad (1)$$

$$[V_{k_{row}} \ D^{(2)} \ V] = SVD(A^{(2)}), \quad (2)$$

When a new image sample A_{new} in the data stream is captured for incremental learning of BDPCA, it means the third mode of A needs to plus one, and the whole tensor updates to

$$A_{tensor} = \{A, A_{new}\}.$$

The parameters A_{new} is used for calculating new mean, B in the *SVD*-updating procedure and then carry out the incremental learning of $A^{(1)}$ and $A^{(2)}$. Recur to the proposition in Section 2A, directly update the left singular vectors of

centered $A^{(1)}$ and $A^{(2)}$, equivalent to the incremental learning of eigenspace of S_{col} and S_{row} , i.e. BDPCA method.

C. Forgetting Factors

Forgetting factor is used to down-weight the contributions of the earlier data. According to [7], to moderate between the old and new trained data, forgetting factor can be incorporate in the SVD subspace update by scaling down the old data by the scalar factor from 0 to 1, whereby there is no effect from forgetting factor when the value is 1. The paper also proved and stated a lemma that ‘‘A forgetting factor of f reduces the contribution of each block of data to the overall covariance modeled by an additional factor of f^2 at each SVD update.’’ The main parameter that affected by forgetting factor is the mean. In the process of incremental learning, mean is updated in every loop. By using the forgetting factor, the contribution of the mean of the previous trained data is reduced to decrease the contribution of the previous covariance matrix. In IPCA that proposed by [5] the mean is multiplied by the forgetting factor on each update with:

$$I_Z = \left(\frac{fn}{fn+m} \right) I_X + \left(\frac{m}{fn+m} \right) I_Y \quad (3)$$

where the I_X , I_Y , and I_Z , are the old, new and total mean respectively, while n is the number of old data represented in rows and m is the latest trained data’s column number.

III. Incremental Bi-Directional Principal Components Analysis with Forgetting Factors (IBDPCA-FF)

In this section, a new IBDPCA algorithm that extended its updating procedure with adopting the forgetting factors is presented. Similar as IBDPCA, the stages of our proposed method is required to have two main parts: Initialization and SVD Updating.

A. Initialization

The algorithm is started with initialization which firstly obtains the center image. The initial set of the training samples are used for computing the initial subspace, singular values for row and column directions, and mean, by using the methods as stated in equation (1) and (2).

B. Mean and Subspace Updating

The incremental learning process begins with mean updating and ends with small subspace updating, and the steps are repeated again until the partitioned batched of added set is completely trained. Similar as IBDPCA, the mean updating step updates the mean by taking account of the mean of the latest training data, $M_{A_{new}}$ and the previous mean value, M_A .

In our proposed method, the number of previously trained images, K is multiplied by the forgetting factor (where $f = 1.0$ indicates does not carry any effect to the updates) to moderate the old data with new one:

$$M = \left(\frac{fK}{fK + 1} \right) M_A + \left(\frac{1}{fK + 1} \right) M_{A_{new}} \quad (4)$$

$$B = M_{A_{new}} - M_A \quad (5)$$

The new mean, B is then obtained from (5) by subtracting the new sample's mean with the old mean.

$$[Q_u, R_u] = qr(B - U_{k_{col}} U_{k_{col}}^T B) \quad (6)$$

$$[Q_v, R_v] = qr(B^T - V_{k_{row}} V_{k_{row}}^T B^T) \quad (7)$$

$$[\tilde{U}, \tilde{D}_1, \tilde{V}_1] = SVD \begin{pmatrix} fD^{(1)} & U_{k_{col}}^T B \\ 0 & R_u \end{pmatrix} \quad (8)$$

$$[\tilde{V}, \tilde{D}_2, \tilde{V}_2] = SVD \begin{pmatrix} fD^{(2)} & V_{k_{row}}^T B^T \\ 0 & R_v \end{pmatrix} \quad (9)$$

After applying the QR decomposition method as shown in (6) and (7), the newly formed small SVD left singular value can be obtained from equation (8) and (9). The forgetting factor is again used to multiply into the input value of the compacted singular value to downsize the contributions of the previous data. The steps of (4) to (9) are repeated for each samples in added set. The final trained subspaces, $U_{k_{col}}$ and $U_{k_{row}}$ can be formed by using the equations (10) and (11).

$$U_{k_{col}} = [U_{k_{col}}, Q_U] \tilde{U} \quad (10)$$

$$U_{k_{row}} = [U_{k_{row}}, Q_V] \tilde{V} \quad (11)$$

IV. Experiment Result and Analysis

The experiments with different approaches have been carried out to simulate and compare the performances of the proposed method with the other related techniques such as IBDPCA, and IPCA with forgetting factor (IPCA-FF). Using the databases – FERET [9] and CMU PIE [10], the three different modes of incremental techniques evaluations are carried out to examine the recognition rate in different situation. The Nearest Neighbor classifier is used for the classification part to calculate the shortest Euclidean distance. The equation for this method can be shown as equation (12):

$$\arg \min_{y_{test}} \|y_{test} - y_{train}\|^2 \quad (12)$$

A. Recognition rate evaluation using FERET database

By using FERET database, the performance of our proposed method can be evaluated from various conditions on subjects' faces. FERET is a database that built up with face images that captured from different angles in different time (up to 5 years). Six frontal images of each subject are

randomly chosen for simulation and totally 72 subjects are selected from the database. Before each simulation is performed, all the images are preprocessed to be cropped to only face region and resized to 46 X 56 resolutions. Histogram equalization is also applied to the images after the resized images are converted to gray color.

1) Sample-by-sample incremental mode simulation

The 432 images that chosen from the database for this simulation are arranged into three different sets: Initial set, added set and test set. Each subject is set to have three images and one image for initial and added training set respectively. Then the remained two images of the subject are used for testing purpose.

This incremental mode is aimed to evaluate the performance of techniques when the initial set is constructed with few samples from different subjects and when the incremental process is updating the mean and subspace with a samples from each subjects. The simulation is carried out by varying the reduced dimensionality (or number of components). From the resulting plotted graphs that showed in **Figure 1**, it can be observed that the recognition rates of both proposed IBDPCA-FF and IBDPCA are quite closed to each other and could not give consistently higher results.

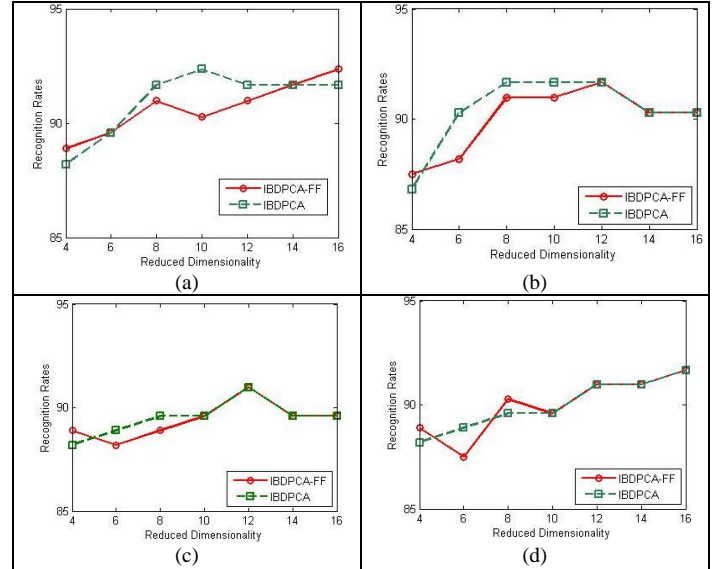


Figure 1. Recognition Rates of IBDPCA-FF with sample-by-sample mode on FERET database, where reduced dimensionality represents k_{row} varied from 4 to 16. (a) $k_{col} = 6$, (b) $k_{col} = 8$, (c) $k_{col} = 10$, (d) $k_{col} = 12$.

2) Class-by-class incremental mode simulation

In the data partition for class-to-class incremental mode, initial set is constructed with 30 classes and added set is constructed with the remained 42 classes. Four images from each class are chosen for training, while the remained two images are used for testing.

This incremental mode is aimed to evaluate the performance of techniques when the initial set is constructed with samples from fixed number of subjects and added set is constructed with the same number of samples but non-repeated subjects. Similar like the previous mode, this simulation is carried out by varying the reduced dimensionality as well. Based on the result shown in plotted graphs of Figure 2, it is easy to observe that the IBDP-CA-FF is relatively in consistent to have better recognition rate than IBDP-CA.

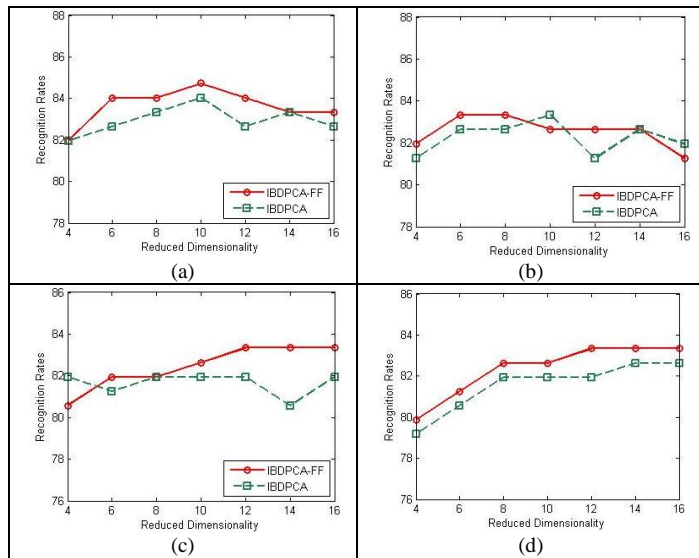


Figure 2. Recognition Rates of IBDP-CA-FF with class-by-class mode on FERET database, where reduced dimensionality represents k_{row} varied from 4 to 16. (a) $k_{col} = 6$, (b) $k_{col} = 8$, (c) $k_{col} = 10$, (d) $k_{col} = 12$.

3) Simulation for evolution with newly added samples

The third mode of simulation aim to show and compare the learning representations updates through time. During the updating process, the face recognition is carried out to examine the current training set with the test set whenever a new sample is added. The simulations are performed in the two modes based on sample-by-sample and class-by class. By setting the number of features $k_{row} = 10$ and $k_{col} = 10$, the performance for techniques IBDP-CA-FF, IBDP-CA and IP-CA-FF are evaluated. By observing from the results in Figure 3 (a) and (b), the performance of IBDP-CA-FF shows the best recognition rates compared to other two. IP-CA although is adopted with forgetting factors, the concatenation of 2-D images to 1-D column vectors makes the training matrix difficult to evaluate accurately. This problem is also called as curse of dimensionality or small size sample (3s) [3]. Compared to the original IBDP-CA, the proposed version shows slightly higher recognition rates during the incremental process.

B. Recognition rate evaluation using CMU PIE database

We also evaluate the performance of our proposed technique based on CMU PIE database. It is constructed by 41368 face

images varied from pose, illuminations, and expressions. Similar as the previous simulations, we have the algorithms tested on sample-by-sample and class-by-class incremental modes. Based on the images under illuminations, 64 classes have chosen and 43 images are randomly selected from each class. For sample-by-sample mode, each class we selected 35 images from for initial set, 3 images for added set and 5 images for testing. For class-by-class mode, each class we selected 38 images for training and 5 images for testing. We constructed the initial set with 33 classes and 31 classes for added set. The simulation result shows in Figure 4 indicates that it remained the good recognition rate as the previous simulation on FERET database,

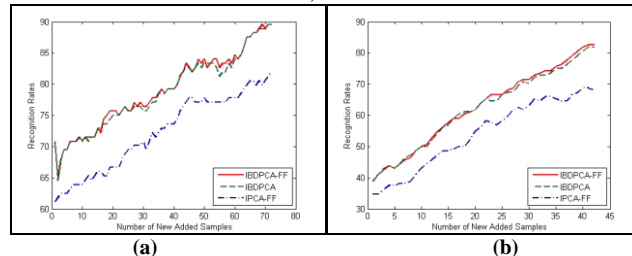


Figure 3. Actual Evolution of new adding samples in: (a) sample-by sample mode on FERET database. (b) class-by-class mode on FERET database

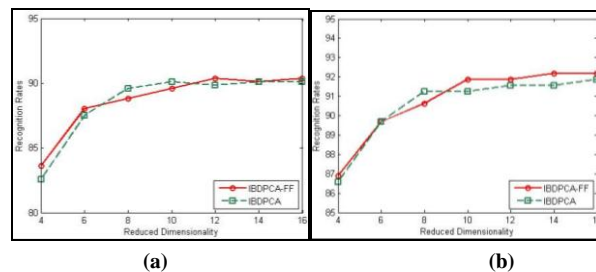


Figure 4. Recognition Rates of IBDP-CA-FF with (a) sample-by sample mode (b) class-by-class mode on CMU PIE database where reduced dimensionality represents k_{row} varied from 4 to 16 and $k_{col} = 6$.

v. Conclusion

In this paper, we proposed an enhanced version of incremental bi-directional principal component analysis with forgetting factors. With the forgetting factors, the performance of the incremental method can be enhanced by moderating the old data with new data and reduced the complexity of the trained data if the amount of new added samples is large. Series of simulations are performed in different modes such as sample-by-sample, class-by-class and actual evolution with newly added samples update, to evaluate the performance of the proposed technique based on the two popular databases, FERET and CMU PIE. In overall, the results proved that the recognition rates can be increased by adopting the forgetting factors in IBDP-CA.

References

- [1] M. Turk , A. Pentland, "Eigenfaces for recognition," *J. Cogn.Neurosci.*, vol. 3, no. 1, pp. 71–86, Mar. 1991.
- [2] Caikou,C.,Jingyu,Y.,"Essence of Two-dimensional Principal Component Analysis," *Int. Conf. on Comp. Inte. & Sec.*, pp. 280-282.2010..
- [3] W. Zuo, D. Zhang, J. Yang, and K. Wang, "BDPCA plus LDA: A novel fast feature extraction technique for face recognition," *IEEE Trans. Syst.,Man, Cybern. B, Cybern.*, vol. 36, no. 4, pp. 946–953, Aug. 2006.
- [4] Choi Y., Tokumoto, T. Minh L., Ozawa, S., "Incremental two-dimensional two-directional principal component analysis (I(2D)²PCA) for face recognition," *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, vol., no., pp.1493-1496, 22-27 May 2011.
- [5] Ross,D.,Jongwoo,L.,R.,Lee,M.,Yang.,"Incremental Learning for Robust Visual Tracking," *Int. Jour. Of Comp. Vis.*, pp. 125-141, 2008..
- [6] Chuan-Xian, R., Dao-Qing, D., "Incremental Learning of Bidirectional Principal Components for Face Recognition," *Patt. Rec.*, vol. 43, no. 1, pp. 318-330, January 2010.
- [7] Levy, A. and Lindenbaum, M., "Sequential Karhunen–Loeve Basis Extraction and its Application to Images," *IEEE Trans. Ima. proc.*, vol. 9, no. 8, pp. 1371-1374, August 2000.
- [8] Muller,N.,Magaia,L.,Herbst,M., Singular value decomposition, eigenfaces, and 3D reconstructions. *SIAM Review*, Vol. 46 No. 3, pp. 518–545, 2004.