Modular Image Principal Component Analysis for Handwritten Digits Recognition

José Francisco Pereira, Victor M. O. Alves, George D. C. Cavalcanti and Tsang Ing Ren
Center of Informatics
Federal University of Pernambuco
Recife, Brazil
{jfp,vmoa,gdcc,tir}@cin.ufpe.br

Abstract—Principal Components Analysis (PCA) is a powerful tool for feature extraction that it is used in different research areas. Many variations of the PCA technique were proposed to improve the standard PCA. In this work, MIMPCA is applied to the problem of handwritten recognition. The method (MIMPCA) extracts local and structural digits information to obtain a representative feature set. Experimental results obtained over MNIST digit database achieved better recognition rates than standard PCA-based techniques.

Keywords-PCA, handwritten digits recognition, feature extraction.

I. INTRODUCTION

Handwritten recognition is a very important research area related to image processing, pattern analysis and classification. It can be applied to problems like: postal code recognition in post offices, signature recognition, digital transposition of historical documents in ASCII format, touch-screen device input recognition and others. It can be used also in bank's checks processing, historical documents indexing [1] and others.

The main contribution of this paper is to improve the handwritten recognition accuracy rates using a novel feature extraction method based on Principal Components Analysis (PCA).

PCA based recognition methods are not very accurate when local information vary considerably. This work applies a new technique that aims to combine the best aspects of Modular Principal Components Analysis (MPCA) and Image Principal Components Analysis (IMPCA).

This paper is organized as follows. In Section II, a review of the used feature extraction methods is presented. Section III describes the proposed method. Section IV describes the experimental setup and discussion of the results. Finally, in Section V, some concluding remarks are presented.

II. PCA BASED FEATURE EXTRACTION METHODS

Principal Component Analysis is a statistical based approach for pattern analysis. In this technique, all training data set is used for extract pattern´s statistical information. Using this statistical information a dimensionality reduction can be done through the axis of major pattern data variations. So, the major amount of discriminatory information is preserved even at low dimensionality.

PCA based methods for feature extraction have been largely used in different pattern recognition tasks, such as: face recognition, industrial robotic and handwritten recognition. However, the standard PCA does not explore local feature changes. Hence, global accuracy can be improved by increasing local feature significance in the pattern analysis process. Furthermore, in standard PCA, the 2D image matrices must be transformed into a 1D vector. As result, the image is represented by a high dimensional vector. Consequently, it is difficult to evaluate the covariance matrix accurately due to its large size and relative small number of training examples.

The next two approaches aim to eliminate the limitation of the PCA-based techniques. In section A. the Two-dimensional approach (IMPCA) is presented followed by the Modular approach (MPCA) explained in section B.

A. *Two-Dimensional Principal Components Analysis (IMPCA)*

Two-Dimensional Principal Component Analysis (or Image PCA, IMPCA) [2] is a PCA-based approach that represents each image as a two-dimensional matrix instead of one dimensional vector. In this technique, the original image matrix does not need to be converted into a vector. In this way, the computational cost of computing the covariance matrix and extract eigenvectors are considerably smaller than the required by standard PCA techniques.

Each principal component is obtained by multiplying the eigenvector extracted from the covariance matrix with the normalized image representation. However, using this method each principal component is a vector instead of a traditional scalar value of others PCA-based techniques. Moreover, in general, is not enough to have only one principal component, consequently, a matrix is obtained in final image representation.

As a result, IMPCA has at least two important advantages over PCA. First, it is easy to calculate the covariance matrix because of its low dimensionality and more accurate statistical representation of the train patterns. Second, less time is required to compute the final image representation. However, the technique is not as efficient as PCA in terms of storage requirements.

B. Modular Principal Components Analysis (MPCA)

The PCA-based techniques for feature extraction are not efficient taking into account local variations of the images. Local feature variations are important in image analysis due to its potential discriminatory power. Under conditions where the local information varies considerably the projections of the images vary significantly from the images in normal circumstances. Hence, under this condition is difficult to identify the digits correctly.

Addressing to overcome this limitation the Modular PCA [3] was proposed. In this technique, the images are divided in smaller regions without any previous image analysis then the feature extraction is done over them. In this way, the local projections will represent more accurately the region its cover. By the way, local variation that affects only small regions of the image, like noise, will damage smoothly the digit representation.

Consequently, it is expected that improved recognition rates can be obtained by following the modular PCA approach. However, if the images were divided into very small regions its global information may be lost and the accuracy of this method may deteriorate.

In MPCA only one mean image and covariance matrix are defined for all regions. The principal components are obtained by multiplying each image region vector by the eigenvector extracted from the covariance matrix. Consequently, if a digit image is divided in N regions each principal component will be a vector of size N because one component will be extracted for each region.

III. Modular Image Principal Component Analysis (MIMPCA)

The PCA-based recognition methods are not very efficient under conditions where local features vary, since it considers the global information of each image. For example, the different shapes for writing the same digit. Some regions of them are very similar while others differ considerably from equivalent regions of original images. Under these conditions the weighted vectors that represent the image vary considerably from that one of normal digit. So, the accuracy of the technique is significantly affected by these changes.

The modular approach is also robust in cases where images are locally affected. This technique takes advantage of the unaffected regions of the digit to improve its accuracy rate. From this point of view, the technique will have good recognition performance comparing to the techniques based on standard PCA due to its modular approach.

The Modular Image Principal Component Analysis (MIMPCA) adopts the modular approach for reducing local variations effects and combines with two-dimensional approach for local feature extraction improvement. Detailed explanation about the technique is described below.

Let $I_1, I_2, \cdots, I_M$ be the training set of digit images. So, I_M denotes an image of size $K \ x \ L$ in the training set, represented by a matrix of the same size.

In this method, each image is divided into X pieces horizontally and Y pieces vertically. Therefore, the original image is divided into N sub-image where $N = X \ x \ Y$ and the size of each sub-image is equal to $K \cdot L \ / \ N$ pixels. These sub-images can be represented mathematically as

$$I_{mij} = I_m(i, j) \ \forall i, j$$

where i varies from 1 to X and j varies from 1 to Y, thus $I_{[m,i,j]}$ represents the sub-images of coordinates i, j of the m-th image in the training set.

In this technique just one average image is obtained for all sub-images. The average image is calculated as

$$\bar{A} = \frac{1}{M} \sum_{m=1}^{M} \bar{A}_m$$

where $\overline{A_m}$ corresponds to the average image of m-th class.

The next step is to normalize all sub-images by subtracting them from the global mean

$$Y_{mij} = I_{mij} - \bar{\bar{A}} \quad \forall \ m, i, j$$

Based on the sub-images matrices the covariance matrix can be calculated as defined below.

$$C = \frac{1}{M} \sum_{m=1}^{M} C_m$$

where C_m corresponds to the class covariance matrix.

It is important to observe that there is just one covariance and mean matrix for all the sub-images. Experimentally it was observed that using only one average and covariance matrix, the final system precision is improved. This result is explained by considering that using N image means and covariance matrixes corresponds to apply the IMPCA technique for independent pictures, which it can to lead to a loss of the global information present in the original image.

The eigenvectors associated with the largest eigenvalues extracted from C will be represented by $E_1, E_2, E_3, \cdots, E_V$. The principal components are computed from the eigenvectors as shown in following equation.

$$W_{mij} = E_v^T \cdot (I_{mij} - A), \quad \forall \ m, i, j, v$$

where v takes the values $1, 2, \cdots, V$, i and j vary from 1 to K and L, respectively, and m represents the m-th image from the reference set and varies from 1 to M. Each weight $W_{[mij]}$ is a L-dimensional vector of coefficients and it corresponds to a 1D projection of the data. But, in general, it is not enough to have only one eigenvector. So, each sub-image will be represented by a $L \ x \ V$ matrix.

In this way, the final projection matrix (P) is defined as $P = [E_1^T, E_2^T, \cdots, v]$. Now each projected sub-image sample can be computed as a simple matrix multiplication as defined in below equation.

$$Y_{test,i,j} = P \cdot (I_{test,i,j} - \bar{A}), \quad \forall \ m, i, j$$

Therefore, for representing an entire image, which is divided in N sub-images, it is necessary N matrix containing $L \times V$ coefficients. This is a disadvantage of MIMPCA compared to the original Modular PCA technique: its storage requirements.

IV. EXPERIMENTS AND DISCUSSION

In this section the database and classifiers are presented and all experiments are discussed.

A. Database

The experiments were conducted using the MNIST numerical digits database [4]. The database contains 70,000 gray-scale handwritten numerical digits images. It is a subset of NIST handwritten database. All images were normalized to 28x28 pixels, based on their center of mass. The database was divided in three datasets: 50,000 samples for training, 10,000 samples for validation and 10,000 samples for testing.

B. Classifiers

For experimental validation of the proposed method, three different classifiers were used: K-nearest neighbor [5], feed-forward multilayer perceptron and pattern distributor [6].

Pattern Distributor: It is a quite recent approach of modular neural network. Modular neural networks aim to divide the problem into subsets of more simply problems. Each sub-problem is handled for one module. In the end, the achievements of all modules are combined for obtain the final response of the classifier. This approach frequently is better than monolitics neural networks, either accuracy rates and training speed.

C. Experimental Setup

The experiments were made as follow. Firstly, the MNIST database was processed with the three features extraction algorithms analyzed (MPCA, IMPCA and MIMPCA). We have used 1 to 8 dimensions for every algorithm. This yields 24 different databases. For modular approaches (MPCA and MIMPCA) the experiments were performed using two different configurations: 2x2 and 2x1. For K-nn classifier the 2x1 partition reaches the best accuracy rate while the 2x2 was the best configuration for non-linear classifiers.

For K-nn only the original database division for train and test samples sets was used, so, no standard deviation was obtained. Experimental results shown that 3-nn reached the best recognition performance over all experiments.

For multilayer perceptron, we have used two layers (hidden and output layers). The training was made with resilient backpropagation algorithm with 200 hidden neurons, obtained empirically. Other parameters were: logistic sigmoid as activation function, maximum of 500 iterations and learning rate equals to 0.1.

For pattern distributor, we have used the grouping algorithm described by Alves and Cavalcanti [7]. This algorithm aims to make the task decomposition through a confusion matrix, obtained by some classifier. The most confusing classes are grouped. The architecture is depicted on Figure 1, where every module is equivalent to a multilayer perceptron. All modules are identical to the monolithic MLP already described.

Every experiment was made 10 times, with different random weights initialization, except for K-nn.

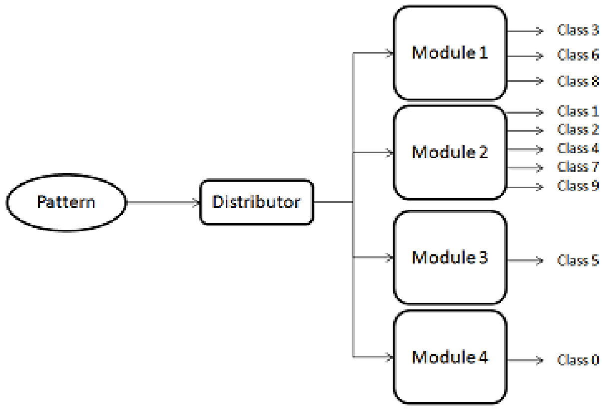

Figure 1. Pattern distributor architecture used. Every module is responsible for a sub-set of the problem.

D. Results

Following the setup described, one obtained the results shown in Table I to III. Because of 10 time execution, for the non-linear classifiers, the results are with the standard deviation in parenthesis. Result tables show the accuracy rate of every feature extraction model with every classifier and 1 to 8 dimensions. According to the results, the proposed method was the best choice for every classifier used. Only the best results were presented in this work.

Comparing the three classifiers, the K-nn presented better results than other non-linear models, for all feature extraction analyzed. The accuracy difference between the three classifiers is depicted on Figure 2.

The results show also that the IMPCA method was better integrated with Pattern Distributor classifier than with K-nn classifier. On the other hand, the MPCA method was better integrated with K-nn than Pattern Distributor classifier.

As expected, in general, the pattern distributor classifier presented better results than simple MLP. Also, the IMPCA feature extraction outperforms MPCA, except for K-nn classification.

As showed in obtained results the MIMPCA technique outperforms the based ones methods in all the experiments. These results evince the feature extraction power of proposed technique for handwritten recognition problem. The performance improvement obtained using MIMPCA can be explained by the combination of 2D and modular feature extraction processes.

Comparing the results of Modular Two-Dimensional approach to the performance of standard PCA the MIMPCA is largely better the PCA using low dimensionality data. However, the best final system accuracy is very similar. For MIMPCA technique the best results (96.98%) was reached using only eight principal components while the PCA technique reaches the best results (97.01%) using thirty principal components.

The processing time for feature extraction of all training database is detailed in Table IV.

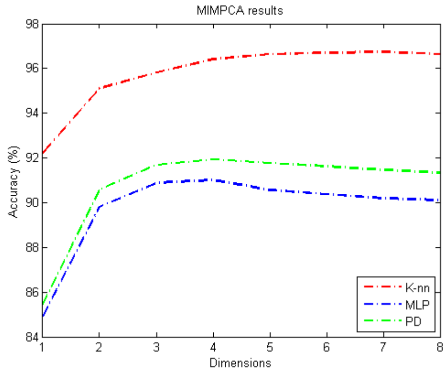

Figure 2. Graphical results obtained by every classifier with MIMPCA feature extraction.

TABLE I. K-NN RESULTS (IN PERCENT).

Dim.	MIMPCA	MPCA	IMPCA
1	**83,12**	84.09	41.84
2	**91,79**	92.07	72.91
3	**95,91**	95.80	86.31
4	**96,58**	96.23	89.47
5	**96,85**	96.42	92.76
6	**96,78**	95.70	93.65
7	**96,95**	96.18	94.63
8	**96.98**	96.12	95.20

TABLE II. MULTILAYER PERCEPTRON RESULTS (IN PERCENT) .

Dim.	MIMPCA	MPCA	IMPCA
1	**84.907 (0.4715)**	46.068 (0.44116)	78.009 (0.43897)
2	**89.804 (0.47794)**	67.969 (0.51067)	84.731 (0.33057)
3	**90.881 (0.43317)**	81.231 (0.32057)	89.255 (0.17765)
4	**91.009 (0.30914)**	84.399 (0.32959)	89.497 (0.20892)
5	**90.561 (0.25449)**	86.981 (0.39411)	89.775 (0.3083)
6	**90.368 (0.33366)**	88.421 (0.32851)	89.825 (0.21225)
7	**90.185 (0.39705)**	88.853 (0.32435)	89.767 (0.25962)
8	**90.113 (0.43581)**	89.593 (0.2275)	89.886 (0.24282)

TABLE III. PATTERN DISTRIBUTOR RESULTS (IN PERCENT).

Dim.	MIMPCA	MPCA	IMPCA
1	**85.454 (0.52007)**	45.462 (0.43951)	77.943 (0.51874)
2	**90.587 (0.26034)**	68.093 (0.37116)	85.076 (0.24901)
3	**91.691 (0.34796)**	81.345 (0.6278)	89.766 (0.22979)
4	**91.935 (0.21083)**	84.708 (0.34058)	90.259 (0.26577)
5	**91.773 (0.32253)**	86.966 (0.37728)	90.491 (0.21553)
6	**91.634 (0.20844)**	88.357 (0.31655)	90.44 (0.25573)
7	**91.462 (0.31675)**	89.249 (0.284)	90.546 (0.20057)
8	**91.334 (0.3839)**	89.904 (0.27625)	90.493 (0.11441)

In handwritten recognition problems based on digit images the local information is very important for final image classification. The same digit can be written with many different shapes by the same person, but, in general, these images keep the same structural pattern. In this way, the proposed technique that explores the local information present in each region improved the final system recognition rate. In this context, MIMPCA explores the local information using a more powerful feature extraction processes base on 2D image representation.

TABLE IV. FEATURE EXTRACTION TIME (S).

Dim.	Time (s)
PCA	**40 (5)**
MPCA	**15 (2)**
IMPCA	**08 (1)**
MIMPCA	**02 (<1)**

V. CONCLUSIONS

The Modular Image PCA method, which is an extension of the Modular PCA and Two-Dimensional PCA methods, was proposed for feature extraction over handwritten recognition problems. The MIMPCA technique was originally proposed for 2D face image feature extraction. However, due to its powerful feature extraction algorithm for local information analysis it can be applied also for handwritten digit recognition. In this kind of problem local image regions are highly discriminatory and can increase the overall system performance.

By the experiments the proposed technique is better than original MPCA and IMPCA for all experiments performed in this work in terms of recognition performance. The computational cost for feature extraction is also much smaller than that one required in based techniques. However, it is not more efficient in terms of storage requirement. One simple strategy to reduce this storage requirement is to use PCA for further data reduction after MIMPCA which, at this moment, will be a quit smaller than the original data.

REFERENCES

[1] V. M. O. Alves, A. L. I. Oliveira, E. R. Silva Jr and C. A. B. Mello, "Handwritten Digit Segmentation in Images of Historical Documents with One-Class Classifiers", 133-140, 2008.

[2] Yang, J., Zhang, D., Frangi, A.F. and Yang, J. Two-dimensional PCA: A new approach to appearance-Based Face Representation and Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, No. 1, pp. 131 - 137, 2004.

[3] Gottmukkal, R. and Asari, V. K. An improved face recognition technique based on modular PCA approach. Pattern Recognition Letters 15 (2004) pp. 429-436, 2004.

[4] Lecun Y., Bottou L., Bengio Y., and Haffner P.. "Gradient-based learning applied to document recognition." *Proceedings of the IEEE*, 86(11):2278-2324, November 1998.

[5] A. Webb, "Statistical Pattern Recognition". John Wiley & Sons, second edition, 2002.

[6] S. U. Guan, C. Bao and T. Neo, "Reduced Pattern Training Based on Task Decomposition Using Pattern Distributor", In: IEEE Transactions on Neural Networks, vol. 18, pp. 1738-1749, 2007.

[7] V. M. O. Alves, G. D. C. Cavalcanti, "Tree Architecture Pattern Distributor: a task decomposition classification approach", In: International Joint Conference on Neural Networks (IJCNN'09), pp.133-140, 2009.

[8] José F. Pereira ; George D. C. Cavalcanti ; Tsang Ing Ren . Modular Image Principal Component Analysis for Face Recognition. In: International Joint Conference on Neural Networks, 2009, Atlanta. IEEE International Joint Conference on Neural Networks, 2009. p. 2481-2486.