Incremental Learning of Locally Orthogonal Subspaces for Set-based Object Recognition

Tae-Kyun Kim[1*], Josef Kittler[2] and Roberto Cipolla[1]
1 : Engineering Department, University of Cambridge, Cambridge, CB2 1PZ, UK
2 : CVSSP, University of Surrey, Guildford, GU2 7XH, UK
* : tkk22@eng.cam.ac.uk, http://mi.eng.cam.ac.uk/∼tkk22

Abstract

Orthogonal subspaces are effective models to represent object image sets (generally any high-dimensional vector sets). Canonical correlation analysis of the orthogonal subspaces provides a good solution to discriminate objects with sets of images. In such a recognition task involving image sets, an efficient learning over a large volume of image sets, which may be increasing over time, is important. In this paper, an incremental learning method of orthogonal subspaces is proposed by updating the principal components of the class correlation and total correlation matrices separately, yielding the same solution as the batch computation with far lower computational cost. A novel concept of local orthogonality is further proposed to cope with non-linear manifolds of data vectors and find a more optimal solution of orthogonal subspaces for a certain neighbouring object image sets. In the experiments using 700 face image sets, the locally orthogonal subspaces outperformed the orthogonal subspaces as well as relevant state-of-the-art methods in accuracy. Note that the locally orthogonal subspaces are also amenable to incremental updating due to their linear property.

1 Introduction

The popularity of the methods of object recognition based image sets has been increasing because of their greater accuracy and robustness as compared with the approaches exploiting a single image as input [1, 2, 3, 4, 5, 7]. Of the methods that compare a set to a set, canonical correlations [1] of linear subspaces have attracted much attention with the benefits of robust and computationally efficient matching when dealing with changing conditions of data acquisition and large volumes of data as input for decision making [4, 5, 7, 8]. The previous method called Constrained Mutual Subspace Method (CMSM) finds the constrained subspace where the entire class populations exhibit small variance [8, 10]. Then, each class subspace is projected on this constrained subspace to create a model and compared by canonical correlations with the new data. They have shown the constrained subspace improves the accuracy of the simple canonical correlation method [4]. However, CMSM does not have any principled way to select the dimensionality of the constrained subspace despite the sensitivity of accuracy on this parameter. In [7], an optimal linear discriminant function is proposed to find the components to maximize the canonical

[1] It is also called canonical angles or principal angles.

BMVC 2006 doi:10.5244/C.20.58

correlations of the within-class subspaces and minimize the canonical correlations of the between-class subspaces with more desirable feature selection. However, the iterative optimization in the method is computationally costly, making incremental update rather difficult. In [5, 6] nonlinear extensions of canonical correlations have been proposed. The non-linear extensions of CMSM have also been noted in [8, 10]. However, the benefits of these methods in terms of high accuracy were compromised by the expensive computational cost in both on-line matching and model learning.

Oja and Kittler introduced the concept of orthogonal subspaces for effective class feature extraction [11]. In their methods the class-specific components which are orthogonal to those of all the other classes, are extracted. Then a new vector is classified into the class which has the minimum distance between the vector and the orthogonal subspaces. It will be shown that these orthogonal subspaces can be effectively combined with the canonical correlation analysis for set-based recognition. Whereas a single vector is a new input in the classical study [11], the classification of a set of image vectors is better handled by the canonical correlation analysis of the orthogonal subspaces. Interestingly, the principle of the Orthogonal Subspace Method (OSM) is very close to that of CMSM. Both methods find the components which maximally represent the class data while minimizing the variances of all the other classes. However, OSM provides the optimal way to choose the number of such components based on the eigenvalues, while CMSM requires an empirical setting for the number of the components, which is practically unfavorable.

In this paper, an incremental method of learning orthogonal subspaces is presented. Practically, the assumptions are that a complete set of training samples is not given in advance and the execution of the batch-computation [2], whenever new data is presented, is too expensive in both time and space. An efficient algorithm for updating the orthogonal subspaces is greatly needed to accumulate the information conveyed by new data so that the method's future accuracy is enhanced. The orthogonal subspace method seeks the class-specific components which maximize the ratio of the variances of the i-th class correlation matrix over the total correlation matrix. The incremental learning algorithm of OSM is proposed by separately updating the principal components of the both correlation matrices then computing the orthogonal components only by the updated few principal components. Each update is benefited in both time and space by the concept of the sufficient spanning set used for the incremental Principal Component Analysis (PCA) in [14]. The proposed incremental OSM solution is identical to that of the batch-mode computation of orthogonal subspaces but at a far lower computational and space cost. The proposed method also allows sets of vectors to be added in a single update, thus avoiding too frequent updates of the orthogonal subspaces. The Locally Orthogonal Subspace Method (LOSM) has been further proposed to deal with non-linear manifolds of data for accuracy improvement of OSM. Each class subspace is better distinguished from its rival classes by constructing the subspace more orthogonal to its neighboring subspaces [3]. The methodology for efficient matching and incremental updating of LOSM is also presented.

The next section summarizes the orthogonal subspaces and canonical correlation methods. The incremental learning of the orthogonal subspaces is proposed in Section 3 and the locally orthogonal subspaces in Section 4. The section 5 presents a comparative evaluation for both accuracy and time-efficiency using the 700 face image sequences.

[2] All existing data is re-used with new data for computing a new model.

[3] In practice, it is hard to achieve orthogonality of a class to all different classes.

Notations	Descriptions
C, N	number of classes, dimension of input data
M_i, M_T	number of data points of the i-th class and total data
R_i, R_T	correlation matrix of the i-th class and total data
U_i	orthogonal component matrix of i-th class
P_i, Λ_i	eigenvector and eigenvalue matrices of R_i
P_T, Λ_T	eigenvector and eigenvalue matrices of R_T
d_i, d_T	number of sufficient components of the i-th class and total data
U_j^i	locally orthogonal component matrix of j-th class to i-th class

Table 1: ***Notations.***

2 Orthogonal Subspaces and Canonical Correlations

See Table 1 for the important notations used throughout the paper.

Orthogonal subspaces. Denote the correlation matrices of the C classes by R_i, $i = 1,...,C$, where $R_i = 1/M_i \sum xx^T$ and M_i is the number of data vectors in the i-th class. Let w_i denote the respective priori probabilities. Then, matrix $R_T = \sum_{i=1}^C w_i R_i$ is the correlation matrix of the mixture of all the classes. The total correlation matrix is decomposed s.t. $P_T^T R_T P_T = \Lambda_T$. Letting $Z = P_T \Lambda_T^{-1/2}$, we have $Z^T R_T Z = I$. This means that matrices $w_i Z^T R_i Z$ and $\sum_{j \neq i} w_j Z^T R_j Z$ have the same eigenvectors but the respective eigenvalues λ_i and $\overline{\lambda_i}$ are related by $\lambda_i = 1 - \overline{\lambda_i}$. Let the matrix, U_i, be constructed from eigenvectors of the i-th class having the eigenvalues equal to unity s.t.

$$w_i U_i^T Z^T R_i Z U_i = I_i, \tag{1}$$

then

$$\sum_{j \neq i} w_j U_i^T Z^T R_j Z U_i = O \quad \text{and} \quad w_j U_i^T Z^T R_j Z U_i = O, \text{ for all } j \neq i, \tag{2}$$

where O is a zero matrix since every matrix $w_j U_i^T Z^T R_j Z U_i$ is positive semi-definite. Assume that the j-th class is also represented by the eigenvectors of $w_j Z^T R_j Z$ having the eigenvalues equal to one s.t. $w_j Z^T R_j Z \simeq U_j U_j^T$. From (2), we have $w_j U_i^T U_j U_j^T U_i = O$, i.e. $U_i^T U_j = O$. This is the definition of the mutually orthogonal subspaces where all the vectors of each subspace are orthogonal to those of the other subspace [11].

Canonical correlations. Of many solutions to compute canonical correlations, which are all equivalent, the Singular Value Decomposition (SVD) solution [12] is presented. Assume that $U_i \in \mathscr{R}^{N \times d}$ and $U_j \in \mathscr{R}^{N \times d}$ form unitary orthogonal bases for two linear subspaces, where N, d are the dimensions of the input vector and the subspaces respectively. Canonical correlations, which are mutually defined as the maximal correlations of any two vectors on the two subspaces, are computed as the singular values of $U_i^T U_j \in \mathscr{R}^{d \times d}$ s.t.

$$U_i^T U_j = Q_L \Lambda Q_R^T \quad \rightarrow \quad Q_L^T U_i^T U_j Q_R = \Lambda = diag(\sigma_1, ..., \sigma_d) \tag{3}$$

where $Q_L^T Q_L = Q_L Q_L^T = Q_R^T Q_R = Q_R Q_R^T = I_d$. Canonical correlations of any orthogonal subspaces are always zeros as $U_i^T U_j = O \rightarrow Q_L^T U_i^T U_j Q_R = O$, where O is a zero matrix. Thus, class discrimination can be performed by analyzing the canonical correlations of the class-wise orthogonal subspaces. The similarity of the subspaces is given as the aggregation of the canonical correlations by

$$\mathscr{S}(U_i, U_j) = \mathrm{tr}(\Lambda), \tag{4}$$

and Nearest Neighbor (NN) classification is performed based on the similarity measure.

3 Incremental Learning of Orthogonal Subspaces

In practice, the eigenvectors having the eigenvalues which are exactly equal to one in (1), do not often exist. Instead, the eigenvectors corresponding to the largest few eigenvalues can be exploited. Note that in the space projected by matrix Z in Section 2, the most important basis vectors for each class which are the eigenvectors corresponding to the largest eigenvalues, are at the same time the least significant basis vectors for the ensemble of the rest of the classes. Thus, the classical orthogonal subspaces (1) can be generalized into the subspaces spanned by the components U_i s.t.

$$w_i U_i^T Z^T R_i Z U_i = \Delta_i, \quad \sum_{j \neq i} w_j U_i^T Z^T R_j Z U_i = I - \Delta_i, \tag{5}$$

where Δ_i is the diagonal matrix corresponding to the largest few eigenvalues. There are many previous studies about incremental PCA, e.g. [14], but the involvement of the successive eigenvalue problems of the correlation matrices in the Orthogonal Subspace Method (OSM) in Section 2 makes the incremental learning difficult. We propose the incremental OSM solution by the three steps: update of the principal components of each class correlation matrix, update of those of the total correlation matrix and the computation of the orthogonal components only using both updated principal component sets. The concept of the sufficient spanning set [14] is conveniently exploited in each step to reduce the dimension of the eigenvalue problems. The proposed method provides the same solution as the batch-mode OSM with far lower computational cost. When new data is added to the existing data set, all existing orthogonal subspace models $U_i,\ i = 1,...,C$ (C is the number of classes) are incrementally updated to get new orthogonal subspaces described by U_i' as follows. Here, we assume the equal priori probabilities for all classes for simplicity.

1) Update of principal components of class correlation matrix. Let the number of samples, eigenvector and eigenvalue matrices corresponding to the first few eigenvalues of the i-th class correlation matrix R_i in the existing data be (M_i, P_i, Λ_i) respectively. The set $(M_i^n, P_i^n, \Lambda_i^n)$ similarly denotes those of the new data. The update is defined as the functional form by

$$\mathscr{F}_1(M_i, P_i, \Lambda_i, M_i^n, P_i^n, \Lambda_i^n) = (M_i', P_i', \Lambda_i'). \tag{6}$$

Note this is applied only to the classes having the new data. As the updated class correlation matrix is $R_i' \simeq \frac{M_i}{M_i'} P_i \Lambda_i P_i^T + \frac{M_i^n}{M_i'} P_i^n \Lambda_i^n P_i^{nT}$ where $M_i' = M_i + M_i^n$, the sufficient spanning set of R_i' can be given as $\Upsilon_i = \mathscr{H}([P_i,\ P_i^n])$, where $\mathscr{H}$ is an orthonormalisation function of column vectors (e.g. QR decomposition). The function $\mathscr{H}$ also eliminates any zero vectors after the orthonormalisation to further reduce the number of the sufficient components. Then, the updated principal components can be written by $P_i' = \Upsilon_i Q_i$, where Q_i is a rotation matrix. By this representation, the eigenproblem of the updated class correlation matrix is changed into a new low dimensional eigenproblem by

$$R_i' \simeq P_i' \Lambda_i' P_i'^T = \Upsilon_i Q_i \Lambda_i' Q_i^T \Upsilon_i^T \quad \rightarrow \quad \Upsilon_i^T (\frac{M_i}{M_i'} P_i \Lambda_i P_i^T + \frac{M_i^n}{M_i'} P_i^n \Lambda_i^n P_i^{nT}) \Upsilon_i \simeq Q_i \Lambda_i' Q_i^T. \quad (7)$$

Note that the new eigenvalue problem requires only $O({d_i}^3)$ computations, where d_i is the number of columns of Υ_i. The total computational cost of this stage takes $O(C^n \times (d_i^3 + min(N, M_i^n)^3))$, where N is the dimension of input space and C^n is the number of classes in the new data given. The latter term is for computing $(M_i^n, P_i^n, \Lambda_i^n)$ from the new data.

2) Update of principal components of total correlation matrix. The subsequent update is described as

$$\mathscr{F}_2(M_T, P_T, \Lambda_T,\ M_i^n, P_i^n, \Lambda_i^n) = (M_T', P_T', \Lambda_T') \qquad i = 1, ..., C^n, \quad (8)$$

where $M_T = \sum_{i=1}^{C} M_i$, P_T, Λ_T are the first few eigenvector and eigenvalue matrices of the total correlation matrix of the existing data. C^n represents the class number of the new data. The updated total correlation matrix is

$$R_T' \simeq \frac{M_T}{M_T'} P_T \Lambda_T P_T^T + \frac{M_T^n}{M_T'} \sum_{i=1}^{C^n} P_i^n \Lambda_i^n P_i^{nT} \quad (9)$$

where $M_T' = M_T + M_T^n$, $M_T^n = \sum M_i^n$. The sufficient spanning set of R_T' can be given as

$$\Upsilon_T = \mathscr{H}([P_T, P_1^n, ..., P_{C^n}^n]) \quad (10)$$

and $P_T' = \Upsilon_T Q_T$,where Q_T is a rotation matrix. Accordingly, the new small dimensional eigenproblem is obtained by

$$R_T' \simeq P_T' \Lambda_T' P_T'^T \quad \rightarrow \quad \Upsilon_T^T (\frac{M_T}{M_T'} P_T \Lambda_T P_T^T + \frac{M_T^n}{M_T'} \sum_{i=1}^{C^n} P_i^n \Lambda_i^n P_i^{nT}) \Upsilon_T \simeq Q_T \Lambda_T' Q_T^T \quad (11)$$

The computation requires $O(d_T^3)$, where d_T^3 is the sufficient number of components of Υ_T. Note that all P_i^n have already been produced at the previous step.

3) Update of orthogonal components of all classes. The final step only exploits the updated principal components of the previous steps, which is defined as

$$\mathscr{F}_3(P_i', \Lambda_i',\ P_T', \Lambda_T') = U_i', \quad i = 1, ..., C. \quad (12)$$

where U_i' denotes the updated orthogonal components of the i-th class data. Let $Z = P_T' {\Lambda_T'}^{-1/2}$, then, $Z^T R_T' Z = I$. The remaining problem is to find the components which maximize the variance of the projected data $Z^T R_i' Z$. The sufficient spanning set of the projection data can be given by $\Phi_i = \mathscr{H}({P_T'}^T P_i')$. Then, the eigenproblem to solve is

$$Z^T R_i' Z = \Phi_i Q_i \Delta_i Q_i^T \Phi_i^T \quad \rightarrow \quad \Phi_i^T Z^T P_i' \Lambda_i' P_i'^T Z \Phi_i = Q_i \Delta_i Q_i^T, \quad (13)$$

where Q_i, Δ_i are a rotation matrix and eigenvalue matrix respectively. The final orthogonal components are given as $U_i' = \Phi_i Q_i,\ i = 1, ..., C$. This computation only takes $O({d_i}^3)$, where d_i is the number of columns of P'_i. Note usually $d_i < d_T$, where d_T is the number of columns of P'_T.

Batch OSM vs. incremental OSM for time and space complexity. The batch computation of OSM for the combined data costs $O(min(N,M'_T)^3 + C \times min(N,M'_i)^3)$, where the former term is for the diagonalization of the total correlation matrix and the latter for the projected data of the C classes (Refer to Section 2 for the batch-mode computation). The batch computation also requires all data vectors or $N \times N$ correlation matrices to be kept track of. By contrast, the proposed incremental solution is much more time-efficient with the costs of $O(C^n \times (d_i^3 + min(N,M_i^n)^3))$, $O(d_T^3)$ and $O(C \times d_i{}^3)$ for the three steps respectively. Note $d_i \ll M'_i$, $d_T \ll M'_T$, $M_i^n \ll M'_i$. The proposed incremental algorithm is also very economical in space costs, which corresponds to the data $(P_i, \Lambda_i, P_T, \Lambda_T)$, $i = 1, ..., C$.

4 Locally Orthogonal Subspace Method (LOSM)

In the generalized orthogonal subspaces (5), the priori probabilities of classes w_j can be better set up to improve the discriminatory powers of the classes with their rival classes. Rather than equal priors for all classes, higher priors are given to the neighboring classes of the i-th class by

$$w_j \rightarrow w_j^i \begin{cases} \propto \mathscr{S}(U_i, U_j) & \text{for} \quad j = 1, ..., C, \ j \neq i, \\ = 0 & \text{for} \quad j = i \end{cases}$$

where $\mathscr{S}$ is the canonical correlation function defined in Section 2. Then, the i-th class locally orthogonal subspace U_i^i is similarly computed as U_i in Section 2 by replacing the total correlation matrix R_T with the class-specific total correlation matrix by $R_T^i = \sum_{j=1}^{C} w_j^i R_j$ and diagonalizing $Z^T R_i Z$. The weights w_j^i can also be binary-valued in the same concept s.t. $w_j^i = 1$, if $\mathscr{S}(U_j, U_i) > thres$, $w_j^i = 0$ otherwise. In this way, the local orthogonality of classes is more emphasized.

Normalization. When a new test set is given, the locally orthogonal components of the new test set are class-wise extracted with R_T^i for $i = 1, ..., C$. If we let U_q^i as the locally orthogonal components of the new test set for the i-th model class, NN recognition is performed with the normalized scores

$$(\mathscr{S}(U_i^i, U_q^i) - m_i)/\sigma_i, \tag{14}$$

where m_i, σ_i are the mean and standard deviation of matching scores of a validation set with the i-th model class. As each class model exploits a different total correlation matrix, the score normalization process is important.

Time-efficient matching. Batch computation of the C locally orthogonal subspaces of a given new test set is time-consuming, which takes $O(C \times min(N, M_q)^3)$, where M_q is the number of vectors in the new test set. This computational cost can be significantly reduced using the update function $\mathscr{F}_3(P_q, \Lambda_q, P_T^i, \Lambda_T^i)$ in Section 3, where P_q, Λ_q are the eigenvector and eigenvalue matrices of the correlation matrix of the new test set and P_T^i, Λ_T^i for the class specific total correlation matrix respectively. Note that this only requires $O(C \times d_q^3)$, d_q is the number of columns of P_q. The subsequent canonical correlation

Figure 1: ***Database:*** *(top) Frames from a typical video sequence from the database used for evaluation. The motion of the user was not controlled, leading to different poses. (bottom) 7 different illumination conditions in databases.*

matching with C orthogonal subspace models is not computationally expensive as it costs $O(C \times d^3)$ (Refer to Section 2), where d is the dimension of the orthogonal subspaces.

Incremental update of LOSM. The computational cost of the incremental locally OSM is increased by that of the update of the components of the C class-specific total correlation matrices, but it is still much cheaper than the batch OSM. First, the principal components of class correlation matrices are updated by $\mathscr{F}_1$ in the previous section. The update of the principal components of the weighted total correlation matrices defined s.t.

$$\mathscr{F}_2'(M_T, P_T^i, \Lambda_T^i, w_j^i, M_j^n, P_j^n, \Lambda_j^n) = (M_T', P_T^{i\,\prime}, \Lambda_T^{i\,\prime}) \qquad i = 1, ..., C, \;\; j = 1, ..., C^n. \tag{15}$$

is achieved as follows. The updated weighted total correlation matrix is given as

$$R_T^{i\,\prime} = \frac{M_T}{M_T'} P_T^i \Lambda_T^i P_T^{i\,T} + \frac{M_T^n}{M_T'} \sum_{j=1}^{C^n} w_j^i P_j^n \Lambda_j^n P_j^{nT}. \tag{16}$$

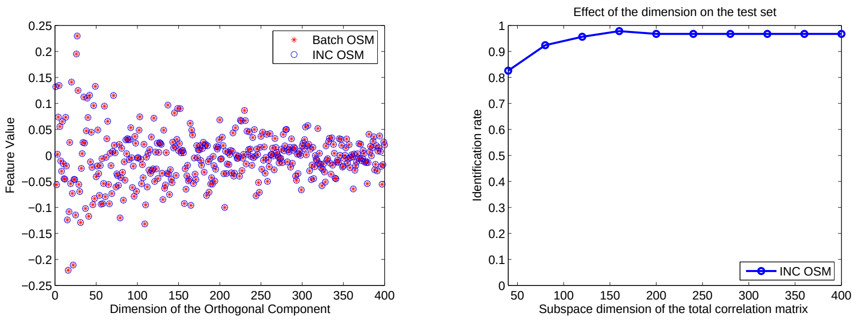
Regardless of the extra weight terms, the sufficient spanning set of $R_T^{i\,\prime}$ is similarly given by $\Upsilon_T^i = \mathscr{H}([P_T^i, P_1^n, ..., P_{C^n}^n])$, then the new eigen-problems and the updated components are similarly given as the second step in Section 3. If we assume that the NN recognition has already been performed for the given new test sets by the scores of $\mathscr{S}(U_i^i, U_j^i), \;\; i = 1, ..., C, \;\; j = 1, ...C_n$, the weights w_j^i can be set up proportionally to these scores. The final locally orthogonal components are also similarly updated by $\mathscr{F}_3$, replacing P_T', Λ_T' with $P_T^{i\,\prime}, \Lambda_T^{i\,\prime}$.

5 Evaluation

We used the face video database with 100 subjects. For each person, 7 video sequences of the individual in arbitrary motion were collected. Each sequence was recorded in a different illumination setting for 10s at 10fps and 320×240 pixel resolution (see Figure 1). Following automatic localization using a cascaded face detector [13] and cropping to the uniform scale, images of faces were histogram equalized. Each sequence is then represented by a set of raster-scanned vectors of the normalized images.

5.1 Accuracy and time complexity of the incremental OSM

The incremental OSM yielded the same solution as the batch-mode OSM for the data merging scenario, where the 100 sequences of 100 face classes of a single illumination



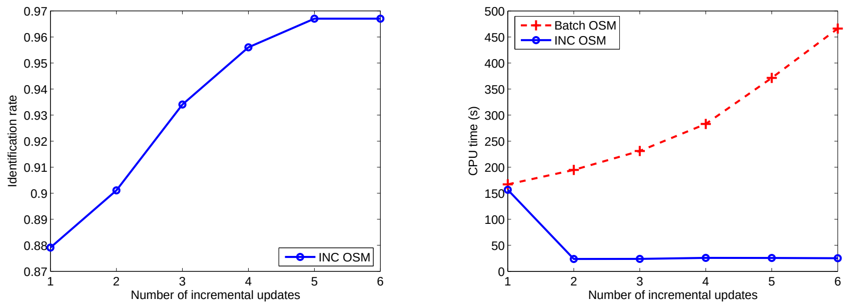
(a) (b)

Figure 2: ***Batch vs. Incremental OSM-1:*** *(a) Example orthogonal components, which are computed by the incremental and the batch-mode, are very alike. (b) Insensitiveness of the incremental OSM to the dimensionality of the subspace of the total correlation matrix. The incremental solution yields the same solution as the batch-mode, just provided the enough dimensionality of the subspaces.*

setting were initially used for learning the orthogonal subspaces. Then, the sets of the 100 face classes of other illumination settings were additionally given for the update. We set the total number of updates including the initial batch computation to be 6 and the number of images to add at each iteration around 10,000. The dimensionality of the uniformly scaled images was 2,500 and the number of orthogonal components was around 10, which varies for more than 99% of the energy from the eigenvalue plot. See Figure 2 (a) for the example orthogonal component computed by the proposed incremental algorithm and the batch-mode. Figure 2 (b) shows the insensitivity of the incremental OSM to the dimensionality of the subspace of the total correlation matrix. The incremental OSM yields the same accuracy as the batch-mode OSM, just provided the enough dimensionality of the subspace. The subspace dimensionality was automatically chosen from the eigenvalues plots of the correlation matrices at each update. Figure 3 (a) shows the accuracy improvement of the incremental OSM according to the number of updates. It efficiently updates the existing orthogonal subspace models over new evidences contained in the additional data sets, giving gradual accuracy improvements. The computational costs of the batch OSM and the incremental OSM are compared in Figure 3 (b). Whereas the computational cost of the bath-mode is largely increased as the data is repeatedly added, the incremental OSM keeps the cost of the update low.

5.2 Accuracy of Locally OSM

Another experiment was designed for comparing accuracy of several methods with the locally orthogonal subspaces. The training of all the algorithms was performed with the data acquired in a single illumination setting and testing with a single other setting. An independent illumination set with both training and test sets was exploited for the validation. We compared the performance of Mutual Subspace Method (MSM) [4] as a gauging method, where the dimensionality of each subspace is 10 representing more than 99% energy of the data, CMSM [8] used in a state-of-the-art commercial system FacePass [9], where the dimension of the constrained subspace was determined to be



(a) (b)

Figure 3: ***Batch vs. Incremental OSM-2:*** *(a) Accuracy improvement of the incremental OSM for the number of updates. (b) Computational costs of the batch and incremental OSM.*

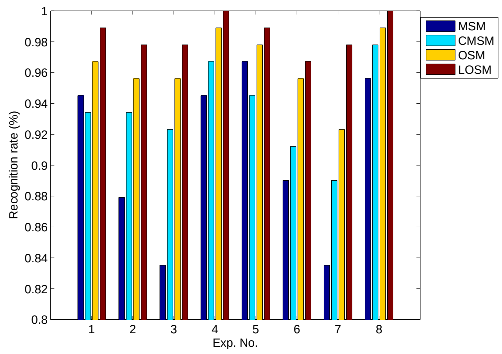


Figure 4: ***Accuracy comparison.***

360, which yielded the best accuracy for the validation set, canonical correlations of Orthogonal Subspace Method (OSM), and canonical correlations of the Locally Orthogonal Subspace Method (LOSM), where the class priori probabilities were set to be binary-valued by a certain threshold. The threshold typically gave a half of the total classes as the neighboring classes. The component numbers of the total correlation matrix and the orthogonal subspaces of OSM and LOSM were 200 and 10 respectively. Figure 4 compares the recognition accuracy of all methods, where the experiment numbers correspond to the combinations of the training/test lighting sets, which were chosen as the most difficult scenarios for MSM. In Figure 4, the OSM was superior to CMSM and the proposed locally orthogonal subspace method (LOSM) outperformed all the other methods. Theoretically, the proposed incremental solution of LOSM provides the same solution of the batch computation of LOSM with slightly more computational costs than that of the incremental OSM.

6 Conclusions

In the object recognition task involving image sets, the development of an efficient incremental learning method for handling increasing volumes of image sets is important.

Image data emanating from environments dramatically changing from time to time is continuously accumulated. The proposed incremental solution of the orthogonal subspaces and the locally orthogonal subspaces facilitates a highly efficient learning to adapt to new data sets. The same solution as the bath-computation is obtained with far lower complexity in both time and space. In the recognition experiments using 700 face image sets, the proposed LOSM delivered the best accuracy over all other relevant methods.

Acknowledgment

The authors would like to thank Ken-ichi Maeda for motivating this study and giving his valuable comments. This study was funded in part by the Toshiba Sponsorship and the Chevening Scholarship.

References

[1] G. Shakhnarovich, J. W. Fisher, and T. Darrel. Face recognition from long-term observations. *ECCV*, pages 851–868, 2002.

[2] O. Arandjelović, G. Shakhnarovich, J. Fisher, R. Cipolla, and T. Darrell. Face recognition with image sets using manifold density divergence. *CVPR*, 2005.

[3] S. Satoh. Comparative Evaluation of Face Sequence Matching for Content-based Video Access. *Int'l Conf. on Automatic Face and Gesture Recognition*, 2000.

[4] O. Yamaguchi, K. Fukui, and K. Maeda. Face recognition using temporal image sequence. *Int'l Conf. on Automatic Face and Gesture Recognition*, pages 318–323, 1998.

[5] L. Wolf and A. Shashua. Learning over sets using kernel principal angles. *Journal of Machine Learning Research*, 4(10):913–931, 2003.

[6] T-K Kim, O. Arandjelović and R. Cipolla, Learning over Sets using Boosted Manifold Principal Angles (BoMPA). *BMVC*, 2005.

[7] T-K Kim, J. Kittler and R. Cipolla, Learning Discriminative Canonical Correlations for Object Recognition with Image Sets. *ECCV*, Graz, 2006.

[8] M. Nishiyama, O. Yamaguchi, and K. Fukui, Face Recognition with the Multiple Constrained Mutual Subspace Method. *In Proc. of Audio Video Based Person Authentication*, 2005.

[9] Toshiba. Facepass. http://www.toshiba.co.jp/rdc/mmlab/tech/w31e.htm.

[10] K. Fukui, B. Stenger and O. Yamaguchi, A Framework for 3D Object Recognition Using the Kernel Constrained Mutual Subspace Method. *In Proc. of ACCV*, 2006.

[11] E.Oja, Subspace Methods of Pattern Recognition. Research Studies Press, 1983.

[12] Å. Björck and G. H. Golub. Numerical methods for computing angles between linear subspaces. *Mathematics of Computation*, 27(123):579–594, 1973.

[13] P. Viola and M. Jones. Robust real-time face detection. *IJCV*, 57(2):137–154, 2004.

[14] P. Hall, D. Marshall and R. Martin. Merging and Splitting Eigenspace Models. *IEEE Trans. on PAMI*, 22(9), 2000.