Dimension-Incremental Subspace Learning for High-Dimensional Data Classification

Tomoya SAKAI†

† Institute of Media and Information Technology, Chiba University
Yayoi 1–33, Inage, Chiba, 263–8522 Japan
E-mail: †tsakai@faculty.chiba-u.jp

Abstract This paper proposes novel methods for learning subspaces using dimension-incremental SVD and random sampling. The most intensive computation in the linear subspace methods is the reduction of dimensionality of the feature space by the eigen decomposition or singular value decomposition. In the present methods, the subspaces are learned by updating their orthonormal basis sets with random increment of the dimension of the feature space. The subspace learning progresses with the similarity measurement of test samples until their classification is completed. This strategy can reduce the computational expense without critical loss of recognition rate especially for the high-dimensional data, and the classification results can be assessed by observing the convergence of the similarity measures. The performance of the present methods was experimentally verified using face recognition datasets.

Key words CLAFIC, SVD, EVD, PCA, feature selection, Monte Carlo, random projection

1. Introduction

I present a substantial improvement in computation of the subspace methods by an incremental approach to the dimension of the feature space. The subspace methods [8], [12] have provided us effective techniques for applications such as optical character recognition, face recognition, and so on. The reduction in the computational costs of the subspace methods contributes to their applicable advances in the technology for large-scale and high-dimensional data.

In the linear subspace methods, the classes of given training samples are basically represented as linear subspaces spanned by the principal components of the samples in the Euclidean feature space. After learning the subspaces, test samples, i.e., queries, are classified into the classes according to (dis)similarity measures between the classes and the queries calculated with the principal components of the learned subspaces. The subspace learning by the principal component analysis is known as the reduction of dimensionality, of which computational expense is quite significant due to high dimensionality of the feature space. In particular, appearance-based vision techniques sometimes have to treat images as intolerably high-dimensional feature vectors with the pixel values in practice.

A possible solution to reduce the computational cost of the subspace learning due to the high dimensionality is incremental learning with respect to the dimension. An iterative algorithm that updates the principal components referring to the feature vectors of the training samples from low to high dimensionality can be computationally cost-effective if it can be terminated at an early iterative stage for the classification of the queries. The subspace learning with the dimensional increment of the feature space can be achieved by application of the incremental singular value decomposition (SVD) [1], [3], [4], [10], [11]. If randomly chosen dimensions are appended to the low-dimensional feature space, the (dis)similarity measures between the learned subspaces and the queries in a low-dimensional feature space are expected to approximate those between them in the high-dimensional feature space due to the same principle as the random projection [2], [5].

In this paper, I first review the traditional linear subspace methods performed in a low-dimensional feature space. Second, I propose classification methods that measure the similarity in low-dimensional feature spaces constructed by dimension increment. I call the subspace methods with random increment of the dimension the *Monte Carlo subspace methods*. Finally, I apply the Monte Carlo subspace methods to the appearance-based face recognition to show the cost effectivity.

2. Linear Subspace Methods in Low-Dimensional Feature Space

Let $\{C_l\}_{l=1}^c$ be a collection of classes, from each of which n_l training samples are given as d-dimensional feature vectors. A data matrix of the class C_l is defined as the matrix with the n_l feature vectors in its columns.

$$\underset{d\times n_l}{\mathbf{X}_l} := \left[\begin{array}{ccc} \mathbf{x}_1 & \dots & \mathbf{x}_{n_l} \end{array}\right], \quad l = 1, \dots, c \tag{1}$$

The linear subspace $S_l := \operatorname{span} \mathbf{X}_l \subseteqq \mathbb{E}^d$ of the class C_l is the image space of the data matrix $\mathbf{X}_l$ in the d-dimensional Euclidean feature space $\mathbb{E}^d$. A fundamental approach to determine the basis of the subspace S_l is the SVD[1)] of the data matrix as $\mathbf{X}_l = \mathbf{U}_l \mathbf{K}_l \mathbf{V}_l^\top$. Here, $\mathbf{K}_l$ is a $k_l \times k_l$ diagonal matrix with nonnegative diagonal elements, i.e., the singular values, arranged in decreasing order. $\mathbf{U}_l$ and $\mathbf{V}_l$ are respectively $d \times k_l$ and $n_l \times k_l$ matrices with orthonormal column vectors spanning the k_l-dimensional subspace S_l and a k_l-dimensional subspace $S_l^* := \operatorname{span} \mathbf{X}_l^\top \subseteqq \mathbb{E}^{n_l}$, satisfying $\mathbf{U}_l^\top \mathbf{U}_l = \mathbf{V}_l^\top \mathbf{V}_l = \mathbf{I}$.

Given the feature vector $\mathbf{q} \in \mathbb{E}^d$ of a query whose class is to be identified, most of the subspace methods geometrically evaluate the (dis)similarity between the subspaces $\{S_l\}_{l=1}^c$ and the query $\mathbf{q}$ in the feature space $\mathbb{E}^d$. The CLAFIC method [12], for example, measures the similarity as the squared l_2-norm of the orthogonal projection of the normalized query $\mathbf{q}/||\mathbf{q}||$ onto a subspace $\operatorname{span} \mathbf{U}$.

$$\mathrm{CLAFIC}(\mathbf{U}, \mathbf{q}) := \left|\left| \mathbf{U}^\top \frac{\mathbf{q}}{||\mathbf{q}||} \right|\right|^2 = \frac{\mathbf{q}^\top \mathbf{U}\mathbf{U}^\top \mathbf{q}}{\mathbf{q}^\top \mathbf{q}} \tag{2}$$

Here, $\mathbf{U}$ is composed of the t_l first column vectors in $\mathbf{U}_l$. One can find antecedent work, such as [9], to fix the truncated dimension t_l of the subspace S_l. The (dis)similarity measurement, however, requires the computation with the high-dimensional vectors in $\mathbb{E}^d$.

I approach this problem by choosing the basis of the feature space itself instead of the bases of the subspaces. A row of the data matrix $\mathbf{X}_l$ corresponds to a coordinate of the feature space. Let $\tilde{\mathbf{X}}_l$ be a low-dimensional data matrix consisting of r row vectors chosen from the row vectors in the data matrix $\mathbf{X}_l$, and let $\tilde{\mathbf{q}} \in \mathbb{E}^r$ be a low-dimensional feature vector of the query with r components chosen from $\mathbf{q}$ in the same manner. Then, the subspace defined as the image space of $\tilde{\mathbf{X}}_l$, i.e., $\tilde{S}_l = \operatorname{span} \tilde{\mathbf{X}}_l \subseteqq \mathbb{E}^r$, is the orthogonal projection of the subspace $S_l = \operatorname{span} \mathbf{X}_l \subseteqq \mathbb{E}^d$ onto the chosen r-dimensional feature space $\mathbb{E}^r$. The basis of $\tilde{S}_l$ is obtained by SVD of $\tilde{\mathbf{X}}_l$ as $\tilde{\mathbf{X}}_l = \tilde{\mathbf{U}}_l \tilde{\mathbf{K}}_l \tilde{\mathbf{V}}_l^\top$. Here, $\tilde{\mathbf{U}}_l$ and $\tilde{\mathbf{V}}_l$ are respectively $r \times \tilde{k}_l$ and $n_l \times \tilde{k}_l$ matrices if the dimension of $\tilde{S}_l$ is $\tilde{k}_l := \dim \tilde{S}_l = \operatorname{rank} \tilde{\mathbf{X}}_l$. The (dis)similarity between $\tilde{S}_l$ and $\tilde{\mathbf{q}}$ in the low-dimensional feature space $\mathbb{E}^r$ can be measured, for example, by the CLAFIC method as $\mathrm{CLAFIC}(\tilde{\mathbf{U}}, \tilde{\mathbf{q}})$. Here, $\tilde{\mathbf{U}}$ is composed of the $\tilde{t}_l$ first column vectors in $\tilde{\mathbf{U}}_l$. If the similarity measured in the low-dimensional feature space approximates that measured in the d-dimensional feature space, the query can be classified without referring to all features of the training samples.

1) : The compact SVD is mainly used to illustrate my method although the eigen decomposition is also available in the same way.

3. Dimension Incremental Subspace Methods

3.1 Framework

Assuming this row-incremental update algorithm, i.e., the row-incremental SVD (RiSVD), I describe a common framework of the dimension-incremental subspace methods in Algorithm 1.

Algorithm 1 Classification by dimension-incremental subspace learning

Input: the row-accessible training data matrices $\{ \underset{d\times n_l}{\mathbf{X}_l} \}_{l=1}^c$, and the query $\mathbf{q} \in \mathbb{E}^d$;

Output: the similarity measures $\{\tilde{g}_l\}_{l=1}^c$, and the learned singular value components $\{ \underset{r\times \tilde{k}_l}{\tilde{\mathbf{U}}_l} \}_{l=1}^c$, $\{ \underset{\tilde{k}_l\times \tilde{k}_l}{\tilde{\mathbf{K}}_l} \}_{l=1}^c$ and $\{ \underset{n_l\times \tilde{k}_l}{\tilde{\mathbf{V}}_l} \}_{l=1}^c$;

1: set $\tilde{\mathbf{q}}$ to be a zero-dimensional vector;
2: **for** all $l = 1$ to c **do**
3:   set $\tilde{\mathbf{U}}_l$, $\tilde{\mathbf{K}}_l$ and $\tilde{\mathbf{V}}_l$ to be 0×0 matrices;
4: **end for**
5: **repeat**
6:   choose the i-th dimension (disallow duplication);
7:   append the i-th component of $\mathbf{q}$ to $\tilde{\mathbf{q}}$;
8:   **for** all $l = 1$ to c **do**
9:     set $\boldsymbol{\xi}^\top$ to be the i-th row of $\mathbf{X}_l$;
10:     update $\tilde{\mathbf{U}}_l$, $\tilde{\mathbf{K}}_l$ and $\tilde{\mathbf{V}}_l$ by *RiSVD* using $\boldsymbol{\xi}$;
11:     measure the similarity $\tilde{g}_l$ of $\tilde{\mathbf{q}}$ using $\tilde{\mathbf{U}}_l$, $\tilde{\mathbf{K}}_l$ and $\tilde{\mathbf{V}}_l$;
12:   **end for**
13: **until** $\arg \max_{l=1,\dots,c} \tilde{g}_l$ is fixed.

We can avoid restoring the low-dimensional data matrix $\tilde{\mathbf{X}}_l$ because its SVD matrices $\tilde{\mathbf{U}}_l$, $\tilde{\mathbf{K}}_l$, and $\tilde{\mathbf{V}}_l$ are directly updated with the chosen row vector $\boldsymbol{\xi}^\top$ from $\mathbf{X}_l$. The low-dimensional data matrix of the class C_l is implicitly stored as $\tilde{\mathbf{X}}_l = \tilde{\mathbf{U}}_l \tilde{\mathbf{K}}_l \tilde{\mathbf{V}}_l^\top$ although the chosen row vector $\boldsymbol{\xi}^\top$ is expired after the RiSVD.

The iteration between Step 5 and Step 13 is terminated when the class with the highest similarity is settled by the similarity measurement such as the CLAFIC in the low-dimensional feature space. In case of multiple queries, the similarity measures between the classes and the queries can be calculated simultaneously in the iteration, which is terminated when the classes of all queries are identified.

The number of rows of $\tilde{\mathbf{U}}_l$, i.e., the dimension r of the low-dimensional feature space, is incremented by one after the RiSVD while the number of its columns, i.e., the dimension $\tilde{k}_l$ of the subspace $\tilde{S}_l \subseteqq \mathbb{E}^r$, is incremented if the degeneration of $\tilde{S}_l$ is relieved by the dimension increment of the feature space. The subspace dimension $\tilde{k}_l$ can be increased up to $\operatorname{rank} \mathbf{X}_l \leqq n_l$ maintaining $\tilde{S}_l = \mathbb{E}^r$, which results in $\forall \tilde{g}_l = 1$ for any queries at the early stages. Therefore, the similarity measurement at Step 11 may be enabled after n_l iterations or after $\tilde{t}_l$ iterations in the case using the CLAFIC method.

3.2 Feature Selection

To perform effectively the subspace method in a low-dimensional feature space, we need a rule for choosing the basis of the feature space, or choosing i-th row from the data matrix $\mathbf{X}_l$ at Step 6 in Algorithm 1. The choice of the dimension is nothing more than the selection of the features. A few types of dimension-incremental subspace methods are derived by different rules of the feature selection.

a) Type-I: Monte Carlo Subspace Method by Equally Random Choice

If we do not have a priori knowledge about which rows store the important features for the similarity measures, random choice may provide us with likely measures. This strategy is well known as the Monte Carlo method. One of the advantages of the random choice is that the reliability of the classification can be tested by repeated trials using random sequences.

b) Type-II: Monte Carlo Subspace Method by Query-Dependent Random Choice

Since large components of the query contributes to the similarity measures, one would expect their faster convergence when the feature is chosen depending on the query. I design such a fast method regarding the magnitude of the query component as relative frequency of the choice. The larger query components are more likely to be chosen by this method.

c) Type-III: Query-Dependent Deterministic Choice

One can also consider the non-random choice of the dimension. For the same reason of the type-II, the dimension is chosen in decreasing order of the magnitude of the query component. Since this method does not take advantage of the random choice, the classification results cannot be assessed by repeated trials.

3.3 Row-Incremental SVD

Algorithm 2 describes the RiSVD for the dimension increment of the feature space. The RiSVD is dual to the column-incremental SVD [1], [3], [4], [10], [11] used for the data increment. Algorithm 2 ensures reconstructivity

$$\begin{bmatrix} \tilde{\mathbf{U}}\tilde{\mathbf{K}}\tilde{\mathbf{V}}^\top \\ \boldsymbol{\xi}^\top \end{bmatrix} = \tilde{\mathbf{U}}_{\text{new}}\tilde{\mathbf{K}}_{\text{new}}\tilde{\mathbf{V}}_{\text{new}}^\top,$$

and inductive orthonormality

$$\tilde{\mathbf{U}}_{\text{new}}^\top\tilde{\mathbf{U}}_{\text{new}} = \tilde{\mathbf{V}}_{\text{new}}^\top\tilde{\mathbf{V}}_{\text{new}} = \mathbf{I} \ \text{ if } \ \tilde{\mathbf{U}}^\top\tilde{\mathbf{U}} = \tilde{\mathbf{V}}^\top\tilde{\mathbf{V}} = \mathbf{I}.$$

3.4 Flop Count

In the usual linear subspace methods, the cost of computing the subspace basis for a class C_l by the SVD of the $d \times n_l$ training data matrix $\mathbf{X}_l$ is $O((d+n_l)\min^2(d,n_l)) \approx O(dn_l^2)$ $(d \gg n_l)$ flops [4], [7], and the similarity measurement, by the CLAFIC for example, costs $O(dt_l)$ a class. On the other hand, if Algorithm 1 requires $r_{\max}$ iterations, the SVD of $\mathbf{B}$ in Algorithm 2 costs $O(r_{\max}\tilde{k}_l^3)$, and the matrix multiplication at Step 12 or 18 and at Step 14 or 20 costs

Algorithm 2 Row-incremental SVD

Input: $\underset{r\times\tilde{k}}{\tilde{\mathbf{U}}}$, $\underset{\tilde{k}\times\tilde{k}}{\tilde{\mathbf{K}}}$, $\underset{n\times\tilde{k}}{\tilde{\mathbf{V}}}$ $(r \geq \tilde{k})$ and $\boldsymbol{\xi} \in \mathbb{E}^n$;

Output: $\tilde{\mathbf{U}}_{\text{new}}$, $\tilde{\mathbf{K}}_{\text{new}}$ and $\tilde{\mathbf{V}}_{\text{new}}$;

1: **if** $r = 0$ **then**

2: $\quad \underset{1\times 1}{\tilde{\mathbf{U}}_{\text{new}}} \leftarrow \begin{bmatrix} 1 \end{bmatrix}$;

3: $\quad \underset{1\times 1}{\tilde{\mathbf{K}}_{\text{new}}} \leftarrow \begin{bmatrix} \|\boldsymbol{\xi}\| \end{bmatrix}$;

4: $\quad \underset{n\times 1}{\tilde{\mathbf{V}}_{\text{new}}} \leftarrow \begin{bmatrix} \dfrac{\boldsymbol{\xi}}{\|\boldsymbol{\xi}\|} \end{bmatrix}$;

5: **end if**

6: $\boldsymbol{\eta} \leftarrow \tilde{\mathbf{V}}^\top\boldsymbol{\xi}$;

7: $\mathbf{p} \leftarrow \boldsymbol{\xi} - \tilde{\mathbf{V}}\boldsymbol{\eta}$;

8: $p \leftarrow \|\mathbf{p}\|$;

9: **if** $p \neq 0$ **then**

10: $\quad \underset{(\tilde{k}+1)\times(\tilde{k}+1)}{\mathbf{B}} \leftarrow \begin{bmatrix} \tilde{\mathbf{K}} & \mathbf{0} \\ \boldsymbol{\eta}^\top & p \end{bmatrix}$;

11: $\quad$ do singular value decomposition of $\mathbf{B}$ to obtain $\underset{(\tilde{k}+1)\times(\tilde{k}+1)}{\mathbf{U}_{\text{B}}}$, $\underset{(\tilde{k}+1)\times(\tilde{k}+1)}{\mathbf{K}_{\text{B}}}$ and $\underset{(\tilde{k}+1)\times(\tilde{k}+1)}{\mathbf{V}_{\text{B}}}$ such that $\mathbf{U}_{\text{B}}\mathbf{K}_{\text{B}}\mathbf{V}_{\text{B}}^\top = \mathbf{B}$ and $\mathbf{U}_{\text{B}}^\top\mathbf{U}_{\text{B}} = \mathbf{V}_{\text{B}}^\top\mathbf{V}_{\text{B}} = \underset{(\tilde{k}+1)\times(\tilde{k}+1)}{\mathbf{I}}$;

12: $\quad \underset{(r+1)\times(\tilde{k}+1)}{\tilde{\mathbf{U}}_{\text{new}}} \leftarrow \begin{bmatrix} \tilde{\mathbf{U}} & \mathbf{0} \\ \mathbf{0}^\top & 1 \end{bmatrix} \mathbf{U}_{\text{B}}$;

13: $\quad \underset{(\tilde{k}+1)\times(\tilde{k}+1)}{\tilde{\mathbf{K}}_{\text{new}}} \leftarrow \mathbf{K}_{\text{B}}$;

14: $\quad \underset{n\times(\tilde{k}+1)}{\tilde{\mathbf{V}}_{\text{new}}} \leftarrow \begin{bmatrix} \tilde{\mathbf{V}} & \dfrac{\mathbf{p}}{p} \end{bmatrix} \mathbf{V}_{\text{B}}$;

15: **else**

16: $\quad \underset{(\tilde{k}+1)\times\tilde{k}}{\mathbf{B}} \leftarrow \begin{bmatrix} \tilde{\mathbf{K}} \\ \boldsymbol{\eta}^\top \end{bmatrix}$;

17: $\quad$ do singular value decomposition of $\mathbf{B}$ to obtain $\underset{(\tilde{k}+1)\times\tilde{k}}{\mathbf{U}_{\text{B}}}$, $\underset{\tilde{k}\times\tilde{k}}{\mathbf{K}_{\text{B}}}$ and $\underset{\tilde{k}\times\tilde{k}}{\mathbf{V}_{\text{B}}}$ such that $\mathbf{U}_{\text{B}}\mathbf{K}_{\text{B}}\mathbf{V}_{\text{B}}^\top = \mathbf{B}$ and $\mathbf{U}_{\text{B}}^\top\mathbf{U}_{\text{B}} = \mathbf{V}_{\text{B}}^\top\mathbf{V}_{\text{B}} = \underset{\tilde{k}\times\tilde{k}}{\mathbf{I}}$;

18: $\quad \underset{(r+1)\times\tilde{k}}{\tilde{\mathbf{U}}_{\text{new}}} \leftarrow \begin{bmatrix} \tilde{\mathbf{U}} & \mathbf{0} \\ \mathbf{0}^\top & 1 \end{bmatrix} \mathbf{U}_{\text{B}}$;

19: $\quad \underset{\tilde{k}\times\tilde{k}}{\tilde{\mathbf{K}}_{\text{new}}} \leftarrow \mathbf{K}_{\text{B}}$;

20: $\quad \underset{n\times\tilde{k}}{\tilde{\mathbf{V}}_{\text{new}}} \leftarrow \underset{n\times\tilde{k}}{\tilde{\mathbf{V}}}\ \underset{\tilde{k}\times\tilde{k}}{\mathbf{V}_{\text{B}}}$;

21: **end if**

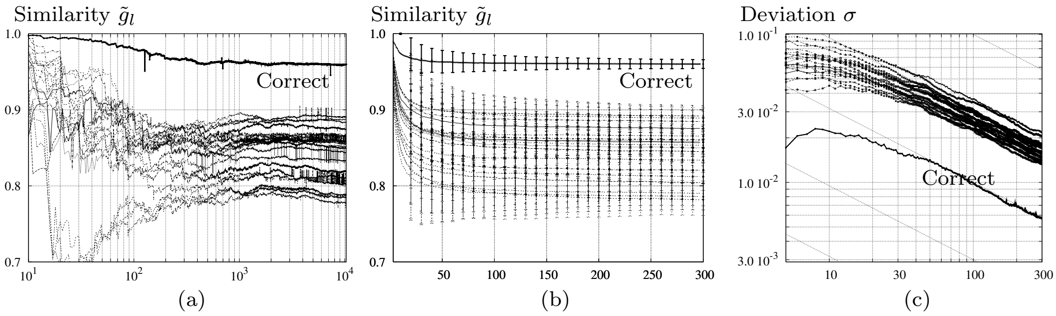

Fig. 1 Similarity measures vs dimension of feature space by the type-I Monte Carlo subspace method. In all three graphs, the horizontal axis is the iteration count r, or the reduced dimension of the feature space. (a) An example of the progress of the similarity measures. (b) Average progress of the similarity measures. The error bar indicates the standard deviation σ. (c) Evolution of σ with respect to r shows the $r^{-\frac{1}{2}}$-asymptotics (the oblique dashed lines).

$O(r_{\max}^2 \tilde{k}_l^2 + r_{\max} n_l \tilde{k}_l^2)$. The similarity measure can be done in $O(r_{\max}^2 \tilde{t}_l)$ flops. Since the total cost is approximately $O(r_{\max}^2)$, the dimension-incremental subspace methods can reduce the computation time if it completes the classification in $r_{\max} \sim O(d^{\frac{1}{2}})$ iterations. Besides, Algorithm 1 can save memory space by out-of-core computation because the input data matrices may be row-accessible.

4. Experiments

The cost-effectiveness for high-dimensional data classification is demonstrated by the appearance-based image recognition using a pre-cropped version of the UMIST face database [6]. A partial set of face images consisting of $n_l = 8$ images each of 20 individuals is used as the training dataset of a class, and the remaining images are used as the queries. The dimension of the feature space is $d = 112 \times 92 = 10304$.

Figure 1(a) shows an example of the progress of the similarity measures $\tilde{g}_l = \mathrm{CLAFIC}(\tilde{\mathbf{U}}_l, \mathbf{q})$ between the 20 classes and a query by the type-I method. The significance of the similarity between the correct class and the query becomes apparent as the dimension of the low-dimensional feature space grows. Figure 1(b) shows the average progress with over 500 trials. Remarkably, only a few dozen times of dimension increment are sufficient to clarify the correct class for the query. The computation time is reduced to 18% of the usual CLAFIC method in case of $r_{\max} = 50$ and to 56% in case of $r_{\max} = 100$. Although calculating the precise similarity requires numerous iterations because the similarity measures slowly converge with order 1/2 as shown in Fig. 1(c), the class with the highest similarity can be determined by the measurement in the low-dimensional feature spaces.

The appearance-based image recognition without any normalisation can be highly dependent on the positions of target objects in the images. It is well-known that the Fourier amplitude, or the power spectrum, is invariant under any spatial shift of the images. I have tested the dimension-incremental subspace methods for the Fourier transform of the UMIST images. As shown in Fig. 2(a), the type-I method could find a correct class in average. However, the similarity measures have large deviations, indicating low reliablity of the classification results at low dimensions. Since every similarity $\tilde{g}_l$ is greater than about 0.92, the differences between the similarity measures are relatively small. Nevertheless, the type-II method could identify the correct class at a few tens of dimensions as shown in Fig. 2(b). This implies that the query-dependent random choice is effective for improving the precision of the similarity measures at low dimensions. Among the present methods, the type-III method shows the fastest convergence as shown in Fig. 2(c).

5. Concluding Remarks

The dimension-incremental approach to the subspace learning allows us to measure the similarity between the classes and queries in low-dimensional feature spaces. The present methods achieve considerable reduction in computation time of the classification, which makes tractable the pattern recognition for large dimensional data. Another distinctive feature of the present methods is that we can observe the progress of the similarity measures with respect to the dimension, and evaluate the reliability of the classification results. The further research on the Monte Carlo scheme for the dimensionality reduction should be pursued from the viewpoint of the random projection [2], [5].

References

[1] M. Brand. Fast online svd revisions for lightweight recommender systems. In *Proc. SDM 2003*, 2003.

[2] E. Brigham and H. Maninila. Random projection in dimensionality reduction: applications to image and text data. In *ACM SIGKDD ICKDDM*, pages 245–250, 2001.

[3] J. R. Bunch and C. P. Nielsen. Updating the singular value decomposition. *Numer. Math.*, 31:111–129, 1978.

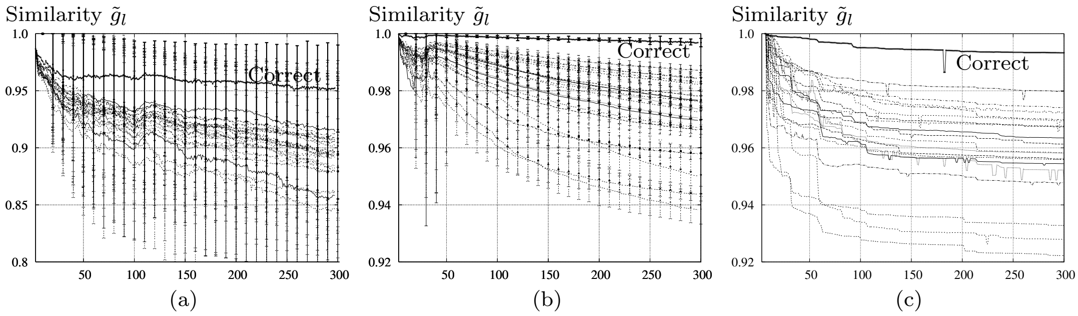

Fig. 2 Same as Fig. 1(b), but for the Fourier transformed data evaluated by (a) the type-I method, (b) the type-II method and (c) the type-III method.

[4] S. Chandrasekaran, B. S. Manjunath, Y. F. Wang, J. Winkeler, and H. Zhang. An eigenspace update algorithm for image analysis. *Graphical Models and Image Processing*, 59(5):321–332, 1997.

[5] D. Fradkin and D. Madigan. Experiments with random projections for machine learning. In *ACM SIGKDD ICKDDM*, pages 517–522, 2003.

[6] D. B. Graham and N. M. Allinson. Characterizing virtual eigensignatures for general purpose face recognition. In *Face Recognition: From Theory to Applications*, volume 163, pages 446–456. NATO ASI Series F, Computer and Systems Sciences, 1998.

[7] M. Gu and S. C. Eisenstat. A stable and fast algorithm for updating the singular value decomposition. In *Tech. Rep. YALEU/DCS/RR-966*. Yale University, 1994.

[8] T. Iijima, H. Genchi, and K. Mori. A theory of character recognition by pattern matching method. In *Proc. 1st IJCPR*, pages 50–56, October 1973.

[9] J. Laaksonen and E. Oja. Subspace dimension selection and averaged learning subspace method in handwritten digit classification. In *Proc. ICANN 1996*, pages 227–232, 1996.

[10] H. Murase and M. Lindenbaum. Partial eigenvalue decomposition of large images using spatial temporal adaptive method. *IEEE Trans. IP*, 4(5):620–629, May 1995.

[11] D. Skocaj and A. Leonardis. Incremental and robust learning of subspace representations. *Image and Vision Computing*, 26:27–38, 2008.

[12] S. Watanabe, P. F. Lambert, C. A. Kulikowski, J. L. Buxton, and R. Walker. Evaluation and selection of variables in pattern recognition. In *Computer and Information Sciences II*. Academic Press, New York, 1967.