



HOKKAIDO UNIVERSITY

Title	POCS-Based Annotation Method Using Kernel PCA for Semantic Image Retrieval
Author(s)	Ogawa, Takahiro; Haseyama, Miki
Citation	IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, E91-A(8), 1915-1923 https://doi.org/10.1093/ietfec/e91-a.8.1915
Issue Date	2008-08-01
Doc URL	https://hdl.handle.net/2115/52943
Rights	copyright©2008 IEICE
Type	journal article
File Information	TFECCe91-a8_1915_1923.pdf



POCS-Based Annotation Method Using Kernel PCA for Semantic Image Retrieval

Takahiro OGAWA^{†a)}, Student Member and Miki HASEYAMA[†], Member

SUMMARY A projection onto convex sets (POCS)-based annotation method for semantic image retrieval is presented in this paper. Utilizing database images previously annotated by keywords, the proposed method estimates unknown semantic features of a query image from its known visual features based on a POCS algorithm, which includes two novel approaches. First, the proposed method semantically assigns database images to some clusters and introduces a nonlinear eigenspace of visual and semantic features in each cluster into the constraint of the POCS algorithm. This approach accurately provides semantic features for each cluster by using its visual features in the least squares sense. Furthermore, the proposed method monitors the error converged by the POCS algorithm in order to select the optimal cluster including the query image. By introducing the above two approaches into the POCS algorithm, the unknown semantic features of the query image are successfully estimated from its known visual features. Consequently, similar images can be easily retrieved from the database based on the obtained semantic features. Experimental results verify the effectiveness of the proposed method for semantic image retrieval.

key words: CBIR, semantic image analysis, POCS, KPCA

1. Introduction

There have recently been many studies on image retrieval due to rapid development of computer technology and Internet technology. Early image retrieval approaches based on keyword features can be traced back to the late 1970s and they were mainly developed in the database management and information retrieval community. In these approaches, query images are first annotated manually by keywords. Then similar images can be retrieved from their corresponding keywords. However, the retrieval performance of these approaches depends on the accuracy of the manual annotation.

To overcome the difficulties in these approaches, an alternative approach, content-based image retrieval (CBIR) [2]–[4], was proposed in the early 1990s. This approach relies on visual feature-based representations, such as color, texture and shape, which can be extracted automatically from images. Traditional approaches thus retrieve images whose visual features are similar to those of query images from the database. This greatly alleviates the difficulties in manual annotation-based approaches since feature extraction is fully automatic. However, the existing approaches

cannot accurately grasp the semantic concepts of images from their visual features. Hence, even after more than a decade of intensified research, their retrieval performance is still not satisfactory.

In order to solve this problem, content-based automatic annotation, which estimates unknown semantic concepts from the known visual features, has been proposed [5], [6]. This approach can estimate the semantic concepts of the query image by utilizing other annotated images as training data. It should be noted that there are many images in a database whose semantics are quite different from those of the query image. Thus, it is desirable that the semantic concepts should be estimated from only the images whose semantics are similar to those of the query image. Unfortunately, those similar images cannot be selected by traditional methods since the semantic concepts of the query image are unknown. Although several methods [7], [8], which select similar images and regions for the annotation of the query image, have been proposed, the selected results are similar to each other only in the visual features. Thus, since the images and the regions, whose semantic features are different from those of the query image, may be utilized for the annotation, their performance tends to become worse.

In this paper, we propose a new annotation method based on a theory of projections onto convex sets (POCS) [9] to realize semantic image retrieval. Our method includes the new semantic feature “confidence factor,” which represents how reliable each keyword is for a target image. In addition, database images annotated by keywords are semantically assigned to several clusters based on the confidence factors. Utilizing the obtained clusters, our POCS-based annotation algorithm estimates unknown confidence factors of a query image from its known visual features. The main contributions of this POCS algorithm are twofold: 1) introduction of a nonlinear eigenspace [10], [11] of visual features and confidence factors in each cluster into the constraint of the POCS algorithm and 2) adaptive selection of the optimal cluster for the query image based on the converged error. The first approach accurately provides the confidence factors for each cluster by using its visual features in the least squares sense, and the second approach can select the optimal cluster including the query image even if its confidence factors are unknown. By combining these two approaches, the proposed method accurately estimates the confidence factors of the query image from its known visual features. Therefore, the semantics of the query image can be successfully represented by them. Consequently, based

Manuscript received December 7, 2007.

Manuscript revised March 7, 2008.

[†]The authors are with the Graduate School of Information Science and Technology, Hokkaido University, Sapporo-shi, 060-0814 Japan.

a) E-mail: ogawa@lmd.ist.hokudai.ac.jp

DOI: 10.1093/ietfec/e91–a.8.1915

on the obtained confidence factors, similar images can be easily retrieved from the database.

This paper is organized as follows. The POCS algorithm is explained in Sect. 2. In Sect. 3, the image annotation method based on the POCS algorithm is presented. Semantic image retrieval realized by the proposed method is also explained in this section. In Sect. 4, the effectiveness of our annotation method is verified by results of experiments. Concluding remarks are presented in Sect. 5.

2. POCS Algorithm

In the POCS algorithm [9], every known property of an original vector v in the Hilbert space can be formulated as a corresponding convex set C_i ($i = 1, 2, \dots, n$). The original vector v is then assumed to lie in the intersection of these convex sets, i.e.,

$$v \in C_0 = \bigcap_{i=1}^n C_i. \quad (1)$$

Since all of the known properties are captured in the intersection of all the convex sets, C_0 , the problem of estimating the original vector v from its n properties is equivalent to that of finding an element in C_0 . If the projection operator P_i of the convex set C_i is given as

$$\|v - P_i v\| = \min_{\mu \in C_i} \|v - \mu\| \quad (i = 1, 2, \dots, n), \quad (2)$$

the problem is recursively solvable. Therefore, the iterative equation given by

$$v_t = P_n P_{n-1} \cdots P_2 P_1 v_{t-1} \quad t = 1, 2, \dots \quad (3)$$

will converge to a limiting point of the intersection $C_0 = \bigcap_{i=1}^n C_i$, as $t \rightarrow \infty$, for an arbitrary initial element.

The POCS framework ensures that the final solution is optimal mathematically. The optimality here is in the sense that the final solution satisfies all of the known constraints about the original vector.

3. POCS-Based Image Annotation Method

A POCS-based image annotation method for semantic image retrieval is presented in this section. It is assumed that every image in the database is annotated by keywords and that the semantic features of every image can be defined. It is also assumed that the query image does not have any keywords and that its semantic features are unknown. Therefore, by estimating the semantic features of the query image from the database images, annotation is performed in order to retrieve similar images.

Figure 1 shows an outline of our POCS-based image annotation method. Note that the images in the database have a large number of keywords. Hence, there are many images with semantic features that are quite different from those of the query image. These images should not affect the annotation of the query image. Thus, in order to perform

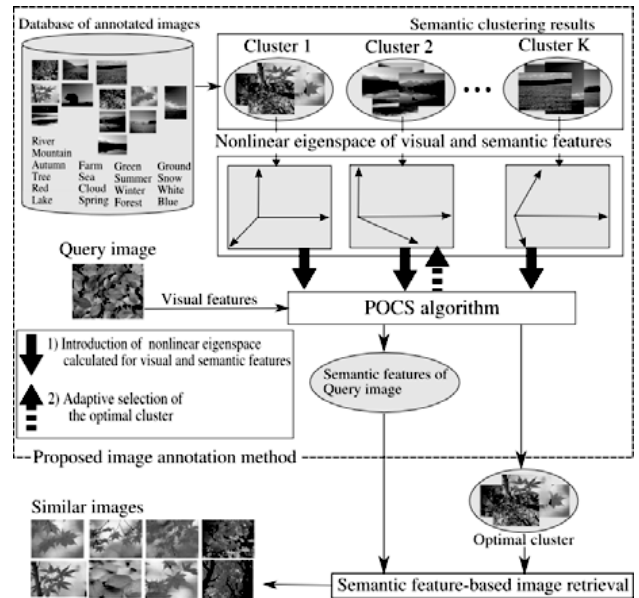


Fig. 1 Outline of the proposed image annotation method.

this annotation more precisely, we first assign the database images to several clusters based on their semantic features.

From the clustered images, the semantic features of the query image are estimated by the POCS algorithm that includes the following two novel approaches: **1)** introduction of a nonlinear eigenspace of the visual and semantic features in each cluster into the constraint and **2)** adaptive selection of the optimal cluster for the query image based on the converged errors. The first approach accurately provides the semantic features for each cluster from its visual features, and the second approach is necessary to select the optimal cluster including the query image from only its known visual features. By using these two approaches, adaptive estimation of the semantic features can be achieved, and successful annotation of the query image should be possible. Therefore, by utilizing the obtained semantic features, images that are semantically similar to the query image can be retrieved from the database.

In order to realize the above annotation, the annotated images in the database must first be assigned to some clusters. In addition, the nonlinear eigenspace of each cluster must be calculated for the constraint of the POCS algorithm. Semantic clustering of the database images and calculation of the nonlinear eigenspace for each cluster are described in Sects. 3.1 and 3.2, respectively. In Sect. 3.3, the POCS-based annotation algorithm is presented and application of the algorithm to semantic image retrieval is shown.

3.1 Semantic Clustering of Annotated Images

In this subsection, semantic clustering of annotated images in the database is presented. Suppose that there are N images f_i ($i = 1, 2, \dots, N$) in the database and they are annotated by L keywords. First, the proposed method generates a vector $\tilde{\mathbf{x}}_i = [\tilde{x}_i^1, \tilde{x}_i^2, \dots, \tilde{x}_i^L]'$, whose elements $\tilde{x}_i^l$

($l = 1, 2, \dots, L$) are one or zero, for each image f_i . Specifically, if the l -th keyword is assigned to the target image f_i , the element $\tilde{x}_i^l$ becomes one. Otherwise, $\tilde{x}_i^l$ becomes zero. Furthermore, since the number of keywords assigned to each image is different, we normalize $\tilde{\mathbf{x}}_i$ and calculate a new vector $\mathbf{x}_i$ as follows:

$$\mathbf{x}_i = \frac{\tilde{\mathbf{x}}_i}{\|\tilde{\mathbf{x}}_i\|}, \quad (4)$$

where each element of $\mathbf{x}_i$ represents the reliability of each keyword for the target image f_i . Therefore, we use this new semantic feature as a “confidence factor.”

The proposed method regards $\mathbf{x}_i$ ($i = 1, 2, \dots, N$) as semantic feature vectors and performs their clustering that minimizes the following new criterion:

$$E = \sum_{i=1}^N \sum_{j=1}^N s(i, j) \|\mathbf{x}_i - \mathbf{x}_j\|^2, \quad (5)$$

where the number of clusters is set to K . In the above equation, $s(i, j)$ is defined below.

$$s(i, j) = \begin{cases} 1 & \text{if } C(i) = C(j) \\ 0 & \text{otherwise,} \end{cases} \quad (6)$$

where $C(i)$ ($i = 1, 2, \dots, N$) represents the cluster including the image f_i and its value is $1, 2, \dots, K$. From Eq. (4), since $\|\mathbf{x}_i\| = 1$ and $\|\mathbf{x}_j\| = 1$, Eq. (5) is rewritten as follows:

$$E = \sum_{i=1}^N \sum_{j=1}^N s(i, j) (2 - 2\mathbf{x}_i' \mathbf{x}_j), \quad (7)$$

where vector/matrix transpose is defined by the superscript $'$ in this paper. In the above equation, criterion E becomes lower when the ratio of the same keywords between the images in the same cluster is higher. Therefore, criterion E is useful for the clustering of annotated images in the database. The proposed method realizes the clustering of the database images f_i ($i = 1, 2, \dots, N$) by searching $C(i)$ ($i = 1, 2, \dots, N$), which minimize criterion E in Eq. (7). The details of its procedures are shown below.

1) Calculation of initial clusters

The database images f_i ($i = 1, 2, \dots, N$) are arbitrarily assigned to K clusters, where the number of the images belonging to each cluster k ($k = 1, 2, \dots, K$) M^k is set to the same value $\frac{N}{K}$.

2) Renewal of clusters

Two images f_i and f_j ($i, j \in \{1, 2, \dots, N | i \neq j\}$) are arbitrarily selected from the database. If criterion E becomes lower by replacing their clusters with each other, these images f_i and f_j are respectively assigned to clusters $C(j)$ and $C(i)$, i.e., the values of $C(i)$ and $C(j)$ are replaced with each other.

3) Determination of clustering results

The procedure 2) is iterated T_c times, and the values of $C(i)$ ($i = 1, 2, \dots, N$) are obtained as the clustering results of the database images f_i ($i = 1, 2, \dots, N$).

As shown in the above procedures, the clustering of the database images can be achieved. In the proposed method, the iteration number T_c is set to 2.0×10^6 .

3.2 Calculation of Nonlinear Eigenspaces

As shown in the previous subsection, the images in the database are assigned to K clusters. In this subsection, calculation of the nonlinear eigenspace of visual and semantic features for each cluster k ($k = 1, 2, \dots, K$) using the proposed method is shown.

For each image f_j^k ($j = 1, 2, \dots, M^k$) in cluster k , we first divide the image into $B \times B$ blocks and calculate their color histograms with Q bins. Then by aligning the elements of the color histograms, a new vector $\mathbf{y}_j^k$ ($\in \mathbf{R}^{QB^2}$) is defined. This vector is normalized as $\|\mathbf{y}_j^k\| = 1$. Next, we denote $\mathbf{x}_i$ ($i = 1, 2, \dots, N$) belonging to cluster k as $\mathbf{x}_j^k$ ($j = 1, 2, \dots, M^k$). Furthermore, the proposed method maps $\mathbf{x}_j^k$ and $\mathbf{y}_j^k$ into the feature space $\mathbf{F}$ via a nonlinear map [10], and $\phi_{\mathbf{x}}(\mathbf{x}_j^k)$ and $\phi_{\mathbf{y}}(\mathbf{y}_j^k)$ are obtained. In contrast to linear space, nonlinear space enables part of the high-order statistics to be captured. According to this characteristic, we utilize the nonlinear map. Note that the mapped results $\phi_{\mathbf{x}}(\mathbf{x}_j^k)$ and $\phi_{\mathbf{y}}(\mathbf{y}_j^k)$ are high-dimensional or infinite-dimensional, and it therefore may not be possible to calculate them directly. Fortunately, it is well known that the following computational procedures depend only on the inner products in the feature space, which can be efficiently obtained from suitable kernel functions $\kappa_{\mathbf{x}}(\cdot, \cdot)$ and $\kappa_{\mathbf{y}}(\cdot, \cdot)$ [12]. Specifically, given arbitrary vectors $\mathbf{x}_a$ and $\mathbf{x}_b$ ($\in \mathbf{R}^L$) in the input space, the inner product of $\phi_{\mathbf{x}}(\mathbf{x}_a)$ and $\phi_{\mathbf{x}}(\mathbf{x}_b)$ is calculated as follows:

$$\phi_{\mathbf{x}}(\mathbf{x}_a)' \phi_{\mathbf{x}}(\mathbf{x}_b) = \kappa_{\mathbf{x}}(\mathbf{x}_a, \mathbf{x}_b). \quad (8)$$

In addition, given arbitrary vectors $\mathbf{y}_a$ and $\mathbf{y}_b$ ($\in \mathbf{R}^{QB^2}$) in the input space, the inner product of $\phi_{\mathbf{y}}(\mathbf{y}_a)$ and $\phi_{\mathbf{y}}(\mathbf{y}_b)$ is also calculated as follows:

$$\phi_{\mathbf{y}}(\mathbf{y}_a)' \phi_{\mathbf{y}}(\mathbf{y}_b) = \kappa_{\mathbf{y}}(\mathbf{y}_a, \mathbf{y}_b). \quad (9)$$

In Eqs. (8) and (9), several functions (e.g., Gaussian, polynomial, or sigmoid) can be applied to $\kappa_{\mathbf{x}}(\cdot, \cdot)$ and $\kappa_{\mathbf{y}}(\cdot, \cdot)$. In the proposed method, we utilize the Gaussian kernel. The Gaussian kernel is a default “general purpose kernel” in the kernel methods community [13]. The parameters of the Gaussian kernel functions $\kappa_{\mathbf{x}}(\cdot, \cdot)$ and $\kappa_{\mathbf{y}}(\cdot, \cdot)$ are respectively $\sigma_{\mathbf{x}}^2$ and $\sigma_{\mathbf{y}}^2$, and they are set to the variance of $\|\mathbf{x}_i - \mathbf{x}_j\|$ and $\|\mathbf{y}_i - \mathbf{y}_j\|$ ($i, j = 1, 2, \dots, N$), respectively.

From the vectors $\phi_{\mathbf{x}}(\mathbf{x}_j^k)$ and $\phi_{\mathbf{y}}(\mathbf{y}_j^k)$ obtained by the above procedures, the proposed method defines a vector $\mathbf{z}_j^k$ as follows:

$$\mathbf{z}_j^k = \begin{bmatrix} \phi_{\mathbf{x}}(\mathbf{x}_j^k) \\ \phi_{\mathbf{y}}(\mathbf{y}_j^k) \end{bmatrix}. \quad (10)$$

Then, given a matrix $\Xi^k = [\mathbf{z}_1^k, \mathbf{z}_2^k, \dots, \mathbf{z}_{M^k}^k]$ from the above

vector $\mathbf{z}_j^k$, Ξ^k satisfies the following singular value decomposition:

$$\Xi^k \mathbf{H}^k \cong \mathbf{U}^k \mathbf{\Lambda}^k \mathbf{V}^{k'}, \quad (11)$$

where $\mathbf{H}^k$ is a centering matrix and is defined as follows:

$$\mathbf{H}^k = \mathbf{I}^k - \frac{1}{M^k} \mathbf{1}^k \mathbf{1}^{k'}. \quad (12)$$

In this equation, $\mathbf{I}^k$ is the $M^k \times M^k$ identity matrix and $\mathbf{1}^k = [1, 1, \dots, 1]'$ is an $M^k \times 1$ vector. Moreover, in Eq. (11),

$$\mathbf{U}^k = [\mathbf{u}_1^k, \mathbf{u}_2^k, \dots, \mathbf{u}_{D^k}^k] \quad (13)$$

and

$$\mathbf{V}^k = [\mathbf{v}_1^k, \mathbf{v}_2^k, \dots, \mathbf{v}_{D^k}^k] \quad (14)$$

are D^k -dimensional eigenvector matrices of $\Xi^k \mathbf{H}^k \mathbf{H}^k \Xi^{k'}$ and $\Xi^{k'} \mathbf{H}^k \mathbf{H}^k \Xi^k$, respectively, and $\mathbf{\Lambda}^k$ is their eigenvalue matrix. In our method, the dimension D^k is set to the value of which the cumulative proportion cp^k is larger than Th. The cumulative proportion cp^k is calculated as follows:

$$\text{cp}^k = \frac{\sum_{d=1}^{D^k} \lambda_d^k}{\sum_{d=1}^{M^k} \lambda_d^k}, \quad (15)$$

where λ_d^k ($\lambda_1^k \geq \lambda_2^k \geq \dots \geq \lambda_{M^k}^k$) is the eigenvalue of the eigenvector $\mathbf{u}_d^k$.

Utilizing the matrix $\mathbf{U}^k \mathbf{U}^{k'}$, the proposed method can project arbitrary vectors $\mathbf{z}$, the dimension of which is the same as that of $\mathbf{z}_j^k$, onto the eigenspace spanned by its eigenvectors $\mathbf{u}_d^k$ ($d = 1, 2, \dots, D^k$). Note that the eigenvectors $\mathbf{u}_d^k$ are high-dimensional, and $\mathbf{U}^k$ therefore cannot be calculated directly. Thus, we introduce the computational scheme shown in Eqs. (8) and Eqs. (9) into the calculation of $\mathbf{U}^k \mathbf{U}^{k'} \mathbf{z}$. From the singular value decomposition in Eq. (11), the matrix $\mathbf{U}^k$ can be rewritten as follows:

$$\mathbf{U}^k \cong \Xi^k \mathbf{H}^k \mathbf{V}^k \mathbf{\Lambda}^{k-1}. \quad (16)$$

Then arbitrary vectors $\mathbf{z}$ can be projected onto the nonlinear eigenspace of cluster k by using the above equation.

Finally, we explain the correlation between the visual features and the semantic features calculated from the nonlinear eigenspace in the proposed method. The eigenvector $\mathbf{u}_d^k$ ($d = 1, 2, \dots, D^k$) utilized in the proposed method is the vector $\mathbf{u}$ maximizing the variance of $\mathbf{u}'(\mathbf{z}_j^k - \bar{\mathbf{z}}^k)$ ($j = 1, 2, \dots, M^k$) under the constraint $\|\mathbf{u}\| = 1$. This means the eigenvector $\mathbf{u}_d^k$ maximizes the correlation between $\mathbf{u}'(\mathbf{z}_j^k - \bar{\mathbf{z}}^k)$ and itself. Denote the variance of $\mathbf{u}'(\mathbf{z}_j^k - \bar{\mathbf{z}}^k)$ ($j = 1, 2, \dots, M^k$) as λ ,

$$\begin{aligned} \lambda &= \mathbf{u}' \Xi^k \mathbf{H}^k \mathbf{H}^k \Xi^{k'} \mathbf{u} \\ &= [\mathbf{u}_x', \mathbf{u}_y'] \begin{bmatrix} \mathbf{C}_{xx}^k & \mathbf{C}_{xy}^k \\ \mathbf{C}_{yx}^k & \mathbf{C}_{yy}^k \end{bmatrix} \begin{bmatrix} \mathbf{u}_x \\ \mathbf{u}_y \end{bmatrix} \\ &= 2\mathbf{u}_x' \mathbf{C}_{xy}^k \mathbf{u}_y + (\mathbf{u}_x' \mathbf{C}_{xx}^k \mathbf{u}_x + \mathbf{u}_y' \mathbf{C}_{yy}^k \mathbf{u}_y). \end{aligned} \quad (17)$$

In the above equation, $\mathbf{u}_x$ and $\mathbf{u}_y$ ($\mathbf{u} = [\mathbf{u}_x', \mathbf{u}_y']'$) are respectively D_x and D_y -dimensional vectors. Note that D_x and D_y represent the dimensions of the nonlinear spaces of the nonlinear maps ϕ_x and ϕ_y , respectively. Furthermore, $\mathbf{C}_{xx}^k$ ($= \Xi_x^k \mathbf{H}^k \mathbf{H}^k \Xi_x^{k'}$), $\mathbf{C}_{xy}^k$ ($= \Xi_x^k \mathbf{H}^k \mathbf{H}^k \Xi_y^{k'}$), $\mathbf{C}_{yx}^k$ ($= \Xi_y^k \mathbf{H}^k \mathbf{H}^k \Xi_x^{k'}$), and $\mathbf{C}_{yy}^k$ ($= \Xi_y^k \mathbf{H}^k \mathbf{H}^k \Xi_y^{k'}$) are respectively $D_x \times D_x$, $D_x \times D_y$, $D_y \times D_x$, and $D_y \times D_y$ matrices, where

$$\Xi_x^k = [\phi_x(\mathbf{x}_1^k), \phi_x(\mathbf{x}_2^k), \dots, \phi_x(\mathbf{x}_{M^k}^k)] \quad (18)$$

and

$$\Xi_y^k = [\phi_y(\mathbf{y}_1^k), \phi_y(\mathbf{y}_2^k), \dots, \phi_y(\mathbf{y}_{M^k}^k)]. \quad (19)$$

From Eq. (17), the eigenvector $\mathbf{u}_d^k$ ($d = 1, 2, \dots, D^k$) is the vector $\mathbf{u}$ maximizing λ containing not only the correlation between the visual features and the semantic features but also the correlation between the visual features and that between the semantic features. Note that when $\mathbf{u} = \mathbf{u}_d^k$ ($d = 1, 2, \dots, D^k$), λ in Eq. (17) is equivalent to λ_d^k .

In the point of view concerning the correlation between the visual features and the semantic features, the kernel canonical correlation analysis (CCA) can be applied to the proposed method. Since the kernel CCA can maximize the correlation between the visual features and the semantic features, it is more suitable than the kernel PCA. However, the proposed method utilizes the kernel PCA in the following reason. In this paper, we assume that all of the visual features can be calculated and all of the semantic features are unknown. However, when the proposed method is actually applied to the annotation of query images, these two features may be partly unknown and known, respectively. In such a case, our method can estimate the unknown semantic features by using the same eigenvectors $\mathbf{u}_d^k$ and eigenvalues λ_d^k ($d = 1, 2, \dots, D^k$) obtained by the kernel PCA. On the other hand, the kernel CCA must recalculate the coefficients, which provide the canonical variates, and the canonical correlation coefficients from the known features and the unknown features of the query image for each cluster. Therefore, since the eigenvectors and the eigenvalues calculated by the kernel PCA are robust to the change of the known and unknown features, it is applied to the proposed method.

In this way, we can calculate the nonlinear eigenspace of visual and semantic features for each cluster. Since the eigenspace can express the correlation between the visual features and the semantic features, estimation of the unknown semantic features of the query image from its known visual features is feasible. The details of this estimation is described in the following subsection.

3.3 POCS-Based Annotation Algorithm and Its Application for Image Retrieval

In this subsection, estimation of the confidence factors of the query image by utilizing the POCS algorithm is shown. Retrieval of images similar to the query image using the estimated confidence factors is also shown.

First, we calculate a vector $\mathbf{y}$ from the query image in the same way as $\mathbf{y}_j^k$. In addition, we denote an unknown vector of the confidence factors as $\mathbf{x}$, and we define $\mathbf{z}$ as follows:

$$\mathbf{z} = \begin{bmatrix} \phi_{\mathbf{x}}(\mathbf{x}) \\ \phi_{\mathbf{y}}(\mathbf{y}) \end{bmatrix}. \quad (20)$$

From this vector $\mathbf{z}$, the proposed method calculates a new vector $\hat{\mathbf{z}}$ satisfying the following two constraints to estimate the unknown vector $\mathbf{x}$.

[Constraint 1]

Since $\phi_{\mathbf{y}}(\mathbf{y})$ is the known vector calculated from the query image, it is fixed in the vector $\hat{\mathbf{z}}$.

[Constraint 2]

In the feature space, the target vector $\hat{\mathbf{z}}$ is in the nonlinear eigenspace spanned by its eigenvectors $\mathbf{u}_1^k, \mathbf{u}_2^k, \dots$, and $\mathbf{u}_{D^k}^k$ of cluster k . Therefore, $\hat{\mathbf{z}}$ satisfies

$$\hat{\mathbf{z}} - \bar{\mathbf{z}}^k = \mathbf{U}^k \mathbf{U}^{k'} (\hat{\mathbf{z}} - \bar{\mathbf{z}}^k), \quad (21)$$

where $\bar{\mathbf{z}}^k$ is a center vector of cluster k and obtained by the following equation:

$$\bar{\mathbf{z}}^k = \frac{1}{M^k} \Xi^k \mathbf{1}^k. \quad (22)$$

Then, from Eqs. (16) and (22), Eq. (21) is rewritten as follows:

$$\begin{aligned} \hat{\mathbf{z}} &\cong \Xi^k \mathbf{H}^k \mathbf{V}^k \Lambda^{k-2} \mathbf{V}^{k'} \mathbf{H}^{k'} \Xi^{k'} \left(\hat{\mathbf{z}} - \frac{1}{M^k} \Xi^k \mathbf{1}^k \right) + \frac{1}{M^k} \Xi^k \mathbf{1}^k \\ &= \Xi^k \mathbf{S}^k \Xi^{k'} \hat{\mathbf{z}} - \frac{1}{M^k} \Xi^k (\mathbf{S}^k \Xi^{k'} \Xi^k - \mathbf{I}^k) \mathbf{1}^k, \end{aligned} \quad (23)$$

where $\mathbf{S}^k$ is defined as follows:

$$\mathbf{S}^k = \mathbf{H}^k \mathbf{V}^k \Lambda^{k-2} \mathbf{V}^{k'} \mathbf{H}^k. \quad (24)$$

Utilizing the POCS algorithm, the proposed method calculates the vector $\hat{\mathbf{z}}$ that satisfies the above two constraints from the initial vector $\mathbf{z}$. Specifically, these constraints are respectively utilized as the closed convex sets C_1 and C_2 in Eq. (1), and $\hat{\mathbf{z}}$ is calculated by their projection operators P_1 and P_2 in Eq. (3). Note that if there is no intersection of the two convex sets, we cannot obtain the result $\hat{\mathbf{z}}$ satisfying both of the constraints. In such a case, the proposed method outputs the result that satisfies Constraint 1. Furthermore, we define $\hat{\mathbf{z}} = [\phi_{\mathbf{x}}(\hat{\mathbf{x}})', \phi_{\mathbf{y}}(\mathbf{y})']'$ from Constraint 1, and Eq. (23) is rewritten as

$$\begin{aligned} \begin{bmatrix} \phi_{\mathbf{x}}(\hat{\mathbf{x}}) \\ \phi_{\mathbf{y}}(\mathbf{y}) \end{bmatrix} &\cong \begin{bmatrix} \Xi_{\mathbf{x}}^k \\ \Xi_{\mathbf{y}}^k \end{bmatrix} \mathbf{S}^k \begin{bmatrix} \Xi_{\mathbf{x}}^{k'} \\ \Xi_{\mathbf{y}}^{k'} \end{bmatrix} \begin{bmatrix} \phi_{\mathbf{x}}(\hat{\mathbf{x}}) \\ \phi_{\mathbf{y}}(\mathbf{y}) \end{bmatrix} \\ &\quad - \frac{1}{M^k} \begin{bmatrix} \Xi_{\mathbf{x}}^k \\ \Xi_{\mathbf{y}}^k \end{bmatrix} \left(\mathbf{S}^k \begin{bmatrix} \Xi_{\mathbf{x}}^{k'} \\ \Xi_{\mathbf{y}}^{k'} \end{bmatrix} \begin{bmatrix} \Xi_{\mathbf{x}}^k \\ \Xi_{\mathbf{y}}^k \end{bmatrix} - \mathbf{I}^k \right) \mathbf{1}^k \\ &= \begin{bmatrix} \Xi_{\mathbf{x}}^k \\ \Xi_{\mathbf{y}}^k \end{bmatrix} \mathbf{S}^k \left(\Xi_{\mathbf{x}}^{k'} \phi_{\mathbf{x}}(\hat{\mathbf{x}}) + \Xi_{\mathbf{y}}^{k'} \phi_{\mathbf{y}}(\mathbf{y}) \right) \\ &\quad - \frac{1}{M^k} \begin{bmatrix} \Xi_{\mathbf{x}}^k \\ \Xi_{\mathbf{y}}^k \end{bmatrix} \left\{ \mathbf{S}^k \left(\Xi_{\mathbf{x}}^{k'} \Xi_{\mathbf{x}}^k + \Xi_{\mathbf{y}}^{k'} \Xi_{\mathbf{y}}^k \right) - \mathbf{I}^k \right\} \mathbf{1}^k. \end{aligned} \quad (25)$$

Then $\Xi_{\mathbf{x}}^{k'} \phi_{\mathbf{x}}(\hat{\mathbf{x}})$ and $\Xi_{\mathbf{y}}^{k'} \phi_{\mathbf{y}}(\mathbf{y})$ in Eq. (25) are respectively obtained by the following equations:

$$\begin{aligned} \Xi_{\mathbf{x}}^{k'} \phi_{\mathbf{x}}(\hat{\mathbf{x}}) &= [\phi_{\mathbf{x}}(\mathbf{x}_1^k), \phi_{\mathbf{x}}(\mathbf{x}_2^k), \dots, \phi_{\mathbf{x}}(\mathbf{x}_{M^k}^k)]' \phi_{\mathbf{x}}(\hat{\mathbf{x}}) \\ &= [\kappa_{\mathbf{x}}(\mathbf{x}_1^k, \hat{\mathbf{x}}), \kappa_{\mathbf{x}}(\mathbf{x}_2^k, \hat{\mathbf{x}}), \dots, \kappa_{\mathbf{x}}(\mathbf{x}_{M^k}^k, \hat{\mathbf{x}})], \end{aligned} \quad (26)$$

$$\begin{aligned} \Xi_{\mathbf{y}}^{k'} \phi_{\mathbf{y}}(\mathbf{y}) &= [\phi_{\mathbf{y}}(\mathbf{y}_1^k), \phi_{\mathbf{y}}(\mathbf{y}_2^k), \dots, \phi_{\mathbf{y}}(\mathbf{y}_{M^k}^k)]' \phi_{\mathbf{y}}(\mathbf{y}) \\ &= [\kappa_{\mathbf{y}}(\mathbf{y}_1^k, \mathbf{y}), \kappa_{\mathbf{y}}(\mathbf{y}_2^k, \mathbf{y}), \dots, \kappa_{\mathbf{y}}(\mathbf{y}_{M^k}^k, \mathbf{y})]. \end{aligned} \quad (27)$$

Furthermore, $\Xi_{\mathbf{x}}^{k'} \Xi_{\mathbf{x}}^k$ and $\Xi_{\mathbf{y}}^{k'} \Xi_{\mathbf{y}}^k$ in Eq. (25) are calculated as follows:

$$\begin{aligned} \Xi_{\mathbf{x}}^{k'} \Xi_{\mathbf{x}}^k &= \begin{bmatrix} \kappa_{\mathbf{x}}(\mathbf{x}_1^k, \mathbf{x}_1^k) & \kappa_{\mathbf{x}}(\mathbf{x}_1^k, \mathbf{x}_2^k) & \dots & \kappa_{\mathbf{x}}(\mathbf{x}_1^k, \mathbf{x}_{M^k}^k) \\ \kappa_{\mathbf{x}}(\mathbf{x}_2^k, \mathbf{x}_1^k) & \kappa_{\mathbf{x}}(\mathbf{x}_2^k, \mathbf{x}_2^k) & \dots & \kappa_{\mathbf{x}}(\mathbf{x}_2^k, \mathbf{x}_{M^k}^k) \\ \vdots & \vdots & \ddots & \vdots \\ \kappa_{\mathbf{x}}(\mathbf{x}_{M^k}^k, \mathbf{x}_1^k) & \kappa_{\mathbf{x}}(\mathbf{x}_{M^k}^k, \mathbf{x}_2^k) & \dots & \kappa_{\mathbf{x}}(\mathbf{x}_{M^k}^k, \mathbf{x}_{M^k}^k) \end{bmatrix}, \\ \Xi_{\mathbf{y}}^{k'} \Xi_{\mathbf{y}}^k &= \begin{bmatrix} \kappa_{\mathbf{y}}(\mathbf{y}_1^k, \mathbf{y}_1^k) & \kappa_{\mathbf{y}}(\mathbf{y}_1^k, \mathbf{y}_2^k) & \dots & \kappa_{\mathbf{y}}(\mathbf{y}_1^k, \mathbf{y}_{M^k}^k) \\ \kappa_{\mathbf{y}}(\mathbf{y}_2^k, \mathbf{y}_1^k) & \kappa_{\mathbf{y}}(\mathbf{y}_2^k, \mathbf{y}_2^k) & \dots & \kappa_{\mathbf{y}}(\mathbf{y}_2^k, \mathbf{y}_{M^k}^k) \\ \vdots & \vdots & \ddots & \vdots \\ \kappa_{\mathbf{y}}(\mathbf{y}_{M^k}^k, \mathbf{y}_1^k) & \kappa_{\mathbf{y}}(\mathbf{y}_{M^k}^k, \mathbf{y}_2^k) & \dots & \kappa_{\mathbf{y}}(\mathbf{y}_{M^k}^k, \mathbf{y}_{M^k}^k) \end{bmatrix}. \end{aligned}$$

From Eq. (25), $\phi_{\mathbf{x}}(\hat{\mathbf{x}})$ of the vector $\hat{\mathbf{x}}$ can be obtained by iterating the following $\phi_{\mathbf{x}}(\hat{\mathbf{x}}_t)$:

$$\begin{aligned} \phi_{\mathbf{x}}(\hat{\mathbf{x}}_t) &\cong \Xi_{\mathbf{x}}^k \mathbf{S}^k \left(\Xi_{\mathbf{x}}^{k'} \phi_{\mathbf{x}}(\hat{\mathbf{x}}_{t-1}) + \Xi_{\mathbf{y}}^{k'} \phi_{\mathbf{y}}(\mathbf{y}) \right) \\ &\quad - \frac{1}{M^k} \Xi_{\mathbf{x}}^k \left\{ \mathbf{S}^k \left(\Xi_{\mathbf{x}}^{k'} \Xi_{\mathbf{x}}^k + \Xi_{\mathbf{y}}^{k'} \Xi_{\mathbf{y}}^k \right) - \mathbf{I}^k \right\} \mathbf{1}^k \\ &\quad (t = 1, 2, \dots, T), \end{aligned} \quad (28)$$

where $\hat{\mathbf{x}}_0 = \mathbf{x}$. Then the proposed method can calculate $\phi_{\mathbf{x}}(\hat{\mathbf{x}})$ of $\hat{\mathbf{x}} = [\hat{x}_1, \hat{x}_2, \dots, \hat{x}_L]'$ whose element $\hat{x}_l$ ($l = 1, 2, \dots, L$) corresponds to the confidence factor of the l -th keyword.

From Constraint 2 in the proposed method, the nonlinear eigenspace correctly approximates $\mathbf{z}_j^k$ ($j = 1, 2, \dots, M^k$) belonging to the same cluster k in the least squares sense[†]. Therefore, if the target vector $\mathbf{z}$ belongs to cluster k , we can accurately estimate $\hat{\mathbf{z}}$ and obtain accurate confidence factors for the query image. Unfortunately, we cannot know which cluster is the optimal for the query image before estimating its confidence factors. This means that the target vector $\mathbf{z}$ cannot be classified by the algorithm shown in 3.1. Thus, in order to solve this problem, the proposed method monitors the following error caused in the known visual features:

$$\tilde{E}^k = \|\mathbf{y} - \hat{\mathbf{y}}\|^2, \quad (29)$$

where $\hat{\mathbf{y}}$ satisfies the following equation:

$$\begin{aligned} \phi_{\mathbf{y}}(\hat{\mathbf{y}}) &\cong \Xi_{\mathbf{y}}^k \mathbf{S}^k \left(\Xi_{\mathbf{x}}^{k'} \phi_{\mathbf{x}}(\hat{\mathbf{x}}) + \Xi_{\mathbf{y}}^{k'} \phi_{\mathbf{y}}(\mathbf{y}) \right) \\ &\quad - \frac{1}{M^k} \Xi_{\mathbf{y}}^k \left\{ \mathbf{S}^k \left(\Xi_{\mathbf{x}}^{k'} \Xi_{\mathbf{x}}^k + \Xi_{\mathbf{y}}^{k'} \Xi_{\mathbf{y}}^k \right) - \mathbf{I}^k \right\} \mathbf{1}^k. \end{aligned} \quad (30)$$

[†]Given sample data, their eigenspace minimizes information loss since it is calculated on the basis of the maximum variance criterion.

Note that the vector $\hat{\mathbf{y}}$ becomes the same as $\mathbf{y}$ when an intersection of the two constraints exists. The criterion $\tilde{E}^k$ is the converged error of the POCS algorithm and the minimum distance between the target vector $\mathbf{z}$ and the nonlinear eigenspace of cluster k . Therefore, this criterion is suitable for classification of the target vector $\mathbf{z}$ as a substitute for Eq. (7). Furthermore, since the following equation is satisfied from Eq. (9):

$$\phi_{\mathbf{y}}(\mathbf{y})' \phi_{\mathbf{y}}(\hat{\mathbf{y}}) = \exp\left(-\frac{\|\mathbf{y} - \hat{\mathbf{y}}\|^2}{\sigma_y^2}\right), \quad (31)$$

Eq. (29) is rewritten as follows:

$$\begin{aligned} \tilde{E}^k \cong & -\sigma_y^2 \log\left\{\phi_{\mathbf{y}}(\mathbf{y})' \Xi_{\mathbf{y}}^k \mathbf{S}^k \left(\Xi_{\mathbf{x}}^{k'} \phi_{\mathbf{x}}(\hat{\mathbf{x}}) + \Xi_{\mathbf{y}}^{k'} \phi_{\mathbf{y}}(\mathbf{y})\right) \right. \\ & \left. - \frac{1}{M^k} \phi_{\mathbf{y}}(\mathbf{y})' \Xi_{\mathbf{y}}^k \left\{\mathbf{S}^k \left(\Xi_{\mathbf{x}}^{k'} \Xi_{\mathbf{x}}^k + \Xi_{\mathbf{y}}^{k'} \Xi_{\mathbf{y}}^k\right) - \mathbf{I}^k\right\} \mathbf{1}^k\right\}. \end{aligned} \quad (32)$$

By minimizing the above equation, we can select the optimal cluster k^{opt} for the query image. Thus, the proposed method outputs the result $\hat{\mathbf{z}}$ obtained by cluster k^{opt} and regards its vector $\phi_{\mathbf{x}}(\hat{\mathbf{x}})$ as the final estimation results of the confidence factors in the feature space.

In this way, the proposed method can estimate the confidence factors of the keywords for the query image. Then the distances of the confidence factors between the query image and the images belonging to cluster k^{opt} can be calculated as follows:

$$P_j^{k^{\text{opt}}} = \|\hat{\mathbf{x}} - \mathbf{x}_j^{k^{\text{opt}}}\|^2 \quad (j = 1, 2, \dots, M^{k^{\text{opt}}}). \quad (33)$$

In the same way as the derivation of Eq. (32), the above equation is rewritten as follows:

$$\begin{aligned} P_j^{k^{\text{opt}}} \cong & -\sigma_x^2 \log\left\{\phi_{\mathbf{x}}(\mathbf{x}_j^{k^{\text{opt}}})' \Xi_{\mathbf{x}}^{k^{\text{opt}}} \mathbf{S}^{k^{\text{opt}}} \left(\Xi_{\mathbf{x}}^{k^{\text{opt}'} } \phi_{\mathbf{x}}(\hat{\mathbf{x}}) + \Xi_{\mathbf{y}}^{k^{\text{opt}'} } \phi_{\mathbf{y}}(\mathbf{y})\right) \right. \\ & \left. - \frac{1}{M^{k^{\text{opt}}}} \phi_{\mathbf{x}}(\mathbf{x}_j^{k^{\text{opt}}})' \Xi_{\mathbf{x}}^{k^{\text{opt}}} \left\{\mathbf{S}^{k^{\text{opt}}} \left(\Xi_{\mathbf{x}}^{k^{\text{opt}'} } \Xi_{\mathbf{x}}^{k^{\text{opt}}} + \Xi_{\mathbf{y}}^{k^{\text{opt}'} } \Xi_{\mathbf{y}}^{k^{\text{opt}}}\right) \right. \right. \\ & \left. \left. - \mathbf{I}^{k^{\text{opt}}}\right\} \mathbf{1}^{k^{\text{opt}}}\right\}. \end{aligned} \quad (34)$$

Consequently, we can retrieve N_p similar images, whose $\mathbf{x}_j^{k^{\text{opt}}}$ minimize the above criterion $P_j^{k^{\text{opt}}}$, from cluster k^{opt} . This means that semantic image retrieval based on the confidence factors can be realized.

4. Experimental Results

The effectiveness of the proposed annotation method for image retrieval is verified in this section. In the experiment, we used a database containing 20000 color images of various sizes. About 30–40 keywords were assigned to each image, and the total number of keywords in the database was 6834. For the database images, we first performed clustering as described in 3.1 and assigned the images to $K (= 10)$ clusters. Next, as shown in Figs. 2–4, we applied the proposed method[†] to three kinds of query images whose keywords were unknown and retrieved similar images. In each figure, the top 23 images retrieved by the criterion in Eq. (34) are

shown. It can be seen that the proposed method realizes the semantic retrieval of images whose contents are similar to those of the query image.

Traditional CBIR approaches [2]–[4] retrieve images whose visual features are the most similar to those of the query image. Thus, since they cannot accurately grasp the semantic features from the known visual features, semantic image retrieval cannot be realized. For example, the retrieved image at the 12th order in Fig. 2 has the same content as that of the query image, but the visual features of the two images are different. This means that this image cannot be retrieved by traditional approaches. On the other hand, the proposed method enables estimation of the confidence factors, which correspond to the semantic features, from only the visual features. Images whose contents are the same as those of the query image can therefore be retrieved even if their visual features are different. Therefore, the proposed method realizes semantic image retrieval as shown in Figs. 2–4.

Note that there are several methods [7], [8], which annotate query images for similar image retrieval. These methods aim at the annotation of the query images from only the images whose semantics are similar to those of the query images. Specifically, several similar images or small regions are selected for estimating the semantics of the query image. However, they are selected based on only the similarity of the visual features. Thus, since the images and the regions, whose semantic features are different from those of the query image, may be utilized for the annotation, their performance tends to become worse. On the other hand, our method estimates the semantic features of the query image from its visual features in such a way that $\hat{\mathbf{z}}$ in Eq. (23) minimizes the distance $\tilde{E}^k$ from the nonlinear eigenspace of the visual and semantic features in each cluster. Furthermore, by searching the cluster minimizing this distance $\tilde{E}^k$, the optimal cluster k^{opt} , that provides the final estimation results of the semantic features, can be selected. Therefore, since the proposed method can select the optimal cluster based on not only the visual features but also the semantic features, the problem of the conventional methods is solvable.

Next, we quantitatively verify the performance of the proposed method. In this simulation, we used 1000 test query images for calculating precision-recall curves. First, for each query image, we retrieved similar images from the database containing 20000 images by the following two schemes:

- (i) Retrieval of $J_{(i)}$ images based on rates of corresponding keywords under the ideal condition of keywords having been assigned to the query image,
- (ii) Retrieval of $J_{(ii)}$ images based on the proposed approach or the traditional approach [2] under the condition of keywords of the query image being unknown.

Next, we regarded the $J_{(i)}$ images obtained by scheme (i) as

[†]In this simulation, we set the parameters of the proposed method as follows: $B = 4$, $Q = 64$, $\text{Th} = 0.9$, and $T = 5$.



Fig. 2 Results of retrieval using the proposed method. The query image contains semantic concepts such as “dancer,” “Yosakoi” and “festival.”

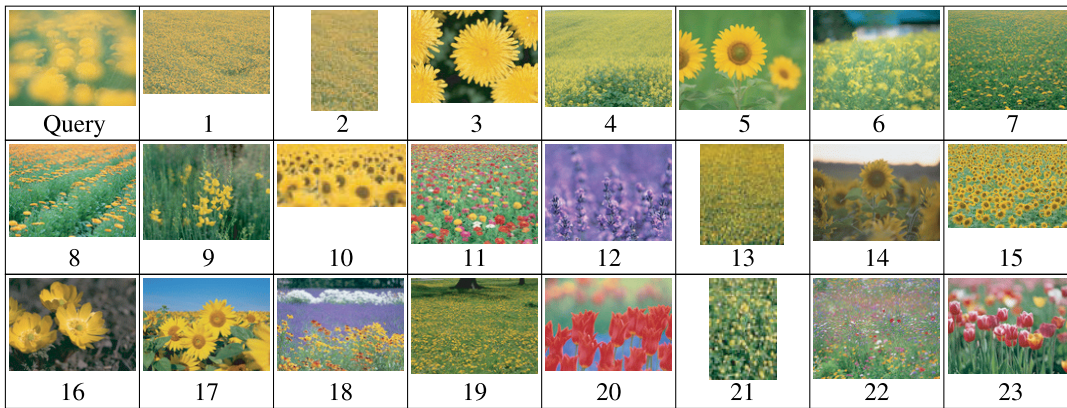


Fig. 3 Results of retrieval using the proposed method. The query image contains semantic concepts such as “dandelion,” “flower garden” and “plant.”



Fig. 4 Results of retrieval using the proposed method. The query image contains semantic concepts such as “waterfall,” “mountain stream” and “rock.”

the relevant images and calculated the precision and the recall from the results of scheme (ii). Therefore, the precision and the recall are defined as follows:

$$\begin{aligned} \text{Precision} &= \frac{\text{Num. of correctly retrieved images}}{\text{Num. of images retrieved by scheme (ii)}} \\ &= \frac{\text{Num. of correctly retrieved images}}{J_{(ii)}}, \quad (35) \end{aligned}$$

$$\begin{aligned} \text{Recall} &= \frac{\text{Num. of correctly retrieved images}}{\text{Num. of relevant images}} \\ &= \frac{\text{Num. of correctly retrieved images}}{J_{(i)}}. \quad (36) \end{aligned}$$

Note that $J_{(ii)}$ is equivalent to N_p in the proposed method. From the above definitions, we respectively show the precision-recall curves of the proposed approach and the tra-

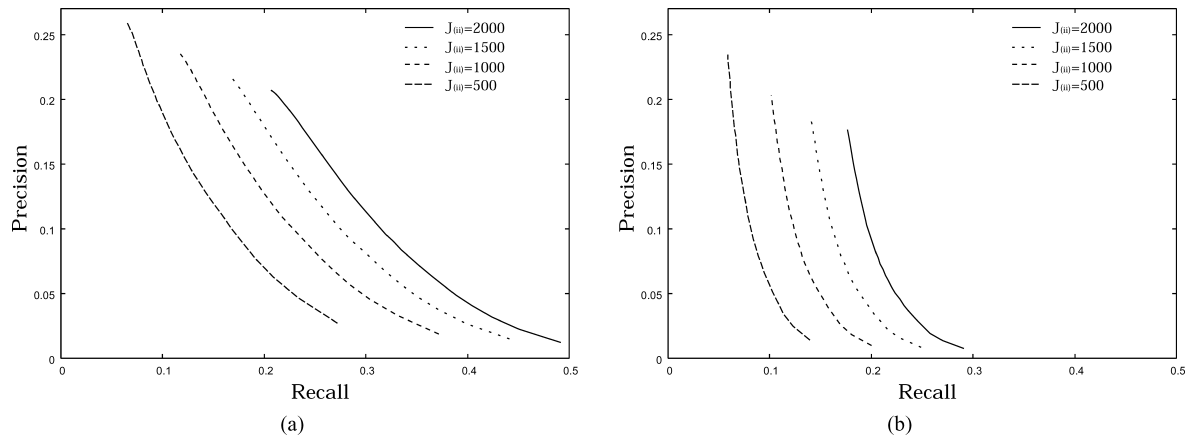


Fig. 5 Quantitative evaluation of the proposed method and the traditional method based on precision and recall: (a) Precision-recall curves of the proposed method and (b) Precision-recall curves of the traditional method [2].

Table 1 Comparison of computation time (sec) between the proposed method and the traditional method.

Retrieval approach	Computation time
Traditional approach [2]	4.89×10^{-1}
Proposed approach	1.33×10^2

ditional approach in Figs. 5(a) and (b). The four precision-recall curves in each figure respectively show the results corresponding to the values $J_{(ii)}$ of 500, 1000, 1500, and 2000. Furthermore, each curve is obtained from the results corresponding to the values $J_{(i)}$ of 50, 100, 150, $\dots$, 2000. From these figures, we can see that the proposed approach has achieved noticeable improvement over the traditional approach. Therefore, the effectiveness of the proposed approach was quantitatively specified by this experiment.

Furthermore, we verify computational cost of the proposed method. Table 1 shows the comparison of the computation time between the retrieval schemes in the proposed method and the traditional method. Each time was obtained by averaging the computation time taken in the similar image retrieval for 1000 query images. This simulation was performed on a personal computer using AMD Athlon(tm) 64 X2 Dual Core Processor 3800+ 2.0GHz CPU with 1.0Gbytes RAM. Both the proposed method and the traditional method were implemented by using Matlab. As shown in this table, the computation time of the proposed method is larger than that of the traditional method. Therefore, from this experiment, we can see that the reduction of the computational cost in the proposed method is necessary for practical use. This issue will be addressed in a future work.

Finally, we explain the determination of the number of the clusters K . We first discuss the settable range of K and further determine its optimal value by verifying the relationship between K and the performance of the proposed method. When the number of the clusters is set to K for the database containing 20000 images, the size of the matrices $\Xi_x^{k'} \Xi_x^k$ and $\Xi_y^{k'} \Xi_y^k$ in the proposed method becomes

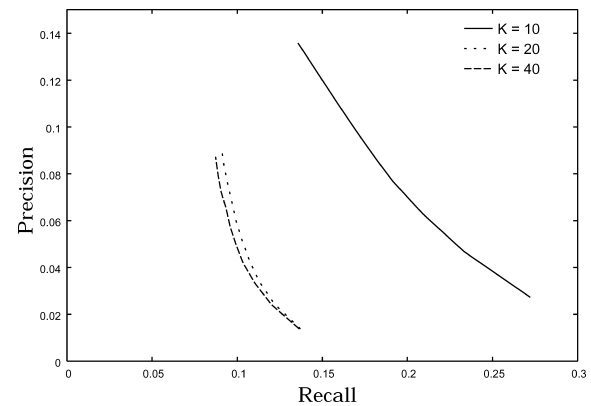


Fig. 6 Relationship between K and the precision-recall curves of the proposed method. The three curves respectively show the results corresponding to the value $J_{(ii)}$ of 500. Each curve is obtained from the results corresponding to the values $J_{(i)}$ of 50, 100, 150, $\dots$, 500.

$(20000/K) \times (20000/K)$. Note that if the size of these matrices becomes more than 2000×2000 , i.e., K is set to less than 10, the computation of Eqs. (28), (32), and (34) becomes difficult. Thus, we must set K to the value equal to or more than 10. Furthermore, we change the value of K and determine its optimal value by verifying the performance of the proposed method. Figure 6 shows the relationship between K and the precision-recall curves obtained by the same experiment as that shown in the quantitative evaluation. From this figure, we can see the precision and the recall become the highest values when $K = 10$. Therefore, we set $K = 10$. However, since it is desirable to adaptively set K to the optimal value from the obtained database, this will be the subject of the subsequent reports.

5. Conclusions

In this paper, we have proposed a POCS-based annotation method for semantic image retrieval. The proposed method clusters images in a database on the basis of their seman-

tic features and introduces a nonlinear eigenspace of the visual and semantic features in each cluster into the constraint of the POCS algorithm. This approach accurately provides semantic features for each cluster by using its visual features. Furthermore, by monitoring the error converged by the POCS algorithm, the optimal cluster for estimating the semantic features of the query image can be selected. Then the semantic features of the query image can be successfully estimated from only its known visual features. Therefore, we can easily retrieve similar images by using the estimated semantic features. Our experimental results verified the superiority of our annotation-based retrieval technique.

In this study, we manually set the parameters of the proposed method. The validity of these values should be verified from a larger image database. Furthermore, the proposed method takes more computational time than that of traditional methods. Therefore, reduction of computational cost is needed for practical use of the proposed method. These topics will be the subject of subsequent studies.

Acknowledgment

This work was supported in part by Grant-in-Aid for JSPS Fellows from the Japan Society for the Promotion of Science.

References

- [1] Y. Rui, T.S. Huang, M. Ortega, and S. Mehrotra, "Relevance feedback: A power tool for interactive content-based image retrieval," *IEEE Trans. Circuits Syst. Video Technol.*, vol.8, no.5, pp.644–655, 1998.
- [2] M. Swain and D. Ballard, "Color indexing," *Int. J. Comput. Vision*, vol.7, no.1, pp.11–32, 1991.
- [3] A.A. Goodrum, "Image information retrieval: An overview of current research," *Inf. Sci.*, vol.3, pp.63–66, 2000.
- [4] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.22, no.12, pp.1349–1380, 2000.
- [5] F. Jing, M. Li, H.-J. Zhang, and B. Zhang, "A unified framework for image retrieval using keyword and visual features," *IEEE Trans. Image Process.*, vol.14, no.7, pp.979–989, 2005.
- [6] E. Chang, K. Goh, G. Sychay, and G. Wu, "CBSA: Content-based soft annotation for multimodal image retrieval using Bayes point machines," *IEEE Trans. Circuits Syst. Video Technol.*, vol.13, no.1, pp.26–38, 2004.
- [7] I. Bartolini and P. Ciaccia, "Towards an effective semi-automatic technique for image annotation," *SEBD 2007*, 2007.
- [8] X. Rui, P. Yuan, and N. Yu, "Image annotations based on semi-supervised clustering with semantic soft constraints," *PCM 2006*, 2006.
- [9] D.C. Youla and H. Webb, "Image restoration by the method of convex projections: Part I — Theory," *IEEE Trans. Med. Imaging*, vol.MI-1, no.2, pp.81–94, Oct. 1982.
- [10] B. Schölkopf, S. Mika, C.J.C. Burges, P. Knirsch, K.-R. Müller, G. Rätsch, and A.J. Smola, "Input space versus feature space in kernel-based methods," *IEEE Trans. Neural Netw.*, vol.10, no.5, pp.1000–1017, 1999.
- [11] S. Mika, B. Schölkopf, A. Smola, K.-R. Müller, M. Scholz, and G. Rätsch, "Kernel PCA and de-noising in feature spaces," *Advances in Neural Information Processing Systems 11*, pp.536–542, 1999.
- [12] J. Mercer, "Functions of positive and negative type and their connection with the theory of integral equations," *Trans. Lond. Phil. Soc.*, (A), vol.209, pp.415–446, 1909.
- [13] K.I. Kim, M.O. Franz, and B. Schölkopf, "Iterative kernel principal component analysis for image modeling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.27, no.9, pp.1351–1366, 2005.



Takahiro Ogawa received his B.S., M.S. and Ph.D. degrees in Electronics and Information Engineering from Hokkaido University, Japan in 2003, 2005 and 2007 respectively. He is currently a JSPS Research Fellow in the Graduate School of Information Science and Technology, Hokkaido University. His research interests are digital image processing and its applications.



Miki Haseyama received her B.S., M.S. and Ph.D. degrees in Electronics from Hokkaido University, Japan in 1986, 1988 and 1993, respectively. She is currently a professor in the Graduate School of Information Science and Technology, Hokkaido University. Her research interests are digital signal processing and its applications.