Local Image Descriptors Using Supervised Kernel ICA

Masaki Yamazaki¹ and Sidney Fels²

¹ Faculty of Information Science and Engineering,
Ritsumeikan University, Shiga, Japan
`rs023018@se.ritsumei.ac.jp`

² Department of Electrical and Computer Engineering,
University of British Columbia, Vancouver, BC, Canada
`ssfels@ece.ubc.ca`

Abstract. PCA-SIFT is an extension to SIFT which aims to reduce SIFT's high dimensionality (128 dimensions) by applying PCA to the gradient image patches. However PCA is not a discriminative representation for recognition due to its global feature nature and unsupervised algorithm. In addition, linear methods such as PCA and ICA can fail in the case of non-linearity. In this paper, we propose a new discriminative method called *Supervised* Kernel ICA (SKICA) that uses a non-linear kernel approach combined with Supervised ICA-based local image descriptors. Our approach blends the advantages of a low dimensionality representation, like PCA-SIFT, with supervised learning based on non-linear properties of kernels to overcome separability limitations of non-linear representations for recognition. Using five different test data sets we show that the SKICA descriptors produce better object recognition performance than other related approaches with the same dimensionality. The SKICA-based representation has local sensitivity, non-linear independence and high class separability providing an effective method for local image descriptors.

1 Introduction

Local image descriptors are widely used in many computer vision applications such as object categorization and recognition [1] and image/video retrieval [2]. Probably one of the most popular and widely utilized local feature descriptors is SIFT [3]. SIFT descriptors have good properties including scale and rotation invariance, robustness against changes in viewpoint and illumination. This makes them an effective descriptor for object recognition applications. However, the SIFT descriptor has relatively high dimensionality (128 dimensions), which limits the performance of matching speed. PCA-SIFT [4] is an extension to SIFT which aims to reduce SIFT's high dimensionality by applying Principal Components Analysis (PCA) and yields a 36-dimensional descriptor. PCA is often used for reducing the dimensionality of an input feature space. PCA does not provide a discriminative representation for matching, however it does provide significant improvements in matching speed due to the dimensionality reduction and in reducing the high frequency noise in the descriptors.

Recently, a method closely related to PCA, Independent Component Analysis (ICA) [5], has received high attention. ICA can be viewed as a generalization of PCA, since it is concerned not only with second-order dependencies but also with high-order dependencies. PCA makes the data uncorrelated while ICA makes the data as independent as possible. Applications of ICA to feature extraction from images have been a topic of research interest. It is known that most of the ICA basis images extracted from natural images are sparse and similar to localized and oriented edges[6], which can capture important information for recognition. In addition, using the extracted ICA features for pattern discrimination has been studied for face recognition [7], texture segmentation [8], and object recognition [9][10]. These results indicate the features estimated by ICA are better than the features estimated by PCA with regard to recognition accuracy.

In pattern recognition problems, it is more desirable that extracted pattern features belonging to different classes are mutually separated as much as possible in the feature space. However ICA algorithm is categorized unsupervised learning, class information is not taken into consideration when feature extraction is carried out. Therefore, high separability of extracted features is not always ensured. To overcome this problem, Supervised ICA(SICA) was proposed [11]. The class separability of the SICA features is enhanced by maximizing the Mahalanobis distance between classes. The results suggest that SICA is more efficient than ICA for recognition.

ICA is based on a linear model, so it is inadequate for ICA to describe complex nonlinear variations in pattern recognition due to illumination changes, viewpoint changes and noise. In addition, there are other non-linear factors due to a camera's properties. One approach to solve this problem is kernel-based methods that as they are effective for such non-linearity. Kernel methods allow for the development of a non-linear extension of some linear algorithms, such as PCA and ICA. Recently, Kernel ICA (KICA)[12] was proposed as a nonlinear extension of ICA, which combines the a nonlinear kernel with ICA. KICA can improve the performance of ICA for pattern recognition.

In this paper we propose a supervised Kernel ICA for local image descriptors (SKICA descriptors), which combines nonlinear kernels with SICA. SKICA descriptors accept the input as normalised and gradient patches (the PCA-SIFT descriptor). In natural images, nearby pixels are statistically related and gradients based on edges are important in object recognition. Therefore, the gradient patches is the appropriate input for SKICA descriptors. Our work has two key contributions. First, we propose a novel dimension reduction method for local image descriptors that is a supervised nonlinear method. Second, we show that SKICA descriptors is better than KICA, SICA, ICA and PCA descriptors for object recognition on various sets of natural images.

The paper is organized as follows. Section 2 presents the SKICA Algorithm. Section 3 describes the SKICA-based local image descriptors (SKICA descriptors). Finally, section 4 shows the experimental results and section 5 summaries our conclusion.

2 Supervised Kernel ICA Algorithm

2.1 ICA and Supervised ICA

Spatial features in an image reflect that the value of one pixel depends on that of its neighbors. In real images, nearby pixels often have common causes and thus are statistically related. Researchers have analyzed the inter-relations between neighbor pixels to find new types of representations. For example, Bell and Sejnowski [6] applied ICA to input data consisting of image patches from natural images. ICA can be described as follow. First, assume we have a set of training images $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)^T$, where each vector $\mathbf{x}_i$ represents an image and the total number of training samples is N . The general model of ICA is:

$$\mathbf{X} = \mathbf{A}\mathbf{S} . \quad (1)$$

where $\mathbf{S} = (s_1, s_2, \dots, s_N)^T$ are the coefficients and each column of $\mathbf{A}$ is a basis image $\mathbf{a}_i$. The purpose of ICA is to seek mutually independent components s_i . The goal of ICA is to find a transformation matrix $\mathbf{W}_{ICA}$, $\mathbf{Y} = \mathbf{W}_{ICA}\mathbf{X}$, to minimize the statistical dependencies between the coefficients s_i . Several methods have been proposed for the $\mathbf{W}_{ICA}$ estimation. For example, Bell and Sejnowski developed an elegant learning algorithm based on information maximization. The learning algorithm for $\mathbf{W}_{ICA}$ can be summarized as:

$$\Delta \mathbf{W}_{ICA} = (\mathbf{I} + g(\mathbf{Y})\mathbf{Y}^T)\mathbf{W}_{ICA} . \quad (2)$$

where $g()$ is sigmoid function. Before the learning procedure, a preprocessing operation $\mathbf{V}_{PCA} = \mathbf{D}^{-1/2}\mathbf{E}^T$, whitens the data using PCA as is typical for most ICA learning algorithms. Here, $\mathbf{D}$ and $\mathbf{E}$ are the eigenvalues matrix and the eigenvectors matrix of the covariance matrix of $\mathbf{X}$, respectively.

Since ICA is an unsupervised learning, the features estimated by ICA are not always useful in recognition. To overcome this problem, Supervised ICA(SICA) based on Linear Discriminant Analysis (LDA) was proposed [11]. The purpose of LDA is to maximize the following objective:

$$J_{lm}(\mathbf{w}_i) = \frac{\mathbf{w}_i^T \mathbf{S}_B \mathbf{w}_i}{\mathbf{w}_i^T \mathbf{S}_W \mathbf{w}_i} . \quad (3)$$

where $\mathbf{S}_B = \sum_{c=l,m} N_c(\mu_c - \bar{\mathbf{x}})(\mu_c - \bar{\mathbf{x}})^T$ is the between class scatter matrix and $\mathbf{S}_W = \sum_{c=l,m} \sum_{i \in c} (\mathbf{x}_i - \mu_c)(\mathbf{x}_i - \mu_c)^T$ is the within class scatter matrix ($l, m = 1, \dots, C$). Here, $\mu_c = \frac{1}{N_c} \sum_{i \in c} \mathbf{x}_i$, $\bar{\mathbf{x}} = \frac{1}{C} \sum_i \mathbf{x}_i = \frac{1}{C} \sum_c N_c \mu_c$, N_c is the number of samples in class c and C is the number of classes, respectively. In SICA, an additional cost function is defined such that the weighted sum of Mahalanobis distances between two classes is maximized. More concretely, using the initial vector $\mathbf{w}_{i0}$ of $\mathbf{w}_i$, the mean features are first calculated from its projection values. Then, their Mahalanobis distances are obtained for all combinations of two classes. Here, the distance between mean features of class l

and m as $J_{lm}(\mathbf{w}_{i0})$ are used for the weights in the cost function. The weights $1/J_{lm}^2(\mathbf{w}_{i0})$ for the Mahalanobis distance of $J_{lm}(\mathbf{w}_i)$ are defined as follows:

$$d(\mathbf{w}_i) = \sum_{l=1}^{c-1} \sum_{m=l+1}^c \frac{1}{J_{lm}^2(\mathbf{w}_{i0})} J_{lm}(\mathbf{w}_i). \quad (4)$$

From this cost function, the following derivative ψ_i is obtained:

$$\psi_i = \frac{\partial d(\mathbf{w}_i)}{\partial \mathbf{w}_i} = \sum_{l=1}^{c-1} \sum_{m=l+1}^c \frac{1}{J_{lm}^2(\mathbf{w}_{i0})} \frac{\partial (J_{lm}(\mathbf{w}_i))}{\partial \mathbf{w}_i}. \quad (5)$$

Equation(5) is added to the ICA algorithm, and then the update formula of $\mathbf{W}_{SICA}$ in the proposed SICA is given as follow:

$$\Delta \mathbf{W}_{SICA} = (\mathbf{I} + g(\mathbf{Y})\mathbf{Y}^T) \mathbf{W}_{SICA} + \alpha \mathbf{\Psi}. \quad (6)$$

where $\mathbf{\Psi} = [\psi_1, \psi_2, \dots, \psi_N]^T$ and α is a positive constant.

2.2 Kernel ICA and Supervised Kernel ICA

Since ICA is based on a linear model, ICA can fail in the case of non-linearity. Recently, Liu et al. [12] proposed Kernel ICA which combines kernel functions with ICA. The basic idea of the KICA is to map the input data into an implicit feature space F using a kernel: $\Phi : \mathbf{x} \in \mathbb{R}^t \rightarrow \Phi(\mathbf{x}) \in F$, and then the ICA algorithm is performed in F to produce the nonlinear independent components of the input data. The input data $\mathbf{X}$ is whitened in feature space F , first. Similar to Kernel PCA (KPCA) [13], the whitening transform is to perform PCA in feature space F . The inner product of two vectors in F is calculated by a kernel function. In this paper, we use a Gaussian kernel $k(x, y) = \exp(-\frac{\|x-y\|^2}{2\delta^2})$. After the whitening transform, the ICA learning iteration algorithm described by Equation (2) is performed as follows:

$$\mathbf{Y}^\Phi = \mathbf{W}_{KICA}^\Phi (\mathbf{V}_{KPCA}^\Phi)^T \Phi(\mathbf{X}) = \mathbf{W}_{KICA}^\Phi (\mathbf{D}^\Phi)^{-1} (\mathbf{E}^\Phi)^T \mathbf{K}. \quad (7)$$

where $\mathbf{K}$ is defined by $\mathbf{K}_{ij} := (\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j))$, $(* \cdot *)$ denote inner product, $\mathbf{D}^\Phi$, $\mathbf{E}^\Phi$ are the Eigenvalues matrix and Eigenvectors matrix of $\mathbf{K}$. The learning algorithm for $\mathbf{W}_{KICA}^\Phi$ can be summarized as the following:

$$\Delta \mathbf{W}_{KICA}^\Phi = (\mathbf{I} + g(\mathbf{Y}^\Phi)(\mathbf{Y}^\Phi)^T) \mathbf{W}_{KICA}^\Phi. \quad (8)$$

According to the derived algorithm above, the new feature representation in feature space, $\mathbf{s}^\Phi$, of a test data $\mathbf{x}'$ can be computed:

$$\mathbf{s}^\Phi = \mathbf{W}_{KICA}^\Phi (\mathbf{V}_{KPCA}^\Phi)^T \Phi(\mathbf{x}') = \mathbf{W}_{KICA}^\Phi (\mathbf{D}^\Phi)^{-1} (\mathbf{E}^\Phi)^T \mathbf{K}(\mathbf{X}, \mathbf{x}'). \quad (9)$$

where $\mathbf{K}(\mathbf{X}, \mathbf{x}') = [k(\mathbf{x}_1, \mathbf{x}'), k(\mathbf{x}_2, \mathbf{x}'), \dots, k(\mathbf{x}_N, \mathbf{x}')]^T$, and k is a kernel function.

We propose Supervised Kernel ICA (SKICA) with a combination of KICA and Kernel Discriminant Analysis (KDA) [14]. As a nonlinear extension of LDA, KDA essentially performs LDA in the feature space F . The conventional between-class scatter operator $\mathbf{S}_B^\Phi$, within-class scatter operator $\mathbf{S}_W^\Phi$ can be expressed as: $\mathbf{S}_B^\Phi = \sum_{c=l,m} N_c (\mu_c^\Phi - \bar{\mathbf{x}}^\Phi)(\mu_c^\Phi - \bar{\mathbf{x}}^\Phi)^T$ and $\mathbf{S}_W^\Phi = \sum_{c=l,m} \sum_{i \in c} (\mathbf{x}_i^\Phi - \mu_c^\Phi)(\mathbf{x}_i^\Phi - \mu_c^\Phi)^T$, where $\mu_c^\Phi = \frac{1}{N_c} \sum_{i \in c} \Phi(\mathbf{x}_i)$, $\bar{\mathbf{x}}^\Phi = \frac{1}{C} \sum_i \Phi(\mathbf{x}_i)$. We maximize the Fisher criterion below to obtain the optimal projection directions $\mathbf{w}_i$ in F :

$$J_{lm}^\Phi(\mathbf{w}_i) = \frac{\mathbf{w}_i^T \mathbf{S}_B^\Phi \mathbf{w}_i}{\mathbf{w}_i^T \mathbf{S}_W^\Phi \mathbf{w}_i}. \quad (10)$$

The weights $1/J_{lm}^{\Phi 2}(\mathbf{w}_{i0})$ for the Mahalanobis distance of $J_{lm}^\Phi(\mathbf{w}_i)$ are defined as follows:

$$d^\Phi(\mathbf{w}_i) = \sum_{l=1}^{c-1} \sum_{m=l+1}^c \frac{1}{J_{lm}^{\Phi 2}(\mathbf{w}_{i0})} J_{lm}^\Phi(\mathbf{w}_i). \quad (11)$$

From this cost function, the following derivative ψ_i^Φ is obtained:

$$\psi_i^\Phi = \frac{\partial d^\Phi(\mathbf{w}_i)}{\partial \mathbf{w}_i} = \sum_{l=1}^{c-1} \sum_{m=l+1}^c \frac{1}{J_{lm}^{\Phi 2}(\mathbf{w}_{i0})} \frac{\partial (J_{lm}^\Phi(\mathbf{w}_i))}{\partial \mathbf{w}_i}. \quad (12)$$

Equation (12) is added to the KICA algorithm, and then the update formula of $\mathbf{W}_{SKICA}^\Phi$ in the proposed SKICA is given as follow:

$$\Delta \mathbf{W}_{SKICA}^\Phi = (\mathbf{I} + g(\mathbf{Y}^\Phi)(\mathbf{Y}^\Phi)^T) \mathbf{W}_{SKICA}^\Phi + \alpha \Psi^\Phi. \quad (13)$$

where $\Psi^\Phi = [\psi_1^\Phi, \psi_2^\Phi, \dots, \psi_N^\Phi]^T$ and α is a positive constant.

3 Supervised Kernel ICA-Based Local Image Descriptors

SKICA is fundamentally a statistical model for natural images that models images as nonlinear superpositions of basis images, with non-Gaussian, independent weighting coefficients optimized for class separability. Here, we argue that the properties of SKICA applied to the local image descriptors retains the advantages of the reduced dimensionality but improves recognition accuracy due to the nature of the SKICA representation. We call SKICA applied to the local image descriptors, SKICA descriptors. To train SKICA, the image patches at the interest points are collected from a common image database. In this paper, we use the normalised patches and the gradient patches (PCA-SIFT descriptors) as the input data and detect the interest points by Difference-of-Gaussians (SIFT detector). SKICA are applied to the input data and the result is used as the projection matrix for SKICA descriptors. This process is done offline. The Euclidean distance between the feature vectors projected by the projection matrix is used for matching. An important and related advantage of using KPCA is that it reduces the dimensionality of the input data prior to applying SKICA. In

Fig. 1. A block diagram for the processing flows in both the learning phase and the running phase of SKICA

Fig. 2. Basis images of natural images

the KPCA-style global filters, noise is automatically associated with the filters with high spatial frequency selectivity whose eigenvectors have small eigenvalues. Thus, KPCA also provides some benefits in reducing the high frequency noise in the descriptors prior to the application of SKICA. A block diagram for the whole process is given in Fig. 1.

In Fig. 2, we show the SKICA basis patches, the ICA basis patches and the PCA basis patches obtained from the normalized image patches in the natural scene data. The PCA-basis consists of global features like 2D Fourier bases. On the other hand, the ICA-basis and the SKICA-basis consist of local features like a sparse coding [15]. Gradient patches are based on the local information such as edges and lines. These edges and lines are both abundant and important in object recognition. Using this information, SKICA can produce nonlinear independent image bases that emphasize important edge information for pattern recognition.

Natural scenes contain many higher-order forms of statistical structure, and they form an extremely non-gaussian distribution that is not at all well captured by orthogonal components. Moreover natural images contain localized, oriented structures with limited phase alignment across spatial frequency. Lines and edges, especially curved and fractal-like edges, cannot be characterized by

linear pairwise statistics. That is to say, higher order dependencies in an image include nonlinear relations among the pixel intensity values, such as the relationships among three or more pixels in an edge or a curve. The features estimated by ICA resemble sparse coding which all have the important property of being localized in frequency and share the spatially localized properties of simple cells in the primary visual cortex [16]. The ICA model based on the receptive fields of simple cells can be described using linear models. Simple cell responses are certainly not completely linear because they show significant nonlinearities such as rectification and response saturation. It is inadequate for ICA to describe complex nonlinear variations in pattern recognition due to illumination changes, viewpoint changes and noise. In addition, there are other non-linear factors specific to image capture due to a camera's lens, shutter, CCD and Gamma correction. The SKICA representation can take into account non-linear higher-order statistical dependences in the data since it has enough information that can be obtained through supervised learning with nonlinear models. Therefore, the SKICA features produce sets of visual filters that are localised and have non-linear independence and high class separability.

4 Experimental Results

We tested SKICA descriptors based on normalised patches and normalised gradient patches for object recognition. We compare SKICA descriptors to KICA, SICA, ICA, PCA and SIFT descriptors. The experiments are performed over a set of classes provided by Caltech101 [17]¹: cars side(720 images), cars rear(651 images), motorbikes(826 images), airplanes(1074 images) and faces(450 images). We used 30 images per object category for training and the rest for testing. To reduce the bias of the number of test images in each category, the mean of the recognition rate of each category is used. This test is repeated 3 times with different training data sets, and the mean recognition rate of 3 runs is used as a final result. All input patches are 32×32 grayscale. Some parameters for the SKICA algorithm were selected by experiment.

4.1 Object Recognition

The object recognition problem from the database is formulated as follows. We compare the images in the test set to all reference images in the training set by matching their respective interest points. Then, we choose the object shown on the reference image with the highest number of matches with respect to the test image as the recognized object. For the matching, an interest point in the test image is compared to an interest point in the reference image by calculating the Euclidean distance between their descriptor vectors. A matching pair is detected if its distance is close enough to be considered a match.

The recognition performance is presented in Fig. 3. Note that the recognition rate of SKICA descriptors was higher than that of KICA, SICA, ICA and PCA

¹ Datasets are available at: <http://www.robots.ox.ac.uk/~vgg/data/data-cats.html>

Fig. 3. Recognition performance: (a) normalised patches, (b) gradient patches

descriptors. The best recognition rate was approximately 87% for SKICA, 84% for KICA, 83% for SICA, 81% for ICA, 79% for PCA and 86% for SIFT. Table 1 shows the confusion matrix for the different methods (gradient patches, 30 dimensions). It turns out from these results that the SKICA representation provides more discrimination power than other representations for object recognition.

4.2 Kurtosis and Class Separability of Extracted Features

In order to examine the independence (sparseness) of features, we evaluate the kurtosis of extracted features, $\mathbf{s} = [s_1, s_2, \dots, s_N]$. The kurtosis is defined as follows:

$$kurtosis = \frac{\sum_i (s_i - \bar{s})^4}{\left(\sum_i (s_i - \bar{s})^2 \right)^2} - 3. \quad (14)$$

Table 2 shows that the kurtosis of KICA features was the highest. The reason why the kurtosis of SKICA features becomes smaller is that the maximization of class separability as well as the maximization of independence are carried out.

We calculated values of the cost function in Eq.(11) in order to estimate the class separability. Table 3 shows that the highest class separability is obtained for the features extracted by SKICA as expected. From these results, increasing both the independence of features and the class separability has good influence on the recognition performance.

4.3 Matching of Extracted Features

Fig. 4 shows the result of matching using SKICA, KICA, SICA, ICA, PCA (gradient patches, 30 dimensions) and SIFT on the motorbike images taken from

Table 1. Confusion matrix(%)

SKICA	Car-side	Car-rear	Motobikes	Airplanes	Face
Car-side	88.2	0.9	4.2	3.5	3.2
Car-rear	0.7	86.9	3.5	2	6.9
Motobikes	1.6	5.1	83.3	5.0	5.0
Airplanes	0.4	0.7	8.0	86.6	4.3
Face	0.5	3.8	8.0	2.2	85.5

KICA	Car-side	Car-rear	Motobikes	Airplanes	Face
Car-side	84.5	1.9	5.0	4.4	4.2
Car-rear	1.5	83.2	4.5	3.0	7.8
Motobikes	2.5	6.0	79.6	6.0	5.9
Airplanes	1.5	2.0	8.5	83.0	5.0
Face	1.5	5.0	9.0	3.2	81.3

SICA	Car-side	Car-rear	Motobikes	Airplanes	Face
Car-side	83.4	2.1	5.5	4.5	4.5
Car-rear	1.9	82.1	4.8	3.2	8.0
Motobikes	2.9	6.3	78.5	6.2	6.1
Airplanes	1.7	2.0	9.1	81.9	5.3
Face	1.9	5.3	9.2	3.4	80.2

ICA	Car-side	Car-rear	Motobikes	Airplanes	Face
Car-side	80.2	3.1	6.1	5.4	5.2
Car-rear	2.6	79.2	5.5	4.0	8.7
Motobikes	3.7	7.0	75.3	7.0	7.0
Airplanes	2.5	2.8	10.0	78.6	6.1
Face	2.7	6.0	10.1	4.0	77.2

PCA	Car-side	Car-rear	Motobikes	Airplanes	Face
Car-side	74.0	5.3	7.5	6.8	6.4
Car-rear	4.0	72.8	6.7	5.3	11.2
Motobikes	6.0	8.3	69.1	8.4	8.2
Airplanes	4.0	4.1	11.3	72.1	8.5
Face	4.0	7.2	12.5	5.4	70.9

SIFT	Car-side	Car-rear	Motobikes	Airplanes	Face
Car-side	87.7	1	4.4	3.6	3.3
Car-rear	0.8	86.4	3.7	2	7.1
Motobikes	1.7	5.2	82.8	5.1	5.2
Airplanes	0.5	0.8	8.1	86.1	4.5
Face	0.7	4	8.3	2.5	84.5

different viewpoints. We manually set the thresholds to have each algorithm return 10 matches for comparison. SKICA descriptors correctly match the features near the edges of the motorbike.

Table 2. Kurtosis of extracted features for training samples

	normalised patches	gradient patches
SKICA	32.18	38.24
KICA	33.87	39.65
SICA	30.24	36.18
ICA	31.68	37.62
PCA	6.82	9.51

Table 3. Class separability of extracted features for training samples

	normalised patches	gradient patches
SKICA	56.34	60.57
KICA	53.72	56.83
SICA	51.41	55.04
ICA	48.32	51.46
PCA	44.27	49.36

Fig. 4. The results of matching. The top ten matches are shown for each algorithm: solid lines denote correct matches while dotted lines show incorrect ones.

5 Conclusion

We have proposed a new discriminative method for local image descriptors. We have shown that the SKICA descriptors perform better than KICA, SICA, ICA and PCA descriptors for object recognition, since the SKICA-based representation is localised, has non-linear independence and high class separability. However SKICA involves a kernel function selecting problem. Selecting an appropriate kernel function for a particular application area can be difficult and remains largely an unresolved issue. We are conducting further investigations extensions of SKICA-based representation to color image patches and other descriptor algorithms.

References

1. Fergus, R., Fei-Fei, L., Perona, P., Zisserman, A.: Learning object categories from google's image search. In: Proc. of IEEE International Conference on Computer Vision., vol. 2, pp. 1816–1823 (2005)
2. Sivic, J., Zisserman, A.: Video google: A text retrieval approach to object matching in videos. In: Proc. of IEEE International Conference on Computer Vision., vol. 2, pp. 1470–1477 (2003)
3. Lowe, D.G.: Distinctive image features from scale invariant keypoints. *International Journal of Computer Vision* 60(2), 91–110 (2004)
4. Ke, Y., Sukthankar, R.: PCA-SIFT: A more distinctive representation for local image descriptors. In: Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition., vol. 2, pp. 506–513 (2004)
5. Hyvarinen, A.: Survey on independent component analysis. *Neural Computing Surveys* 2, 94–128 (1999)
6. Bell, A.J., Sejnowski, T.J.: The independent components of natural scenes are edge filters. *Vision Research* 37(23), 3327–3338 (1997)
7. Bartlett, M.S., Movellan, J.R., Sejnowski, T.J.: Face recognition by independent component analysis. *IEEE Transactions on Neural Networks* 13(6), 1450–1464 (2002)
8. Jenssen, R., Eltoft, T.: Independent component analysis for texture segmentation. *Journal of Pattern Recognition Society* 36(10), 2301–2315 (2003)
9. Bressan, M., Guillaumet, D., Vitria, J.: Using an ICA representation of local color histograms for object recognition. *Journal of Pattern Recognition Society* 36(3), 691–701 (2003)
10. Sahambi, H.S., Khorasani, K.: A neural-network appearance-based 3-D object recognition using independent component analysis. *IEEE Transactions on Neural Networks* 14(1), 138–149 (2003)
11. Sakaguchi, Y., Ozawa, S., Kotani, M.: Feature Extraction Using Supervised Independent Component Analysis by Maximizing Class Distance. *IEEJ Transactions on Electronics, Information and Systems* 124(1), 157–163 (2004) (in Japanese)
12. Liu, Q., Cheng, J., Lu, H., Ma, S.: Modeling Face Appearance With Nonlinear Independent Component Analysis. In: Proc. of IEEE International Conference on Automatic Face and Gesture Recognition, pp. 761–766 (2004)
13. Schalkopf, B., Smola, A., Muller, K.: Nonlinear Component Analysis as a Kernel Eigenvalue Problem. *Neural Computation* 10(5), 1299–1319 (1998)

14. Baudat, G., Anouar, F.: Generalized discriminant analysis using a kernel approach. *Neural Computation* 12(10), 2385–2404 (2000)
15. Olshausen, B.A., Field, D.J.: Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381, 607–609 (1996)
16. van Hateren, J.H., van der Schaaf, A.: Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc. of Royal Society London, ser. B.* 265, 359–366 (1998)
17. Fei-Fei, L., Fergus, R., Perona, P.: Learning generative visual models from few training examples: an incremental bayesian approach testing on 101 object categories. In: *Proc. of IEEE CVPR Workshop of Generative Model Based Vision*, pp. 178–187 (2004)