Topology-Preserving Dimension-Reduction Methods for Image Pattern Recognition

Hayato Itoh[1], Tomoya Sakai[2], Kazuhiko Kawamoto[3], and Atsushi Imiya[4]

[1] School of Advanced Integration Science, Chiba University,
1-33 Yayoicho, Inage-ku, Chiba, 263-8522, Japan
[2] Graduate School of Engineering, Nagasaki University,
1-14 Bunkyo-cho, Nagasaki, 852-8521, Japan
[3] Academic Link Center, Chiba University,
1-33 Yayoicho, Inage-ku, Chiba, 263-8522, Japan
[4] Institute of Media and Information Technology, Chiba University,
1-33 Yayoicho, Inage-ku, Chiba, 263-8522, Japan

Abstract. In this paper, we experimentally evaluate the validity of dimension-reduction methods which preserve topology for image pattern recognition. Image pattern recognition uses pattern recognition techniques for the classification of image data. For the numerical achievement of image pattern recognition techniques, images are sampled using an array of pixels. This sampling procedure derives vectors in a higher-dimensional metric space from image patterns. For the accurate achievement of pattern recognition techniques, the dimension reduction of data vectors is an essential methodology, since the time and space complexities of data processing depend on the dimension of data. However, the dimension reduction causes information loss of geometrical and topological features of image patterns. The desired dimension-reduction method selects an appropriate low-dimensional subspace that preserves the topological information of the classification space.

1 Introduction

Pattern recognition techniques are applied to image patterns. In image pattern recognition, images are sampled so that they can be embedded in a vector space, and dimension-reduction is operated to reduce the dimensions of image patterns [1–3].

In practice, as shown in Fig. 1, there are two paths for the dimension reduction. One method reduces the dimension of data in a sampled image space using image compression methods such as the pyramid transform, wavelet transform and low-pass filtering. The other method is data compression in a vector space after vectorisation of sampled image patterns using operations such as random projection. The reduction and vectorisation operations are generally noncommutative as shown in Fig. 1.

In this paper, we evaluate the differences in the effects and performance among these three data compression techniques. We adopted the following dimension

J.-K. Kämäräinen and M. Koskela (Eds.): SCIA 2013, LNCS 7944, pp. 195–204, 2013.

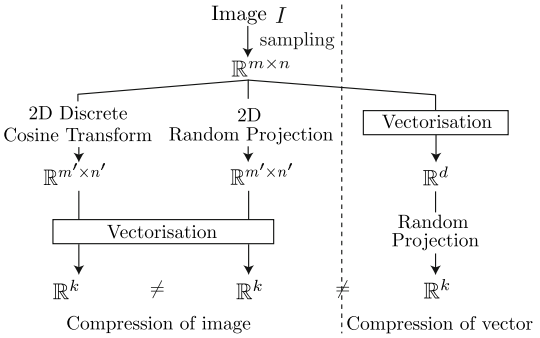


Fig. 1. Differences in the dimension-reduction path among the two-dimensional discrete transformation, two-dimensional random projection and random projection. After sampling of an original image, dimension-reduction methods mainly separated follow two paths. In the first path on the left of the dashed line, after the reduction of the image, the reduced image is converted to a vector. In the second path on the right of the dashed line, after vectorisation, the feature vector is reduced. Here, $m, m', n, n', d, k \in \mathbb{Z}$ and $n' < n, m' < m, k < d$.

reduction techniques: random projection, two-dimensional random projection and the two-dimensional discrete cosine transform. For classification, we adopted the subspace method and two-dimensional tensorial subspace method. We tested each pair of these dimension-reduction techniques and classifiers for face recognition, spacial object recognition, and character recognition.

2 Mathematical Preliminary

Setting H to be the space of patterns, we assume that in H the inner product (f, g) is defined. Furthermore, we define the Schatten product $\langle f, g\rangle$, which is an operator from H to H. Let $f \in H$ and $P_i,\ i = 1, \ldots, N$ be a pattern and an operator for the ith class, where the ith class is defined as

$$\mathcal{C}_i = \{f \,|\, P_i f = f,\ P_i^* P_i = I\}. \tag{1}$$

Since patterns have perturbations, we define the ith class as

$$\mathcal{C}_i(\delta) = \{f \,|\, \|P_i f - f\|_2 \ll \delta,\ P_i^* P_i = I\} \tag{2}$$

where δ is a small perturbation of the pattern. For input $g \in H$ and class $\mathcal{C}_i$, we define the similarity and classification criteria as

$$\theta_i = \angle(C_i(\delta), g),\ 0 < \theta_i < {}^{\exists}\theta_0 \rightarrow g \in \mathcal{C}_i(\delta), \tag{3}$$

since we define the angle between an input pattern g and the space of the pattern as

$$\theta_i = \cos^{-1} \frac{\|P_i g\|_2}{\|g\|_2}. \tag{4}$$

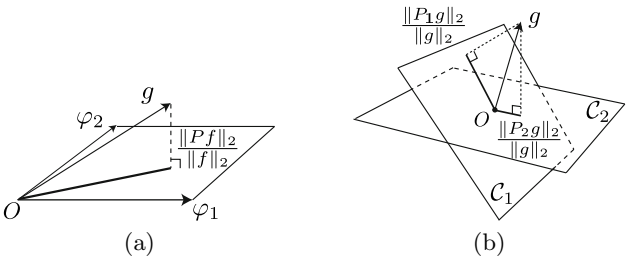


Fig. 2. (a) Geometric property of the subspace method. Let φ_1 and φ_2 be the bases of a class pattern. For input g, similarity is defined as the orthogonal projection to the pattern space. (b) Multiclass recognition using the subspace method. Let P_1 and P_2 be operators for subspaces $\mathcal{C}_1$ and $\mathcal{C}_2$, respectively. Input g is labeled as beign in the 1st class, since the subspace $\mathcal{C}_1$ has the longest projection length of g.

The angle between the input pattern and the pattern space represents their similarity [4–6].

Figure 2(a) shows the basic idea of the subspace method. To identify whether the input data are in the subspace of the class or not, we calculate the angle between the input data and the subspace of the classes. If g belongs to the space, the length of the orthogonal projection is close to 1. Figure 2(b) shows multiclass recognition using the subspace method. For multiclass recognition, we construct a operator P_i for $f_i \in \mathcal{C}_i$ such that

$$E\left(\|f - P_i f\|_2\right) \to \min, \quad P_i^* P_i = I, \tag{5}$$

where $f \in \mathcal{C}_i$, I is the identity operator and E is the expectation on H.

3 Construction of Projections

3.1 Karhunen-Loeve Expansion

For practical calculation of P and P_i in Eqs. (3) and (5), we adopt the Karhunen-Loeve expansion. The Karhunen-Loeve expansion approximates the subspace of data in a Hilbert space. We set $\{\varphi_j\}_{j=1}^n$ to be the eigenfunction of $M = E\langle f, f\rangle$. We define the eigenfunction of M as $\|\varphi_j\|_2 = 1$ for eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_j \geq \cdots \geq \lambda_n$. Therefore, operator P is defined as $P_k = \sum_{j=1}^k \langle \varphi_j, \varphi_j \rangle$ for $k \leq n$.

3.2 Two-Dimensional Discrete Cosine Transform

For a real image $\boldsymbol{X} \in \mathbb{R}^{m \times n}$, the discrete Fourier transformation can be replaced with the discrete cosine transform (DCT), since the image $\boldsymbol{X}$ is a real matrix. Selecting low-frequency bases, we can compress an image as

$$\boldsymbol{Y} = \boldsymbol{S}_L \boldsymbol{\Phi}^{-1} \boldsymbol{P}_\Omega \boldsymbol{\Phi} \boldsymbol{X} \boldsymbol{\Phi} \boldsymbol{P}_\Omega \boldsymbol{\Phi}^{-1} \boldsymbol{S}_R^\top. \tag{6}$$

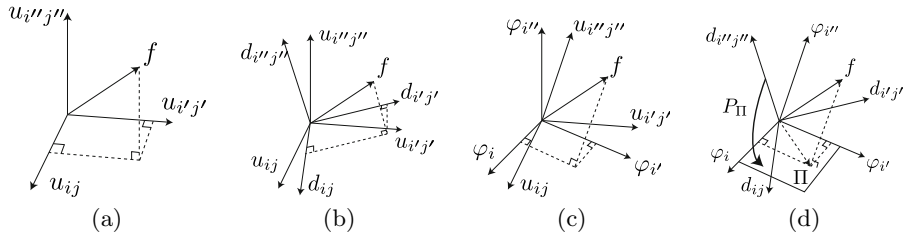


Fig. 3. Image representations in three coordinates. (a) $u_{ij}, u_{i'j'}, u_{i''j''}$ are the bases which represent each pixel of an image. (b) $d_{ij}, d_{i'j'}, d_{i''j''}$ are the bases of the DCT. (c) $\varphi_i, \varphi_{i'}, \varphi_{i''}$ are the bases of the PCA. (d) There is a projection P_Π which projects the image f to the linear subspace $\Pi = \{\varphi_i, \varphi_{i'}\}$ from the space spanned by the DCT.

Here, $\boldsymbol{Y} \in \mathbb{R}^{k_1 \times k_2}$, $k_1 \leq m$, $k_2 \leq n$ $\boldsymbol{\Phi} \in \mathrm{R}^{m \times n}$ is a DCT matrix, $\boldsymbol{P}_\Omega$ is a band limitation matrix, and $\boldsymbol{S}_L \in \mathbb{R}^{k_1 \times m}$ and $\boldsymbol{S}_R \in \mathbb{R}^{k_2 \times n}$ are the sampling matrices. The eigenfunction and eigendistribution of the DCT approximately coincide with those of the Karhunen-Loeve expansion for images. For a discrete image, the Karhunen-Loeve expansion is computed by the principal component analysis (PCA). Figure 3 illustrates the representation of an image by the DCT and PCA and a special case. Since the DCT and PCA are unitary transforms, these bases are related to a rotation transformation.

3.3 Two-Dimensional Tensorial Subspace Method

As an extension of the subspace method for vector data, we introduce a linear subspace method for a bilinear array. For a bilinear array $\boldsymbol{X} \in \mathbb{R}^{m \times n}$, setting $\boldsymbol{P}_L \in \mathbb{R}^{k_1 \times m}$ $(k_1 \leq m)$ and $\boldsymbol{P}_R \in \mathbb{R}^{k_2 \times n}$ $(k_2 \leq n)$ to be orthogonal projections, we call the operation

$$\boldsymbol{Y} = \boldsymbol{P}_L \boldsymbol{X} \boldsymbol{P}_R \tag{7}$$

the orthogonal projection of $\boldsymbol{X}$ to $\boldsymbol{Y} \in \mathbb{R}^{k_1 \times k_2}$. Therefore, using this expression for a collection of bilinear forms $\{\boldsymbol{X}\}_{i=1}^n$, such that $E_i(\boldsymbol{X}_i) = 0$, the solutions of

$$J(\boldsymbol{P}_L, \boldsymbol{P}_R) = E_i \left(\frac{\|\boldsymbol{P}_L \boldsymbol{X}_i \boldsymbol{P}_R\|_2}{\|\boldsymbol{X}_i\|_2} \right) \rightarrow \max, \ w.r.t. \ \boldsymbol{P}_L^* \boldsymbol{P}_L = \boldsymbol{I}, \ \boldsymbol{P}_R^* \boldsymbol{P}_R = \boldsymbol{I} \tag{8}$$

define a bilinear subspace which approximates $\{\boldsymbol{X}\}_{i=1}^n$. Using the solutions of Eq. (8), if an input data array $\boldsymbol{G}$ satisfies the condition

$$\arg \left(\max_i \frac{\|\boldsymbol{P}_{Li} \boldsymbol{G} \boldsymbol{P}_{Ri}\|_2}{\|\boldsymbol{G}\|_2} \right) = \{\boldsymbol{P}_{Lk}, \boldsymbol{P}_{Rk}\}, \tag{9}$$

we conclude that $\boldsymbol{G} \in \mathcal{C}_k(\delta)$ when $\mathcal{C}_k = \{\boldsymbol{X} \mid \|P_{Lk} \boldsymbol{X} P_{Rk} - \boldsymbol{X}\|_2 \ll \delta\}$.

3.4 Marginal Eigenvalue

In practical computation to find the projection P_L and P_R in Eq. (8), we adopt the marginal eigenvalue (MEV)[7]. This is a projection considering the distribution of column and row vectors of sampled images. For two matrices

$$\boldsymbol{M}_{\mathrm{r}} = \frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{P}_L\boldsymbol{X}_i\boldsymbol{P}_R\right)\left(\boldsymbol{P}_L\boldsymbol{X}_i\boldsymbol{P}_R\right)^{\top} = \boldsymbol{P}_L\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{X}_i^{\top}\right)\boldsymbol{P}_L^{\top}, \tag{10}$$

$$\boldsymbol{M}_{\mathrm{c}} = \frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{P}_L\boldsymbol{X}_i\boldsymbol{P}_R\right)^{\top}\left(\boldsymbol{P}_L\boldsymbol{X}_i\boldsymbol{P}_R\right) = \boldsymbol{P}_R^{\top}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i^{\top}\boldsymbol{X}_i\right)\boldsymbol{P}_R, \tag{11}$$

using the Lagrange multipliers Λ_r and Λ_c, we find the projections satisfying

$$J(\boldsymbol{P}_L) = tr\left(\boldsymbol{M}_r\right) - tr\left(\left(\boldsymbol{P}_L\boldsymbol{P}_L^{\top} - \boldsymbol{I}\right)\Lambda_r\right), \tag{12}$$

$$J(\boldsymbol{P}_R) = tr\left(\boldsymbol{M}_c\right) - tr\left(\left(\boldsymbol{P}_R^{\top}\boldsymbol{P}_R - \boldsymbol{I}\right)\Lambda_c\right) \tag{13}$$

where $\boldsymbol{I}$ is the identity matrix. The solutions of Eqs. (12) and (13) are given as the solutions of the eigenproblems of $\boldsymbol{M}_r$ and $\boldsymbol{M}_c$, respectively. We set $\{\varphi_j^r\}_{j=1}^{k_1}$ and $\{\varphi_j^c\}_{j=1}^{k_2}$ be the eigenfunctions of $\boldsymbol{M}^r$ and $\boldsymbol{M}^c$, respectively.. We define the eigenfunctions of $\boldsymbol{M}_r$ and $\boldsymbol{M}_c$ as $\|\varphi_j^r\| = 1$ and $\|\varphi_j^c\| = 1$ for eigenvalues $\lambda_1^r \geq \lambda_2^r \geq \cdots \geq \lambda_j^r \geq \cdots \geq \lambda_n^r$ and $\lambda_1^c \geq \lambda_2^c \geq \cdots \geq \lambda_j^c \geq \cdots \geq \lambda_n^c$, respectively. Therefore, operators $\boldsymbol{P}_L^{\top}$ and $\boldsymbol{P}_R$ are defined as $\boldsymbol{P}_{L,k_1} = \sum_{j=1}^{k_1}\varphi_j^r{\varphi_j^r}^{\top}$ and $\boldsymbol{P}_{R,k_2} = \sum_{j=1}^{k_2}\varphi_j^c{\varphi_j^c}^{\top}$, respectively..

3.5 Random Projection

Let $\boldsymbol{R}$ be a $k \times d$ matrix whose k row vectors span a k-dimensional linear subspace of $\mathbb{R}^k$ $(k < d)$. We obtain a low-dimensional representation $\hat{\boldsymbol{x}}$ for each $\boldsymbol{x}_i \in X$ as

$$\hat{\boldsymbol{x}}_i = \sqrt{\frac{d}{k}}\boldsymbol{R}\boldsymbol{x}_i. \tag{14}$$

Figure 4(a) shows the basic idea of the random projection[8]. For the random projection, we have the following embedding property from the Johnson-Lindenstrauss lemma[9, 10].

Theorem 1 *(Johnson-Lindenstrauss embeddings). For any* $0 < \epsilon \ll 1$*, set* X *of* N *points* $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ *and* $k < d$*, one can map* X *to* $\hat{X} = \{\hat{\boldsymbol{x}}_1, \ldots, \hat{\boldsymbol{x}}_N \in \mathbb{R}^{\hat{d}}\}$ *by the random projection in Eq. (14) with probability* $(1 - \mathrm{e}^{-\mathcal{O}(k\epsilon^2)})$ *when*

$$(1-\epsilon)\|\boldsymbol{x}_j - \boldsymbol{x}_i\|_2 \leq \|\hat{\boldsymbol{x}}_j - \hat{\boldsymbol{x}}_i\|_2 \leq (1+\epsilon)\|\boldsymbol{x}_j - \boldsymbol{x}_i\|_2, \tag{15}$$

that is,

$$P\left(\left|1 - \frac{\|\hat{\boldsymbol{x}}_i - \hat{\boldsymbol{x}}_j\|_2}{\|\boldsymbol{x}_j - \boldsymbol{x}_i\|_2}\right| < \epsilon\right) = 1 - \mathrm{e}^{-\mathcal{O}(k\epsilon^2)}. \tag{16}$$

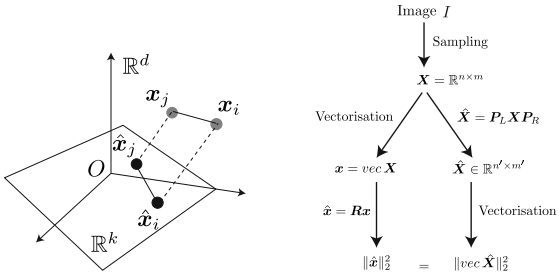


Fig. 4. (a) Random projection. Let $\boldsymbol{x}_i \in X$ be a point and $\hat{\boldsymbol{x}}_i = \boldsymbol{R}\boldsymbol{x}_i$. The distance between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ is preserved in the projected space $\mathbb{R}^k$. (b) Differences in two random projection paths.

The random projection preserves the local topological structure of the vectors of an image pattern.

An efficient random projection is proposed as an improved version of the random projection[10]. Using spectrum spreading and circular convolution, we can speed up the random projection with $\mathcal{O}(d \log d)$ computational time and $\mathcal{O}(d)$ memory storage requirement.

3.6 Two-Dimensional Tensorial Random Projection

For a set of two-dimensional arrays $\{\boldsymbol{X}_i | \boldsymbol{X}_i \in \mathbb{R}^{m \times n}\}_{i=1}^N$ such that $E_i(\boldsymbol{X}_i) = 0$, setting $\boldsymbol{R}_L \in \mathbb{R}^{k_1 \times m}$ and $\boldsymbol{R}_R \in \mathbb{R}^{k_2 \times n}$ to be random projection matrices, we define the transform

$$\hat{\boldsymbol{X}}_i = \boldsymbol{R}_L \boldsymbol{X}_i \boldsymbol{R}_R^\top. \tag{17}$$

For the set $\hat{X} = \{\hat{\boldsymbol{X}}_i\}_{i=1}^N$, we have the following theorem.

Theorem 2 *$\hat{\boldsymbol{X}}_i \in \hat{X}$ and $\boldsymbol{X}_i \in X$ satisfy the Johnson-Lindenstrauss property.*

(*Proof*) Since $\hat{\boldsymbol{X}}_i = \boldsymbol{R}_L \boldsymbol{X}_i \boldsymbol{R}_R$, we have the relation

$$vec\hat{\boldsymbol{X}}_i = (\boldsymbol{R}_L \otimes \boldsymbol{R}_R) vec\boldsymbol{X}_i, \tag{18}$$

where $\boldsymbol{R}_L \otimes \boldsymbol{R}_R = \boldsymbol{R} \in \mathbb{R}^{k \times d}$ is a random projection matrix. Here, $k = k_1 \times k_2$ and $d = m \times n$. Therefore, for any $0 < \epsilon \ll 1$ and set X of N images $\{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N\}$, $\hat{\boldsymbol{X}}_i$ and $\hat{\boldsymbol{X}}_j$ satisfy the property

$$(1 - \epsilon)\|\boldsymbol{X}_j - \boldsymbol{X}_i\|_2 \leq \|\hat{\boldsymbol{X}}_j - \hat{\boldsymbol{X}}_i\|_2 \leq (1 + \epsilon)\|\boldsymbol{X}_j - \boldsymbol{X}_i\|_2. \tag{19}$$

Here, setting $\|\boldsymbol{A}\|_2$ to be the Frobenius norm of a matrix $\boldsymbol{A}$, the relation

$$\|\boldsymbol{x}_i\|^2 = \|vec\boldsymbol{X}_i\|^2 = \|\boldsymbol{X}_i\|_2^2 \tag{20}$$

is satisfied for $\boldsymbol{x}_i = vec\boldsymbol{X}_i$. Therefore, by replacing the Euclidean norm of $vec\boldsymbol{X}_i$ with the Frobenius norm of $\boldsymbol{X}_i$, we have the statement of the theorem. Q.E.D.

Table 1. Details of each database

	# class	# data/class	image size [pixel]	vectorised size
Yale B	38	64	192×168	32,256
ETH80	30	41	128×128	16,384
MNIST	10	7,000	28×28	784

(a) (b) (c)

Fig. 5. Examples of data. (a) Yale B. (b) ETH80. (c) MNIST.

Considering the two-dimensional array as a second-order tensor, we can reduce the dimension of the tensorial data to an arbitrary dimension. The random projection preserves the topology of the tensor in the function space, since the Frobenius norm of a tensor is approximately preserved.

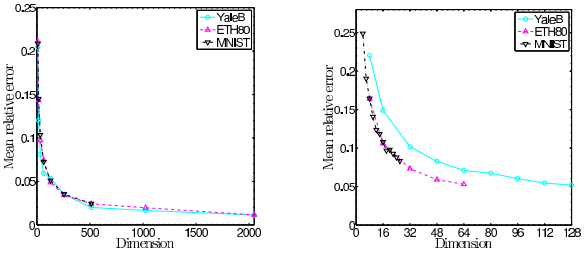
4 Experiments

For the performance evaluation of the two-dimensional discrete cosine transform (2DDCT), the random projection (RP) and the two-dimensional random projection (2DRP), we calculate the relative errors between original images and compressed images projected to low-dimensional space. For the computation of the relative errors by the three dimension-reduction methods, we adopt cropped versions of the extended Yale B database[11], the ETH80 database[12] and the MNIST dataset[13]. Table 1 lists the details of the three databases. Figure 5 shows examples of images in each database. We calculate the mean relative error of 1000 distances between images for each dimension-reduction method. In each process, two different images are randomly chosen. Figure 6 lists the mean relative errors of the RP and 2DRP. Figure 7 gives a comparison of the mean relative errors of 2DDCT for the three databases. In Figs. 6(a) and 6(b), the dimensions represent the sizes of the reduced vectors, and the width and height, respectively.

Next, using the same databases, we calculate the recognition rates of the three dimension-reduction methods. We adopt the subspace method (SM) and two-dimensional tensorial subspace method (TDTSM) as the classification methods.

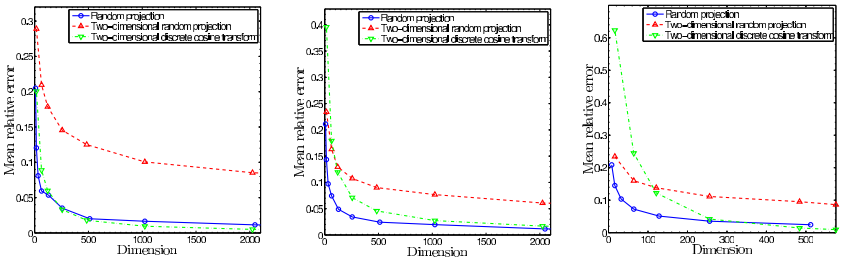
Table 2. Dimensions of the class subspace in the subspace method

	# query	# basis	dimension of reduced vector
Yale B	1	1-32	1024
ETH80	1	1-20	1024
MNIST	1	1-225	225



(a) RP. (b) 2DRP.

Fig. 6. Relateve error of distances among images in three databases



(a) Yale B. (b) ETH80. (c) MNIST.

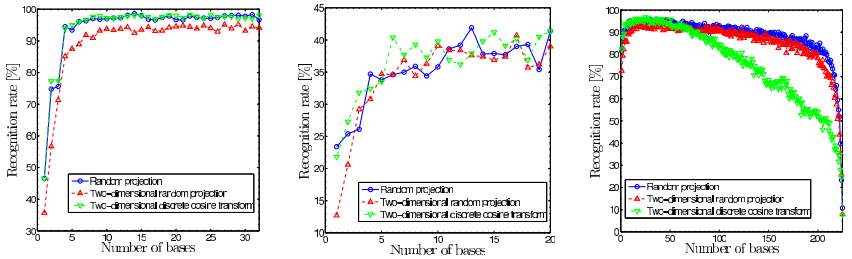
Fig. 7. Comparison of relative errors of distances among images

For the Yale B and ETH80 database, the images labeled with even numbers and odd numbers are used for training data and test data, respectively. The MNIST database is divided into training and test data in advance. The recognition rates are the successful label-estimation ratios of 1000 iterations of the estimation processes, and queries are randomly chosen from the test data. Tables 2 and 3 summarise the dimensions of the class subspace in the recognition processes using the SM and TDTSM, respectively. Figures 8 and 9 illustrate the recognition rates of the SM and TDTSM, respectively. In Figs. 8 and 9, the dimensions represent the sizes of the reduced vectors, and the width and height, respectively.

Figure 6(a) illustrates that the relative errors of distances of the RP are almost the same for the three datasets. These results imply that the results of the RP do not depend on the properties of the database. Figure 6(b) illustrates that the

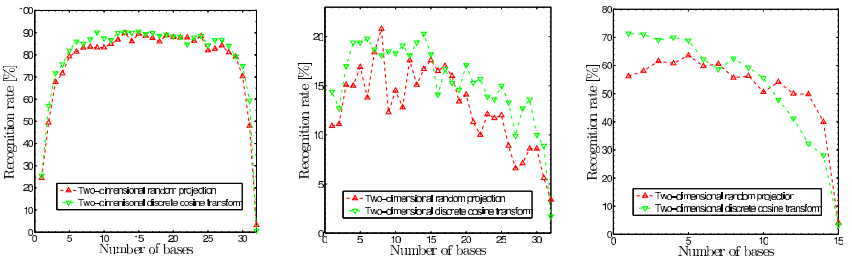
Table 3. Dimensions of the class subspace in the two-dimensional tensorial subspace method

	# query	# basis	dimension of reduced image
Yale B	1	1×1 - 32×32	32 × 32
ETH80	1	1×1 - 32×32	32 × 32
MNIST	1	1×1 - 15×15	15 × 15



(a) Yale B. (b) ETH80. (c) MNIST.

Fig. 8. Recognition rates of the subspace method



(a) Yale B. (b) ETH80. (c) MNIST.

Fig. 9. The recognition rate of the two-dimensional tensorial subspace method

relative errors of the 2DRP are higher than those of the RP, since the RP can efficiently compress the data when the dimension of the data is extremely high.

The results in Fig. 7 show that for every dataset, the relative errors of distances between feature vectors of the RP are smaller than those of the 2DDCT at low dimensions. For the PR and 2DRP, the same property for the relative errors between vectors can be observed.

The results in Fig. 8 show that the recognition rates of the 2DDCT and RP are almost the same in every database. Furthermore, the recognition rates of the 2DRP are the lowest in every dataset. Figure 9 shows that the recognition rates of the 2DDCT and 2DRP are almost the same. However, if the target dimension is small, the 2DDCT has a higher recognition rate than the 2DRP. This property originates from the fact that the energy of images is concentrated in the low-frequency cosine bases.

Figures 8 and 9 show that the recognition rates of the vector-based classification are higher than those of the image-based classification. These results lead to the conclusion that the relative errors and recognition rates of the RP are the lowest and highest, respectively, among the three dimension-reduction methods.

5 Conclusions

We experimentally evaluated the validity of three dimension-reduction methods for image pattern recognition. The desired dimension-reduction method selects

an appropriate low-dimensional subspace that preserves the topology among images.

By experimental evaluation of the reduction operation, we clarified the following properties. First, the relative errors of distances between vectors for the random projection are the lowest in all three databases. Second, the recognition rates of the random projection are the highest in every database. Third, the recognition rates of the vector-based classification method are higher than those of the image-based classification method. Therefore, the random projection is the best dimension-reduction method to preserve the topology of images in the classification space.

References

1. Viola, P., Jones, M.: Robust real-time face detection. International Journal of Computer Vision 57(2), 137–154 (2004)
2. Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., Gong, Y.: Locality-constrained linear coding for image classification. In: Proc. of Computer Vision and Pattern Recognition, pp. 3360–3367 (2010)
3. Bruna, J., Mallat, S.: Classification with scattering operators. In: Proc. of Computer Vision and Pattern Recognition, pp. 1561–1566 (2011)
4. Iijima, T.: Theory of pattern recognition. Electronics and Communications in Japan, 123–134 (1963)
5. Watanabe, S., Labert, P.F., Kulikowski, C.A., Buxton, J.L., Walker, R.: Evaluation and selection of variables in pattern recognition. Computer and Information Science II, 91–122 (1967)
6. Oja, E.: Subspace Methods of Pattern Recognition. Research Studies Press (1983)
7. Otsu, N.: Mathematical Studies on Feature Extraction in Pattern Recognition. PhD thesis, Electrotechnical Laboratory (1981)
8. Vempala, S.S.: The Random Projection Method, vol. 65. American Mathematical Society (2004)
9. Johnson, W., Lindenstrauss, J.: Extensions of lipschitz maps into a hilbert space. Contemporary Mathematics 26, 189–206 (1984)
10. Sakai, T., Imiya, A.: Practical algorithms of spectral clustering: Toward large-scale vision-based motion analysis. Machine Learning for Vision-Based Motion Analysis, 3–26 (2011)
11. Georghiades, A., Belhumeur, P., Kriegman, D.: From few to many: Illumination cone models for face recognition under variable lighting and pose. IEEE Trans. Pattern Anal. Mach. Intelligence 23(6), 643–660 (2001)
12. Leibe, B., Schiele, B.: Analyzing appearance and contour based methods for object categorization. In: Proc. of Computer Vison and Pattern Recognition, vol. 2, pp. 409–415 (2003)
13. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. of the IEEE 86, 2278–2324 (1998)