EFFICIENT IMAGE REDUCTION FOR FAST INTELLIGIBLE CLASSIFICATION

Marc Joliveau
CIRRELT, Université de Montréal, C.P. 6128, succ. Centre-Ville, Montréal, Québec, H3C 3J7, Canada

Keywords: Image reduction, Image learning, Classification, Dimensionality problem, Image databases.

Abstract: In the past decades, many domains collected great amounts of data, particularly multimedia files, and stored them in large databases. Therefore, area such as similarity search for image learning have received much attention in the recent years. This paper presents an innovative way to strongly reduce dimension and keep relations between components of an image data set. Our method is validated on the Mnist learning database containing 70000 pictures of handwritten digits. Results demonstrate that the proposed approach is very efficient. It allows to accurately classify, learn, and identify digits using very short computation time in comparison with those obtained with original full-size images.

1 INTRODUCTION

For the past few decades, lots of application domains collected large amounts of data, resulting in very large databases. Among them, multimedia files like sounds, videos or images received lots of interest. These files, and specifically images, interact with a large variety of domains like medicine, traffic, security, etc. Aim is generally to automatically identify elements shown in images to label and classify them. However, the great complexity of multimedia files complicates comparisons and similarity searches between them.
This paper introduces an innovative way to strongly reduce dimension of images stored in a database, while conserving relations between them. Proposed method is based on the adaptation of a spatio-temporal dimension reduction method called Space-Time Principal Component Analysis (STPCA), that has been introduced in (Joliveau, 2008).
STPCA method has already been accurately exploited in several contexts with various objectives. First, in (Joliveau and DeVuyst, 2007) authors proposed a variant of STPCA that allows for the handling of missing data with a very weak loss of accuracy. Different applications of STPCA with sensor-based traffic data to interact with Intelligent Transportation Systems (ITS) have been presented in (Joliveau and DeVuyst, 2008; Joliveau, 2008; Bauzer-Medeiros et al., 2008; Bauzer-Medeiros et al., 2009). More precisely, STPCA method is used for the derivation of both typical and atypical spatio-temporal patterns that emit accurate predictions on traffic atypical behaviour through a network.
STPCA allows both to reduce dimension and to summarize data. However, until now, mostly summarizing abilities of the method (i.e., the so-called STPCA estimate denoted by $\hat{\mathbf{X}}_n$) have been exploited. This paper proposes the adaptation of STPCA to databases that contain images in order to benefit from its accuracy of dimensionality reduction, which is provided by the STPCA very low-dimension descriptor (i.e., the so-called reduced-order matrix denoted by $\tilde{\mathbf{X}}_n$). To validate our approach, experiments are performed on the Mnist data set that contains 70000 images of handwritten digits.
The paper is organized as follows. Section 2 gives a brief reminder of STPCA and explains adaptation of this spatio-temporal method to image processing. Then, section 3 presents numerical experiments on Mnist grayscale handwritten images data set. This section proves the high accuracy of our method in reducing dimension of images while conserving enough relations between them. It enables us to classify pictures and identify digits drawn on them. Section 4 briefly presents related methods and discusses possible combinations with our proposed approach. Finally, section 5 concludes the paper and indicates further research directions.

Joliveau M. (2010).
EFFICIENT IMAGE REDUCTION FOR FAST INTELLIGIBLE CLASSIFICATION.
In *Proceedings of the 2nd International Conference on Agents and Artificial Intelligence - Artificial Intelligence*, pages 157-162
DOI: 10.5220/0002691801570162

2 SPACE-TIME PRINCIPAL COMPONENT ANALYSIS

In (Joliveau, 2008), author introduced the Space-Time Component Analysis (STPCA), a new method to develop descriptors of spatio-temporal data series. This method is based on the simultaneous application of a Principal Component Analysis (PCA) (Jolliffe, 1986) to both spatial and temporal dimensions. A brief summary of the method and its adaptation to image reduction is presented in this section.

2.1 Data Conditions to Process Images with STPCA

To be applied, STPCA method needs some conditions on the data to be respected. The method was first introduced in spatio-temporal data reduction context to process traffic sensor-based time series. In this context, STPCA assumes that there are data measured at I different fixed locations (e.g., georeferenced sensor), available for N realizations (e.g., days), and that data are collected at the same frequency for all locations on a time duration J. Measurements can thus be stored into a matrix $\mathbf{X}_n$, where n symbolizes the realization index. Each line of $\mathbf{X}_n$ corresponds to a sensor and each column to a time of measurement.
In order to apply STPCA to a grayscale image data set, we consider that each realization n corresponds to an image. Each image can easily be represented by a matrix $\mathbf{X}_n$ containing the shade of gray of each pixel. Thus, when applying method to images, I represents the number of rows in the image, J its number of columns, and N the number of images in the data set. As STPCA needs all realization matrices $\mathbf{X}_n$ to have the same size, the only condition to apply STPCA to an image data set is that all images have to share the same resolution.

2.2 STPCA Definition

The STPCA method can be decomposed in three steps:

1. Assemble realization matrices horizontally (for realizations row analysis) in a single matrix $\mathbf{Y}$, and vertically (for realizations column analysis) in a matrix $\mathbf{Z}$. According to this definition, matrix $\mathbf{Y}$ represents a big image corresponding to the concatenation of all data set images one beside the other, while matrix $\mathbf{Z}$ represents a big image corresponding to the concatenation of all data set images on top of each other. Matrix $\mathbf{Y}$ thus contains I rows and $(J \times N)$ columns, whereas matrix $\mathbf{Z}$ contains $(I \times N)$ rows and J columns.
2. Compute the singular value decomposition for matrices $\mathbf{Y}$ and $\mathbf{Z}$, as follows. For Gram matrix $\mathbf{G}^{row} = \mathbf{Y}\mathbf{Y}^T$, compute the K first eigenvectors $(\Psi_k)_{k=1\ldots I}$, with $K \ll I$, storing them in a matrix $\mathbf{P}$.
$$\mathbf{P} = col\,(\Psi_1, \Psi_2, \ldots, \Psi_K)$$
For Gram matrix $\mathbf{G}^{col} = \mathbf{Z}^T\mathbf{Z}$, compute the L first eigenvectors $(\Phi_l)_{l=1\ldots J}$, with $L \ll J$, storing them in a matrix $\mathbf{Q}$.
$$\mathbf{Q} = col\,(\Phi_1, \Phi_2, \ldots, \Phi_L)$$
3. Finally, the STPCA estimate $\hat{\mathbf{X}}_n$ of a realization matrix $\mathbf{X}_n$ is defined by:
$$\hat{\mathbf{X}}_n = \mathbf{P}\mathbf{P}^T\,\mathbf{X}_n\,\mathbf{Q}\mathbf{Q}^T.$$

Let us emphasize that the corresponding reduced-order coefficient matrix, also called STPCA descriptor, of the realization n is given by:
$$\tilde{\mathbf{X}}_n = \mathbf{P}^T\,\mathbf{X}_n\,\mathbf{Q},$$
of size $K \times L$ where K and L are chosen to be small.
Until now, all applications using STPCA focuses on STPCA estimate $\hat{\mathbf{X}}_n$ in order to efficiently summarize data (Joliveau, 2008; Bauzer-Medeiros et al., 2008), estimate missing values (Joliveau and DeVuyst, 2007) or extract knowledge (Joliveau, 2008; Joliveau and DeVuyst, 2008; Bauzer-Medeiros et al., 2009). However, reduced-order matrices $\tilde{\mathbf{X}}_n$ offer a very strong dimensionality reduction that could be efficiently exploited for classification or fast similarity search. As already demonstrated, STPCA can be easily adapted to pictures. So we propose to test and validate the use of STPCA reduced-order descriptor $\tilde{\mathbf{X}}_n$ on a grayscale images data set.

3 REDUCTION OF GRAYSCALE IMAGES

The aim of the proposed approach is to efficiently reduce dimension of grayscale images while conserving the most possible of their intelligible relations. All proposed experiments have been processed using a 2.8 GHz processor with 2.5 GB of memory.

3.1 The Mnist Database

The Mnist database[1] contains 70000 images of handwritten digits. This image database is decomposed

[1]Mnist database is available on website http://yann.lecun.com/exdb/mnist/

in two parts: a training set of 60000 pictures and a test set of 10000 pictures. The size of each image is 28×28.

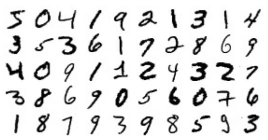

Figure 1: Example of the 50 first Mnist handwritten digit images.

Figure 1 illustrates the 50 first images of Mnist database. Image database comes with a label file containing the corresponding digit of each picture. Mnist is thus considered as "a good database for people who want to try learning techniques and pattern recognition methods on real-world data" (Y. LeCun, 1998, http://yann.lecun.com/exdb/mnist/).

3.2 Image Reconstruction

To give a survey of STPCA performance with images, we first took interest on the STPCA estimate $\hat{\mathbf{X}}_n$ of each picture n from the Mnist training set. An STPCA has thus been performed on this data set with input parameters $I = 28$, $J = 28$ and $N = 60000$. Reduction parameter K and L have been determined such that reduced-order matrices keep the relation between images height and width. So in our case, we only consider values such that $K = L$.

Figure 2: STPCA estimates of Mnist images according to square reduced-order matrices size from 1×1 to 28×28.

Figure 2 shows the STPCA estimate of the 16 first Mnist images according to values of reduction parameters. Each row in the figure refers to one image and each column illustrates the STPCA estimate according to chosen value for K and L. On the first column, we can see the image obtained with $K = L = 1$, on the second column the image obtained with $K = L = 2$ and so on.

The initial analysis indicates that using very weak values of reduction parameters, such as reduced-order matrices are of size 4×4 or 5×5, leads to a sufficient degree of precision for a human to recognize represented digits. These reductions respectively correspond to compression factors of order 50 and 30.

3.3 Inter-Image Relations Conservation

STPCA estimate is interesting as it provides an understandable representation of images. However, in order to think about memory economy and fast computation, we now will consider only reduced-order matrix (or STPCA descriptor) $\tilde{\mathbf{X}}_n$ of each image n.

What we measure first is the impact of the reduction process on the similarity relations between images. We thus perform a STPCA to Mnist training set and then we compute the k-nearest neighbors of each STPCA descriptor matrix $\tilde{\mathbf{X}}_n$ (for $n = 1, \ldots, N$) among the $N-1$ other reduced-order matrices. Distance between two matrices is computed according to the Frobenius norm of their difference. The Frobenius norm $\|\mathbf{A}\|_F$ of a matrix $\mathbf{A}$ is given by:

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^{I} \sum_{j=1}^{J} (\mathbf{A}_{i,j})^2},$$

where $\mathbf{A}_{i,j}$ corresponds to the value at row i and column j of matrix $\mathbf{A}$.

Quality of the obtained classification is then measured by the recognition rate $\tau_{rec}(n)$ of each image n. Recognition rate of an image n indicates the proportion of neighbors of n that share the same label as it, i.e., the proportion of its neighbors that shows the same digit as n. Recognition rate of an image n is given by:

$$\tau_{rec}(n) = \frac{|\mathcal{N}(n, L(n))|}{|\mathcal{N}(n)|}$$

where $\mathcal{N}(n)$ is the set that contains the k-nearest neighbors of n, $\mathcal{N}(n, L(n))$ is a subset of $\mathcal{N}(n)$ concerning only the neighbors with label equal to $L(n)$, the label of n, and $|\mathcal{N}|$ symbolizes the number of element contained in a set $\mathcal{N}$.

Figure 3 illustrates the average recognition rate obtained on Mnist training set according to the size of STPCA descriptor matrices for four different values of k. The corresponding recognition rate when using the original Mnist images is also represented on each figure by a dashed line.

These tests validate STPCA descriptor accuracy to both strongly reduce dimension and to conserve relations between images. For any value of k, since reduced-order matrices are at least of size 3×3, recognition rate is greater or equal than 80%. Thanks

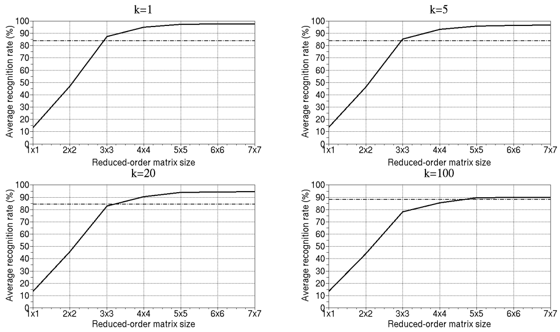

Figure 3: Average recognition rate according to reduced-order matrix size for different values of number of neighbors k.

to noise reduction achieved by our approach, with very small values of k, STPCA descriptor reaches a better recognition rate than those obtained with original images. When $k = 20$ and $k = 100$, since reduced-order matrices size is equal to 4×4 and 5×5, method manages to reach better recognition rates with original data.
Moreover, it must be emphasized that the proposed values of k are very weak (less or equal than 100) in comparison to the number of images in the data set (60000).

3.4 Fast Learning of Digits

Good quality of previous results leads us to use STPCA descriptor for fast learning tasks. Learning the digits using STPCA descriptor can be done as follows.
First, an STPCA is performed on the Mnist training set. This allows us both to learn the eigenmodes matrices $\mathbf{P}$ and $\mathbf{Q}$ and to compute STPCA descriptor of each image in the training set. Then, STPCA descriptor of test set images is computed from matrices $\mathbf{P}$ and $\mathbf{Q}$ learned in the previous step. We assume that we know digit associated to each training set image but, in the process, we ignore digits associated to test set images. The last step consists of computing the k-nearest training set reduced-order matrices of each reduced test image. We finally allocate to each image in the test set the most frequently represented digit among its neighbors. The obtained classification quality is measured by the test error rate that indicates the proportion of irrelevant digit allocations.

Figure 4 illustrates evolution of test error rate according to reduced-order matrices size for different values of k (the number of neighbors considered). Test error rate level obtained with original images of size 28×28 is also represented by a dashed line on each subfigure. Once again, results are surprisingly

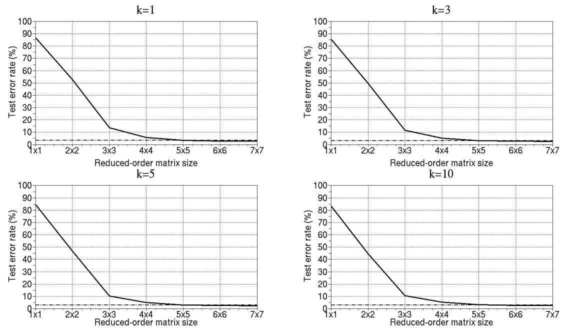

Figure 4: Test error rate according to reduced-order matrices size for different values of k.

good, as for every values of k test error rate is less than 5% since reduced-order matrices are at least of size 4×4. Moreover, for all values of k, since reduced matrices size is at least 5×5, test error rate is weaker than those obtained with original full-size images, reaching values less than 2.5%.

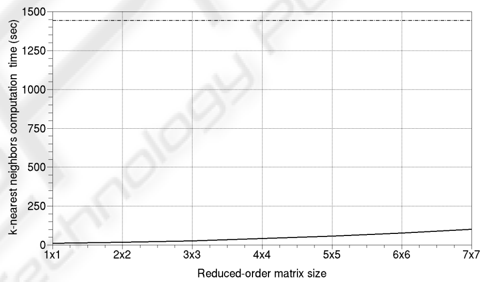

Figure 5: k-nearest neighbors computation times according to reduced-order matrices size.

Figure 5 shows the k-nearest neighbors computation times (in seconds) according to reduced-order matrices size. As chosen values of parameter k are very weak (less than 10) in comparison to the number of element in the data set, k does not significantly influence the computation times. On this figure, we can also see the k-nearest neighbors computation time when using the original 28×28 images represented by a dashed line.
More than providing an accurate learning of digits, STPCA allows to learn them much faster than with original images. For example, computing the k-nearest neighbors of reduced-order matrices of size 5×5 takes 25 times less computation time than with original images. If we refer to figure 4, learning digits with $k = 1$ and reduced-order matrices of this size leads to a very weak test error rate of order 3%. These results are even more surprising when we consider that application of STPCA to both learn eigenmodes and to compute reduced-order matrices is performed in less than 1 minute !

4 COMPARISON TO RELATED WORK

This paper concerns a method to dramatically reduce images while keeping relations between them. Numerical experiments to validate our approach are done through the evaluation of its abilities to learn handwritten digit pictures. This section briefly comments on related work in image reduction and Mnist images learning. It presents the different relations between our work and methods introduced in the literacy.

4.1 Image Reduction

In the last decades, the problem of dimensionality reduction – i.e., finding meaningful low-dimensional structures hidden in high-dimensional observations, received high interests from scientists. When working with images, the aim of methods solving this problem is generally to estimate the underlying geometry of the data set.
Classical techniques principally concern Principal Component Analysis (PCA) and MultiDimensional Scaling (MDS). PCA (Jolliffe, 1986) applies a singular value decomposition on the data, whereas MDS (Kruskal and Wish, 1978) considers pairwise distances between data points (i.e., images) and projects them on a euclidean space such that two similar objects are represented by points close to each other, and two dissimilar objects by faraway points.
More recently, new methods such as Isomap (Tenenbaum et al., 2000) or Local Linear Embeddings (LLE) (Roweis and Saul, 2000) have been introduced to reduce images. Unlike classical approaches, these methods are capable to discover the non-linear degrees of freedom that underlie complex natural observations (e.g., handwritten digits). Isomap applies a MDS from the matrix of geodesic distances between points while LLE uses a weighted neighborhood approach.
Due to the too important level of memory required by MDS, LLE and Isomap when performed on the 70000 images of Mnist, we can only compare our approach to PCA. PCA reduces the dimension of objects in only one direction (i.e., row or column for images) while STPCA reduces both images directions simultaneously, which allows to reach better compression factor (Joliveau, 2008). This property is validated by results shown on table 1 that compares performances of STPCA and PCA when learning digits of Mnist. PCA-Row reduces the dimension in rows direction according to reduction parameter K and PCA-Col reduces the dimension in columns direction according to reduction parameter L. For the three proposed configurations of reduction parameters, $K = L = 5$ (Tab. 1a), $K = L = 6$ (Tab. 1b) and $K = L = 7$ (Tab. 1c), STPCA outperforms PCA for both test error rate and computation time. Thus, STPCA allows to reduce dimensionality more strongly than PCA while conserving more intelligible information.

Table 1: Comparison between PCA and STPCA of test error rate and k-nearest neighbors computation time obtained while learning digits of Mnist.

	PCA-Row	PCA-Col	STPCA
Test error rate	4.91%	5.06%	3.24%
knn CPU-Time	265 sec	285 sec	57 sec

(a) - Parameters $K = 5$ & $L = 5$

	PCA-Row	PCA-Col	STPCA
Test error rate	4.49%	4.89%	2.84%
knn CPU-Time	315 sec	331 sec	77 sec

(b) - Parameters $K = 6$ & $L = 6$

	PCA-Row	PCA-Col	STPCA
Test error rate	4.39%	4.88%	2.70%
knn CPU-Time	367 sec	380 sec	100 sec

(c) - Parameters $K = 7$ & $L = 7$

4.2 Digits Learning

Since Mnist data set publication, several methods have been introduced to learn the handwritten digits it contains. Authors of (LeCun et al., 1998) give an assessment of performances of many learning techniques such as linear classifiers, k-nearest neighbors, virtual SVM, multi-layer neural networks, and convolutionnal net, to identify Mnist digits. Some of these techniques require preprocessing on images such as deskewing or subsampling. Other methods have also been proposed: the combination of trainable feature extractor with Support Vector Machines (SVM) in (Lauer et al., 2007), k-nearest neighbors with non-linear deformation in (Keysers et al., 2007), and the combination of unsupervised sparse feature and a SVM in (Labusch et al., 2008). At the moment, best results are obtained by (Ranzato et al., 2006) whose large convolution net reached an error test of only 0.39%.
Compared to the mentioned methods, best test error obtained with our learning approach (2.42%) is not as good as those presented in (Ranzato et al., 2006) but it is better than those obtained by other proposed approaches, including all multi-layer neural networks from (LeCun et al., 1998). It must also be noted that our approach is not directly designed for learning tasks but to intelligibly reduce dimensionality of images. Thus, it can be used as preprocessing for these learning approaches to reduce the data size, to

improve their computation times, and probably to increase results quality due to the denoising effect of STPCA.

5 CONCLUSIONS

This paper presents the adaptation of the Space Time Principal Component Analysis (STPCA), a bi-dimensional innovative reduction method, to images data sets. Unlike to the other STPCA applications proposed in the literacy that essentially exploit summarizing abilities of the method, our approach focuses more on its dimension reduction capacities provided by the STPCA descriptor.

Numerical experiments on the Mnist data set containing 70000 handwritten digit pictures validate accuracy of the proposed approach to both strongly reduce dimension and to conserve relations between images. Denoising effect when reducing images dimension even allows more accurate and intelligible classification with low-dimension images descriptors than with original pictures. Identification of represented digits on Mnist test set images after a learning process on the Mnist training set also demonstrates a very good behaviour of the introduced approach. Such an identification process leads to a weaker test error rate when it is performed from the 5×5 reduced-order matrices computed for each image than those obtained with original images of size 28×28. Using the proposed approach, we thus obtain results at least as accurate and intelligible from reduced data than from original full-size ones with weaker computation times.

Future works concern the use of reduced images as input data in more complex digit learning systems like support vector machines or large convolutional nets, that could improve both their computability and accuracy, as much as its combination to non-linear feature extractors such as LLE or Isomap. Moreover, it could also be interesting to integrate proposed approach to complex databases or datawarehouses management systems.

REFERENCES

Bauzer-Medeiros, C., Joliveau, M., Jomier, G., and De-Vuyst, F. (2008). Managing sensor data on urban traffic. *Advances in Conceptual Modeling – Challenges and Opportunities*, 5232/2008:385–394.

Bauzer-Medeiros, C., Joliveau, M., Jomier, G., and De-Vuyst, F. (2009). Managing sensor traffic data and forecasting unusual behaviour propagation. *Geoinformatica*.

Joliveau, M. (2008). *Reduction of urban traffic time series from georeferenced sensors, and extraction of spatio-temporal series*. PhD thesis, Ecole Centrale Des Arts Et Manufactures (Ecole Centrale Paris).

Joliveau, M. and DeVuyst, F. (2007). Space-time summarization of multisensor time series. case of missing data. In *Proc. of 2007 International Workshop on Spatial and Spatio-Temporal Data Mining*, pages 631–636.

Joliveau, M. and DeVuyst, F. (2008). Recherche de motifs spatio-temporels de cas atypiques pour le trafic routier urbain. *Extraction et Gestion de Connaissances EGC 08, Revue des Nouvelles Technologies de l'Information - RNTI - E11, F. Guillet et B. Trousse*, 2:523–534.

Jolliffe, I. (1986). *Principal component analysis*. Springer-Verlag, New York.

Keysers, D., Deselaers, T., Gollan, C., and Ney, H. (2007). Deformation models for image recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(8):1422–1435.

Kruskal, J. B. and Wish, M. (1978). *Multidimensional Scaling*. Sage Publications.

Labusch, K., Barth, E., and Martinetz, T. (2008). Simple method for high-performance digit recognition based on sparse coding. *IEEE transactions on neural networks*, 19:1885–1889.

Lauer, F., Suen, C., and Bloch, G. (2007). A trainable feature extractor for handwritten digit recognition. *Pattern Recognition*, 40:1816–1824.

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. In *Proc. of the IEEE*, volume 86, pages 2278–2324.

Ranzato, M., Poultney, C., Chopra, S., and LeCun, Y. (2006). Efficient learning of sparse representations with an energy-based model. In *Proc. of Advances in Neural Information Processing System NIPS 2006*.

Roweis, T. and Saul, L. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290:2323–2326.

Tenenbaum, J., de Silva, V., and Langford, J. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, 290:2319–2323.