Eigenspace merging for model updating

Annalisa Franco[§], Alessandra Lumini[§] and Dario Maio[§]
§ DEIS - CSITE - CNR - Università di Bologna, viale Risorgimento 2, 40136 Bologna - ITALY.
E-mail: {afranco, alumini, dmaio}@deis.unibo.it

Abstract

The Karhunen-Loève transform (KLT) is an optimal method for dimensionality reduction, widely applied in image compression, reconstruction and retrieval, pattern recognition and classification. The basic idea consists in evaluating, starting from a set of representative examples, a reduced space, which takes into account the structure of the data distribution as much as possible, and representing each element in such uncorrelated space. Unfortunately KLT has the drawback of requiring a periodical recomputation in presence of a dynamic dataset. This work presents a novel efficient approach to merge multiple eigenspaces, which provides an incremental method to compute an eigenspace model by successively adding new sets of elements. Experimental results show that the merged model grants performances as good as a one obtained by a batch procedure.

1. Introduction

In recent years an increased interest has arisen in developing methods for feature extraction, data representation and effective indexing of high dimensional datasets. Due to the "dimensionality curse", the performance of traditional access methods and data structures rapidly degrade as the dimensionality of the data increases. The most feasible solution to this problem is given by dimensionality reduction, i.e. the extraction of the principal features of the data, possibly minimizing the loss of precision. The most widely used method of dimensionality reduction is the Karhunen-Loève transform (KL transform) [1], a linear transformation that computes the reduced space as the eigenspace spanned by the first eigenvectors (associated to largest eigenvalues) of the data covariance matrix. The drawback of this approach is that it is expensive to compute, and therefore not readily applicable to dynamic datasets.

Current techniques [2][3][4] are mainly suited for static datasets and are inefficient for dynamic environments; only a few works in the literature address this problem and propose incremental methods, which compute a model by successively updating an earlier eigenspace each time new examples are received. Some methods are based on the recomputation of the eigenvectors from a simplified covariance matrix [5][6][7], and only a work [8] takes into account a possible shift of the mean of the space.

In this paper, we propose a new efficient approach for eigenspace model updating in dynamic databases. Two-space and multi-space merging are proposed for successively updating an initial model as new data become available rather than performing a batch recomputation from the entire dataset. Several case studies are presented and experimental results are given to prove that this technique reduces the computation time without compromising the representation precision of the model: the space calculated by the merging procedure does not differ considerably from the one obtained by a batch method. Moreover eigenspace updating can successfully resolve the problem of dealing with very large datasets which requires a lot of memory and CPU to compute the batch eigenspace: the global space can be created by merging two or more eigenspaces constructed on the sets of a partition of the training set.

The paper is organized as follows: in section 2 a brief theoretical background on the KL transform and formal definition of eigenspace is given; section 3 presents the new approach for eigenspace merging, in section 4 experimental results obtained on databases of real and randomly generated data are reported. Finally, in section 7 some concluding remarks are given.

2. Eigenspaces: fundamentals and notation

Given a set of m n-dimensional observations $P = \{\mathbf{x}_i \in \Re^n \mid i = 1...m_P\}$, derived from the patterns of interest (for example, an image may be vectorized by postponing its rows), the k dimensional eigenspace S_P related to P is obtained by selecting the first k eigenvectors from the KL transformed space of P. $S_P = [\bar{\mathbf{x}}_P, \Phi_P, \Lambda_P, m_P]$, where:

$\bar{\mathbf{x}}_P = \frac{1}{m}\sum_{\mathbf{x}\in P}\mathbf{x}$ is the mean vector,

1051-4651/02 $17.00 (c) 2002 IEEE

$\Phi_P = [\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \ldots \boldsymbol{\varphi}_k]$ and $\Lambda_P = [\lambda_1, \lambda_2, \ldots, \lambda_k]^{\mathrm{T}} \cdot \mathbf{I}$ are the matrices of the first k eigenvectors and eigenvalues ($\lambda_1 \geq \lambda_2 \geq \ldots \lambda_k$) of the data covariance matrix $\mathbf{C}_P$.

A map between the original space and the reduced eigenspace is performed by means of two operators of projection and back-projection:

- the projection $\mathbf{y}$ of a vector $\mathbf{x} \in \Re^n$ into the space S_P, is $\mathbf{y} = \Phi_P{}^{\mathrm{T}} (\mathbf{x} - \bar{\mathbf{x}}_P)$
- the back-projection $\mathbf{x}'$, into the original space, of a vector $\mathbf{y} \in \Re^k$ belonging to $\boldsymbol{S_P}$ is $\mathbf{x}' = \Phi_P \mathbf{y} + \bar{\mathbf{x}}_P$

Typically k is chosen in order to retain the most significant eigenvectors and to discard the ones that are presumed to be negligible, since they are associated with very low eigenvalues. A measure of the amount of information lost is given by the mean-square error ε_P:

$$\varepsilon_P = \frac{1}{m} \sum_{\mathbf{x} \in P} \left\| \mathbf{x} - \bar{\mathbf{x}}_P \right\|_2^2 - \left\| \Phi_P{}^T (\mathbf{x} - \bar{\mathbf{x}}_P) \right\|_2^2$$

where the argument of the summation is the reconstruction error of the vectors $\mathbf{x} \in P$ in the space S_P.

The mean-square error can be conceived as a measure of how much S_P is well-suited to represent P; it can be shown [1] that this error corresponds to the sum of the $(n-k)$ smallest $\mathbf{C}_P$'s eigenvalues: $\varepsilon_P = \sum_{j=k+1}^{n} \lambda_j$

3. Eigenspace merging

This section concerns the problem of merging two or more eigenspaces: first the case of two spaces is analyzed, which is of great interest in pattern classification for the updating of an eigenspace representation when a new set of observations is given, then the more general case of multi-space merging is investigated, which can be useful for eigenspace construction from many reduced sets of elements of the training set. In figure 1 an example of two-space merging in $\Re^2$ is given: each space is represented by 2 lines, corresponding to its eigenvectors, and an ellipse, whose axes are proportional to its eigenvalues.

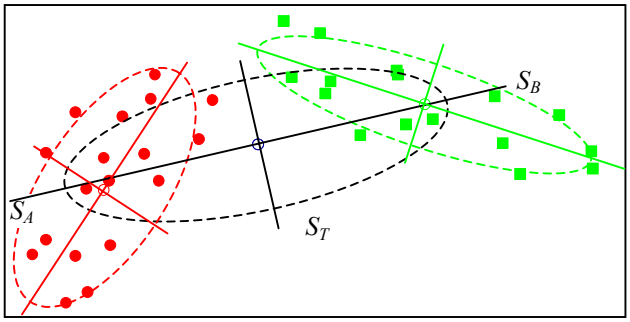

Figure 1. An example of multi-space merging.

3.1. Two-space merging

The mathematical formulation of the two-space merging problem is the following: given two datasets $A = \{\mathbf{x}_{i_A} \in \Re^n \mid i_A = 1,..,m_A\}$, $B = \{\mathbf{x}_{i_B} \in \Re^n \mid i_B = 1,..,m_B\}$ and their related eigenspaces $S_A = [\bar{\mathbf{x}}_A, \Phi_A, \Lambda_A, m_A]$, $S_B = [\bar{\mathbf{x}}_B, \Phi_B, \Lambda_B, m_B]$, we want to calculate the eigenspace $S_T = [\bar{\mathbf{x}}_T, \Phi_T, \Lambda_T, m_T]$ associated to the set $A+B$. Moreover, we suppose that the two original eigenspaces have been reduced to dimension k_A and k_B, respectively; this means that the merged space could be slightly different from the optimal one derived from the whole set of observations. We can derive the total number of observations, the combined mean and covariance matrix as follows:

$$m_T = m_A + m_B \qquad \bar{\mathbf{x}}_T = \frac{\bar{\mathbf{x}}_A \cdot m_A + \bar{\mathbf{x}}_B \cdot m_B}{m_T}$$

$$\mathbf{C}_T = \frac{m_A}{m_T} \mathbf{C}_A + \frac{m_B}{m_T} \mathbf{C}_B + \frac{m_A m_B}{m_T{}^2} (\bar{\mathbf{x}}_A - \bar{\mathbf{x}}_B)(\bar{\mathbf{x}}_A - \bar{\mathbf{x}}_B)^{\mathrm{T}}$$

In the evaluation of the last step only an approximated version of $\mathbf{C}_A$ and $\mathbf{C}_B$ can be used, since we suppose that some eigenvectors of the original space have been discarded: $\mathbf{C}_X \approx \tilde{\mathbf{C}}_X = \Phi_X \Lambda_X \Phi_X{}^{\mathrm{T}}$, for $X \in \{A, B\}$. By substituting these matrices in the equation above we can obtain a good approximation $\tilde{\mathbf{C}}_T$ of the covariance matrix of the total dataset. A direct computation of eigenvectors and eigenvalues of $\tilde{\mathbf{C}}_T \in \Re^{n \times n}$ is very time-consuming, so that the advantages of an updating disappear.

In order to reduce this drawback we propose to transform the problem into a small sample size problem and then calculate the eigenvalues and the eigenvectors of a reduced matrix [1]. Let us rewrite the two correlation matrices as ($X \in [A, B]$):

$$\tilde{\mathbf{C}}_X = \frac{1}{m_X} \Psi_X \Psi_X{}^{\mathrm{T}} \text{ where } \Psi_X = \Phi_X (m_X \cdot \Lambda_X)^{\frac{1}{2}}$$

and let $\mathbf{p}$ be a new vector spanning the subspace defined by the difference of the two mean vectors:

$$\mathbf{p} = \left(\frac{m_A \cdot m_B}{m_T} \right)^{\frac{1}{2}} \cdot (\bar{\mathbf{x}}_A - \bar{\mathbf{x}}_B)$$

then we can express the covariance matrix $\tilde{\mathbf{C}}_T$ as:

$$\tilde{\mathbf{C}}_T = \frac{1}{m_T} \left[\Psi_A \Psi_A{}^{\mathrm{T}} + \Psi_B \Psi_B{}^{\mathrm{T}} + \mathbf{p}\mathbf{p}^{\mathrm{T}} \right] = \frac{1}{m_T} \Psi_T \Psi_T{}^{\mathrm{T}}$$

where $\Psi_T = [\Psi_A \mid \Psi_B \mid \mathbf{p}]$ is a $n \times r$ matrix ($r = k_A + k_B + 1$). Therefore, the rank of $\tilde{\mathbf{C}}_T$ is less or equal to $min(n, r)$. Since the intrinsic dimension (rank) of the eigenproblem is lower than the dimension of the covariance matrix, owing to dimensionality reduction inside the two original

eigenspaces ($r<<n$), the following procedure is more efficient than a direct calculation:

- Construction of the matrix $\mathbf{R}$ as

  $$\mathbf{R} = \frac{1}{m_T}{\Psi_T}^{\mathrm{T}}\Psi_T = \frac{1}{m_T}[\Psi_A \mid \Psi_B \mid \mathbf{p}]^{\mathrm{T}}[\Psi_A \mid \Psi_B \mid \mathbf{p}],$$

  which can be calculated by simply performing the following operations:

  $$\Gamma = {\Psi_A}^{\mathrm{T}}\Psi_B \qquad \mathbf{q} = {\Psi_A}^{\mathrm{T}}\mathbf{p} \qquad \mathbf{l} = {\Psi_B}^{\mathrm{T}}\mathbf{p}$$

  $$\mathbf{R} = \frac{1}{m_T}\begin{bmatrix} m_A\Lambda_A & \Gamma & \mathbf{q} \\ \Gamma^{\mathrm{T}} & m_B\Lambda_B & \mathbf{l} \\ \mathbf{q}^{\mathrm{T}} & \mathbf{l}^{\mathrm{T}} & \mathbf{p}^{\mathrm{T}}\mathbf{p} \end{bmatrix}$$

- Computation of the eigenvalues Λ_R and the eigenvectors Φ_R of $\mathbf{R}$
- Calculating eigenvalues and eigenvectors of $\tilde{\mathbf{C}}_T$ as $\Lambda_T \equiv_r \Lambda_R$ and $\Phi_T = \Psi_T\Phi_R \cdot (m_T \cdot \Lambda_R)^{-\frac{1}{2}}$.

The resulting eigenspace has dimension equal to *r,* which can be further reduced to a lower value k_T.

The merging operator is clearly commutative, as it is easy to prove from the steps used to derive the total space. The merging operator is also associative, but only if the partial space has not been reduced to a lower dimension before the last merging.

3.2. Multi-space merging

Given s sets of points $A_j = \left\{\mathbf{x}_{i_j} \in \Re^n \,\middle|\, i_j = 1,\ldots,m_j\right\}$, j=1..s and their related eigenspaces $S_j = [\bar{\mathbf{x}}_j, \Phi_j, \Lambda_j, m_j]$ (where the subscript j indicates the set A_j), each reduced to the dimension k_j, the merging operator should return the eigenspace $S_T = [\bar{\mathbf{x}}_T, \Phi_T, \Lambda_T, m_T]$ associated to the set $T = \sum_{j=1..s} A_j$. The total number of observations, the combined mean and the total covariance matrix may be obtained by a simple extension of the equations in section 3.1 as follows:

$$m_T = \sum_{j=1}^{s} m_j \qquad \bar{\mathbf{x}}_T = \frac{\sum_{j=1}^{s} \bar{\mathbf{x}}_j \cdot m_j}{m_T}$$

$$\mathbf{C}_T = \sum_{j=1}^{s}\left[\frac{m_j}{m_T}\cdot\mathbf{C}_j\right] + \sum_{i=1}^{s-1}\sum_{j=i+1}^{s}\left[\frac{m_i m_j}{{m_T}^2}(\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_j)(\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_j)^{\mathrm{T}}\right]$$

As in the two-space case, the explicit calculation of the combined covariance matrix is not required. The complete procedure is an extension of the previous one:

- Calculate $\Psi_j = \Phi_j\left(m_j \cdot \Lambda_j\right)^{1/2}$ for each S_j, j = 1..s
- Compute $\mathbf{p}_{ij} = \left(\frac{m_i \cdot m_j}{m_T}\right)^{1/2} \cdot (\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_j)$ for each couple (S_i, S_j), i = 1.. $s-1$, j = (i+1)..s
- Construct $\Psi_T \in \Re^{n\times r}$, $r = \left(\sum_{j=1}^{s} k_j + \frac{s(s-1)}{2}\right)$

  $$\Psi_T = \left[\Psi_1 \mid \Psi_2 \mid \ldots \mid \Psi_s \mid \mathbf{p}_{12} \mid \ldots \mid \mathbf{p}_{(s-1)s}\right]$$

- Resolve the small-size eigenproblem $\mathbf{R} = \frac{1}{m_T}{\Psi_T}^{\mathrm{T}}\Psi_T$ obtaining eigenvalues Λ_R and eigenvectors Φ_R of $\mathbf{R}$
- Calculate the eigenvalues and eigenvectors of $\tilde{\mathbf{C}}_T$ as $\Lambda_T \equiv_r \Lambda_R$ and $\Phi_T = \Psi_T\Phi_R \cdot (m_T \cdot \Lambda_R)^{-\frac{1}{2}}$

The resulting eigenspace has dimension equal to *r*.

We studied a new mathematical model to estimate the mean-square error for the resulting eigenspace, not reported here for the sake of brevity. Interested readers can refer to [9] for a complete error analysis.

4. Experimental results

In this section we present some experiments carried out in order to evaluate the new approach for eigenspace merging with respect to the batch procedure, considering efficacy in terms of approximation errors and efficiency in terms of computational costs. We have conducted our experiments on two real datasets: the ORL database of faces [10] and a set of handwritten digits [11].

4.1. Convergence tests

We first test the convergence of the merging procedure by comparing the eigenspace S_{BATCH} calculated by the batch algorithm with the one S_{MERGE} resulting from merging two or more subspaces. We compare eigenspaces using the following measures:

Approximation error: the mean-square error (section 2) measures how much a space is well suited for representing a set of points.

Classification accuracy: when dimensionality reduction is used for classification/recognition, coupled with a nearest neighbor criterion in the reduced space, we can compare the accuracy α (percentage of correct classification).

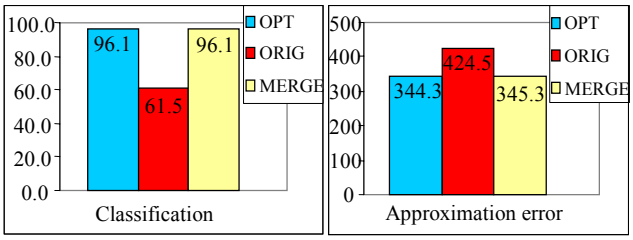

Figure 2. ORL dataset.

The ORL dataset has been used to simulate person identification in a dynamical environment: we used 60 images (6 for each person) to construct an eigenspace, and

we tested accuracy after 3 new persons have been added.

The approximation and classification errors reported in figure 2 are related to 3 eigenspaces: the original S_{ORIG}, an updated one S_{MERGE} obtained by merging S_{ORIG} with the subspaces related to the 3 new faces and the optimal one S_{OPT} obtained by batch recomputation.

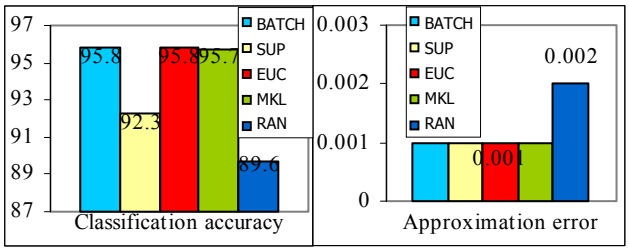

Figure 3. Digits dataset.

The convergence of the merging procedure with respect to the batch procedure is strictly related to the quality of the representation given by the single subspaces merged. We used the set of handwritten digits in order to compare different merged eigenspaces related to the same set of observations. We made several partitions of the set into 13 classes: *SUP* is a supervised partition (each class contains images of the same digit), *MKL* is obtained by the MKL clustering algorithm [12], *EUC* is an Euclidean partition and *RAN* is a random partition. Figure 3 makes a comparison between the batch eigenspace S_{BATCH} related to the whole dataset and the 4 eigenspaces S_{SUP}, S_{MKL}, S_{EUC} and S_{RAN} obtained by merging the eigenspaces related to the partitioned set (k=10 for all spaces). Only S_{RAN} presents a noticeable difference from S_{BATCH} , proving that a good partitioning of the dataset before constructing and merging subspaces is a fundamental requirement in order to achieve a good approximation of the batch eigenspace.

4.2. Execution time tests

We compare (figure 4) the execution time taken for the evaluation of the complete eigenspace for the ORL dataset by following 4 different strategies: the batch procedure over the whole dataset (named T) and 3 different merging approaches, starting from 4 equal-size subsets (named A, B, C, D).

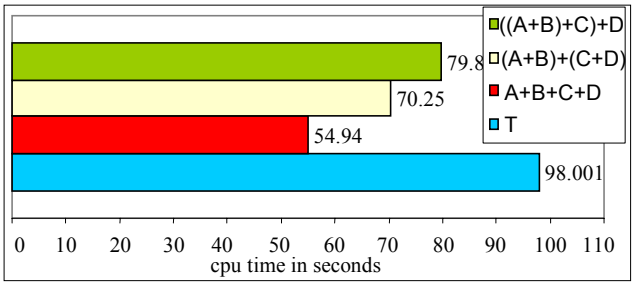

Figure 4. Time to obtain a complete eigenspace following different merging approaches.

The proposed approaches consist in a multiple merging of the 4 eigenspaces at once (A+B+C+D), in two partial plus one final merging ((A+B)+(C+D)) and in a sequential two-space merging (((A+B)+C)+D). The final space is the same in all the cases, since no dimensionality reduction has been performed during the intermediate steps.

5. Conclusion and future work

In this work we have introduced a new approach to compute an eigenspace by successively updating an initial space as new examples become available. This approach is shown to work well for dynamic environments where an eigenspace model has to be updated as new data become available. The concepts here discussed have been developed in the context of research aimed to define a hierarchical data structure [13] where eigenspaces are used to model the nodes and merging is used to update the structure.

References

[1] K. Fukunaga, *Statistical Pattern Recognition*, Academic Press, San Diego, 1990.

[2] M. Turk and A. Pentland, "Eigenfaces for recognition", *J. of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71-86, 1991.

[3] M. Uenohara and T. Kanade, "Use of Fourier and Karhunen-Loeve Decomposition for Fast Pattern Matching with a Large Set of Templates", *IEEE tPami*, vol. 19, no. 8, pp. 891-898, Aug. 1997.

[4] A. Pentland, R. Picard, and S. Sclaroff, "Photobook: Tools for Content-Based Manipulation of Image Databases", *Proc. SPIE Storage and Retr. of Image & Video Databases II*, vol. 2185, SPIE, Bellingham, Wash., pp. 34-47. 1994.

[5] J.R. Bunch, C.P. Nielsen and D.C. Sorenson. "Rank-one modification of the symmetric eigenproblem". *Numerische Mathematik*, vol. 31, pp. 31-48, 1978.

[6] S. Chandrasekaran et al., "An eigenspace update algorithm for image analysis". *Graphical Models and Image Processing*, vol. 59, no. 5, pp. 321–332, 1997.

[7] R.D. DeGroat and R. Roberts. "Efficient, numerically stabilized rank-one eigenstructure updating". *IEEE tASSP*, vol. 38, no. 2, pp. 301-316, 1990.

[8] P. M. Hall, A. D. Marshall and R. R. Martin, "Merging and Splitting Eigenspace Models", *IEEE Trans. PAMI*, vol. 22, no. 9, pp. 1042-1049, 2000.

[9] A. Franco, A. Lumini, D. Maio, "An efficient algorithm for eigenspace updating and merging", *submitted for publication.*

[10] ORL DB of faces. AT&T Laboratories, Cambridge, UK.

[11] E. Alpaydin, C. Kaynak, Dep. of Comp. Eng., Bogazici University, 80815 Istanbul Turkey

[12] R. Cappelli, D. Maio and D. Maltoni, "Multi-space KL for Pattern Representation and Classification", *IEEE Trans. PAMI* vol.23, no.9, pp.977-996, September 2001

[13] R. Cappelli, A. Lumini e D. Maio, "MKL-tree: una struttura gerarchica per l'indicizzazione di dati Multidimensionali", *SEBD2001*, Venice (I), June 2001.