

PAPER

Color Independent Components Based SIFT Descriptors for Object/Scene Classification

Dan-ni AI^{†a)}, Student Member, Xian-hua HAN[†], Xiang RUAN^{††}, Nonmembers, and Yen-wei CHEN[†], Member

SUMMARY In this paper, we present a novel color independent components based SIFT descriptor (termed CIC-SIFT) for object/scene classification. We first learn an efficient color transformation matrix based on independent component analysis (ICA), which is adaptive to each category in a database. The ICA-based color transformation can enhance contrast between the objects and the background in an image. Then we compute CIC-SIFT descriptors over all three transformed color independent components. Since the ICA-based color transformation can boost the objects and suppress the background, the proposed CIC-SIFT can extract more effective and discriminative local features for object/scene classification. The comparison is performed among seven SIFT descriptors, and the experimental classification results show that our proposed CIC-SIFT is superior to other conventional SIFT descriptors.

key words: CIC-SIFT descriptor, object/scene classification, ICA-based transformation

1. Introduction

Object/scene classification, which is important for concept detection or metadata annotation, is one of the most challenging problems in computer vision and pattern recognition.

Feature extraction, as one of the most important steps, can capture a certain visual property of an image. There are two types of image features: global features (color, texture, and shape) for the entire image and local features for a small group of pixels [8]. Recently, local features have been commonly employed in real-world applications. Typically, SIFT (Scale-Invariant Feature Transform) descriptor, which is a local feature proposed by Lowe [2], has been widely used for object/scene classification. The SIFT descriptors can extract distinctive features from an image and describe the local shape using edge gradient histograms. The main drawback of the standard SIFT descriptor is that light color changes are ignored [1], since only gray-scale intensities are used for computation. Thus, several color SIFT descriptors have been proposed recently by combining color information with local SIFT descriptors, such as RGB-SIFT, HSV-SIFT [3], HueSIFT [9], W-SIFT [1], rgSIFT [1], Transformed color SIFT [1], CSIFT [18] and Opponent-SIFT [1]. What they all have in common is that they first transform an RGB color image into other color space

with physical model-based transformations. Physical model based transformations, such as opponent transformation, are very useful color transformations for the representation of some physical properties. Koen E.A. van de Sande et al. presented a comparative study of several global and local descriptors including histograms (RGB, Opponent, Hue, rg, transformed color), color moments, moment invariants [20], standard SIFT and the color SIFT represented previously. Their experiments showed that the accuracy of ranked category classification results of the color SIFT descriptors was much better than those of other descriptors.

The existing color transformations, mentioned above, can be considered as generic or physical model-based methods. They are defined uniquely and are not dependent on input images. However, in some practical situations, training set has been already given in advance. Therefore, it is much better to take into account the object specific knowledge included in the training set, and impose it to SIFT descriptors.

Interest is growing in developing image coding (spatial and spectra) by statistical analysis. These approaches follow the idea of Barlow that the goal of vision information processing is to transform input signals to reduce redundancy between inputs [16]. In an analysis of color information, Buchsbaum et al. found that the use of opposing coding is the most efficient way to encode human photoreceptor signals [13]. By analysing spectral distributions of natural images using principal component analysis (PCA), Rudernam et al. [14] found that the principal components were close to the opposing coding proposed by Buchsbaum et al. [13]. Rudernam also pointed out that principal components depend on experimental data and are not uniquely defined. Since PCA uses only second-order statistics for decorrelation solution, methods involving higher-order statistics such as independent component analysis (ICA) are attracting increased attention recently [4], [12], [15]. In ICA, transformed independent components are non-correlated and as statistically independent of each other as possible. In our previous work, we have shown that among the three color independent components obtained from RGB color space, two are in an opposing-color model by which the responses of R, G and B cones are combined in opposing fashions. This coincides with the idea of contrasting reflected in many color systems [10]. Meanwhile, it has been proven that independent component analysis can be applied to images for enhancing the contrast of different objects [26].

This paper describes a novel color independent components based SIFT descriptor (termed CIC-SIFT). Our

Manuscript received October 14, 2009.

Manuscript revised March 31, 2010.

[†]The authors are with Ritsumeikan University, Kusatsu-shi, 525-8577 Japan.

^{††}The author is with Omron Corporation, Kusatsu-shi, 525-0035 Japan.

a) E-mail: gr041083@ed.risumei.ac.jp

DOI: 10.1587/transinf.E93.D.2577

proposed CIC-SIFT can be divided into two steps. We first find an adaptive transformation using independent component analysis based on the learning rule of Bell and Sejnowski [12]. Then we compute CIC-SIFT descriptors over all three transformed color independent components (channels). Compared to conventional physical model-based color transformations, our proposed ICA-based transformation can be considered as an object-specific method which needs training set. It can enhance contrast between object and background in an image. Consequently, the proposed CIC-SIFT can extract more effective and discriminative local features for object/scene classification. On the other hand, CIC-SIFT is a learning based object-specific method, so it needs an extra learning phase to obtain the transformation matrix. However, all the learning phase is carried out offline, and once the transformation matrix is obtained, there are no significant differences between our proposed method and the conventional physical model-based methods for object/scene classification.

ICA or SKICA is applied to SIFT features for dimension reduction. [24], [25]. While, in our paper, ICA is used to find an efficient color space for each category. And then, the SIFT features are extracted in this new transformed color space.

This paper is organized as follows. In Sect. 2, six conventional SIFT descriptors are summarized briefly. In Sect. 3, we review the theory of independent component analysis and present a novel color independent components based SIFT descriptor (CIC-SIFT). The detailed experimental results are shown in Sect. 4. Finally, we conclude the paper in Sect. 5.

2. Conventional Color SIFT Descriptors

To date, SIFT descriptors are improved by combining the color information for object/scene classification. Several color SIFT descriptors have been proposed by researchers based on conventional color space transformations, such as RGB-SIFT, Transformed color SIFT, HSV-SIFT, CSIFT and Opponent-SIFT.

2.1 SIFT

The SIFT descriptor, proposed by Lowe, extracts the local features that are invariant to the image scale, rotation, and viewpoint [2], and describes the local shape using orientation histograms. The traditional SIFT descriptor is only used for gray-scale image. A color image is first transformed into a gray-scale image and then SIFT descriptors are calculated over the obtained gray-scale image. Usually, a 4×4 array of histograms with 8 orientation bins is used to form the SIFT descriptor for one key point. Therefore, a typical dimensionality of conventional SIFT feature vectors is 128 ($4 \times 4 \times 8$).

2.2 RGB-SIFT

A pixel of a color image is usually given as three intensities

in R, G and B channels. RGB-SIFT descriptors are proposed by combining RGB color information with conventional local SIFT descriptors. 384-dimensional RGB-SIFT descriptors can be obtained by arranging three SIFT descriptors together, which is shown in Eq. (1).

$$\mathbf{D} = [h_{R_1} h_{R_2} \cdots h_{R_{128}} h_{G_1} h_{G_2} \cdots h_{G_{128}} h_{B_1} h_{B_2} \cdots h_{B_{128}}]^T \quad (1)$$

where \mathbf{D} denotes the RGB-SIFT descriptor of a key point, and h_{R_n} , h_{G_n} , h_{B_n} ($n = 1, 2, \dots, 128$) denote orientation bins of a key point for each channel, respectively.

2.3 Transformed Color SIFT

A transformed color SIFT is proposed by van de Sande et al. [1], in which the pixel value distributions are normalized independently as Eq. (2) shows.

$$\begin{bmatrix} R_t & G_t & B_t \end{bmatrix}^T = \begin{bmatrix} \frac{R - \mu_R}{\sigma_R} & \frac{G - \mu_G}{\sigma_G} & \frac{B - \mu_B}{\sigma_B} \end{bmatrix}^T \quad (2)$$

Transformed color SIFT descriptors are scale-invariant, shift-invariant and invariant to light color changes and shift. It is calculated in the transformed color space. Three 128-dimensional descriptors in each channel of the same position are combined to represent a feature vector with a dimension of 384 like RGB-SIFT (3 channels \times 128-dimensional orientation bins).

2.4 HSV-SIFT

HSV-SIFT is given by Bosch et al. [3] to calculate the descriptors over three channels in HSV color space. Since the three channels of HSV color space are not correlated and HSV color space is similar to the human cognitive system, HSV color space could give more information than RGB color space. The transformation of HSV is shown in the following equations.

$$H = \begin{cases} 0, & \text{if } \max = \min \\ \left(60^\circ \times \frac{G-B}{\max-\min} + 360^\circ\right) \bmod 360^\circ, & \text{if } \max = R \\ 60^\circ \times \frac{B-R}{\max-\min} + 120^\circ, & \text{if } \max = G \\ 60^\circ \times \frac{R-G}{\max-\min} + 240^\circ, & \text{if } \max = B \end{cases} \quad (3)$$

$$S = \begin{cases} 0, & \text{if } \max = 0 \\ \frac{\max-\min}{\max} = 1 - \frac{\min}{\max}, & \text{otherwise} \end{cases}$$

$$V = \max$$

Each key point of HSV-SIFT has 384 dimensions.

2.5 CSIFT

Colored SIFT (CSIFT), proposed by Abdel-Hakim et al., does not only embed the color information in the descriptors, but also in respect to the geometrical invariance and color invariance [18]. The Gaussian color model is used to calculate geometrical invariance and color invariance from the RGB color space. It can be shown as Eq. (4) [19]

$$\begin{bmatrix} GC1 \\ GC2 \\ GC3 \end{bmatrix} = \begin{bmatrix} 0.06 & 0.63 & 0.27 \\ 0.3 & 0.04 & -0.35 \\ 0.34 & -0.6 & 0.17 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (4)$$

CSIFT descriptors are calculated from this working space as RGB-SIFT descriptors are.

2.6 Opponent-SIFT

In the Opponent-SIFT, an input color image is converted from RGB color space to opponent color space by using Eq. (5) [1]:

$$\begin{bmatrix} O1 \\ O2 \\ O3 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & -\sqrt{6}/3 \\ 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (5)$$

The third component O3 is a summation of RGB representing intensity, while the first and second components O1 and O2 are opposing coding, which represent color information. The opposing coding is considered as the most efficient way to encode human photoreceptor signals [13]. The Opponent-SIFT descriptor is a combination of three 128-dimension descriptors calculated from each channel of the opponent color space. Consequently, the dimension of the Opponent SIFT descriptor is also 384 similar to RGB-SIFT. As Eq. (5) shows, the transformation matrix of opponent-transform is defined uniquely and doesn't depend on the input images. Thus, it is not adaptive for each category. It motivates us to learn an efficient and suitable transformation matrix (color coding) for each category in the following section.

3. Color Independent Components Based SIFT Descriptors (CIC-SIFT)

3.1 Independent Component Analysis

Independent Component Analysis (ICA) is a linear transformation that transforms a set of random data to be statistically independent of each other [4]. The application of ICA for object/scene classification in our study can be briefly described as follows.

Let us assume that components in RGB color space are represented as a vector $\mathbf{x}_{RGB} = [x_R, x_G, x_B]^T$. For simplicity's sake, it is also assumed that both the mixture variables and the independent components have a zero mean. Therefore, we can do some pre-processing before performing ICA. We calculate the mean value of all pixels in each component: $mean(\mathbf{x}_R)$, $mean(\mathbf{x}_G)$, $mean(\mathbf{x}_B)$. After pre-processing, the mean value of every element of \mathbf{x}_{RGB} is subtracted, which can be represented as a vector $\mathbf{x} = [x_1, x_2, x_3]^T = [(x_R - mean(\mathbf{x}_R)), (x_G - mean(\mathbf{x}_G)), (x_B - mean(\mathbf{x}_B))]^T$. Here, \mathbf{x} is considered a combination of three independent source components $\mathbf{s} = [s_1, s_2, s_3]^T$ with the 3×3 mixing matrix \mathbf{A} :

$$\mathbf{x} = \mathbf{A}\mathbf{s} \quad (6)$$

The goal of ICA is to find the transformation matrix \mathbf{W} , so that each element of the resulting vector \mathbf{y} becomes as independent as possible.

$$\mathbf{y} = [y_1, y_2, y_3]^T = \mathbf{W}\mathbf{x} \quad (7)$$

Bell and Sejnowski have proposed a neural learning algorithm for ICA [12]. The approach is to maximize joint entropy by utilizing the stochastic gradient ascent. The updating formula for \mathbf{W} is:

$$\mathbf{W}^{(t+1)} = \mathbf{W}^{(t)} + \mu \left[\mathbf{I} - \phi(\mathbf{y}^{(t)}) (\mathbf{y}^{(t)})^T \right] \mathbf{W} \quad (8)$$

where μ is a learning coefficient, and t is an iteration number, and $\phi(y) = 1 - 2/(1 + e^{-y})$ is calculated for each component of \mathbf{y} . A whitening technique was used to accelerate this computation [17]. If the iteration converges, \mathbf{y} is considered to be equivalent to \mathbf{s} except in scale and permutation.

3.2 Color Independent Components Based SIFT Descriptors (CIC-SIFT)

Color independent components based SIFT descriptors (termed CIC-SIFT) are proposed to extract image features in the ICA-based color space. It is a two-step feature extraction method (see Fig. 1). In the first learning step, we learn a transformation matrix for each category using ICA. The second step is the color transformation step. We transform original images components (x_R, x_G, x_B) into three independent components using the ICA-based transformation matrix. And then, the transformed color independent components are utilized to extract color SIFT features for object/scene classification (CIC-SIFT descriptor).

Assuming that there are N categories in a database and the information of each image is carried by RGB components. For ICA transformation matrix learning, M images are chosen randomly as training samples from each category. The size of the sample image is $h \times w$. Since each pixel of an image is represented by a 3×1 vector $\mathbf{x}_{RGB} = [x_R, x_G, x_B]^T$, all the sample images can be concatenated together into a $3 \times (h \times w \times M)$ matrix \mathbf{x} . With the sample matrix \mathbf{x} , a 3×3 transformation matrix \mathbf{W} can be calculated by using the learning algorithm proposed by Bell and Sejnowski [12]. Therefore, from categories in the database, N adaptive transformation matrices \mathbf{W}_i ($i \in [1, 2, \dots, N]$) corresponding to each category can be learned. Then, we transform the original components (R, G, B) of training images into three independent components using the corresponding adaptive transformation matrix \mathbf{W}_i , which can be shown in the following equation

$$[x_{IC1}, x_{IC2}, x_{IC3}]^T = \mathbf{W}_i \times [x_R, x_G, x_B]^T \quad (9)$$

where vector $[x_{IC1}, x_{IC2}, x_{IC3}]^T$ represents color independent components which can be used as the new color representation for the color SIFT descriptors.

The CIC-SIFT descriptor is calculated in color independent components space. We combine three 128-dimension descriptors in each component with the same

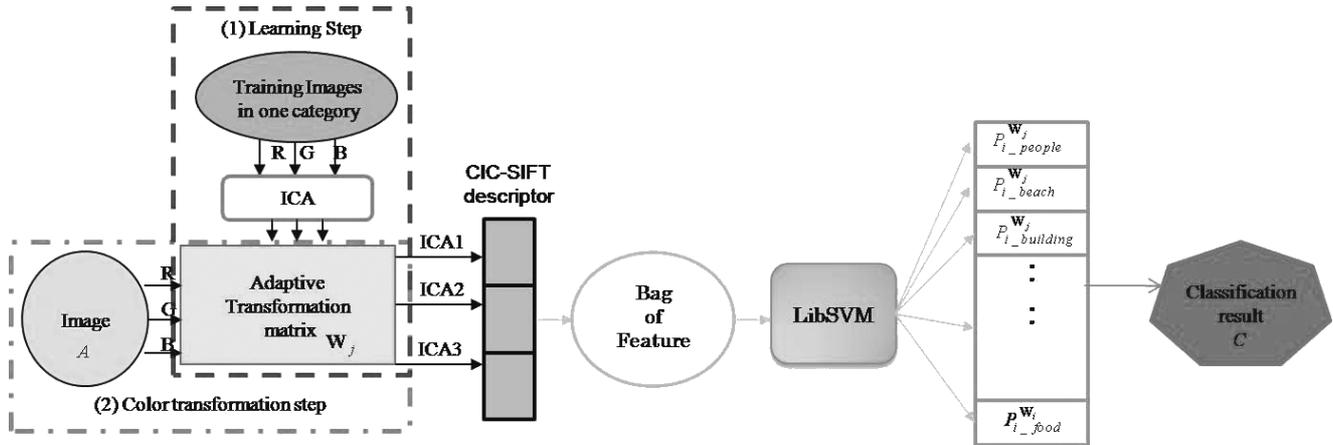


Fig. 1 Flowchart of object/scene classification by using CIC-SIFT descriptors.

key point to represent a 384 dimension feature vector (3 components \times 128-dimensional orientation bins).

4. Object/Scene Classification Experiments

The effectiveness of our proposed CIC-SIFT descriptor is discussed by comparative experiments with several other SIFT descriptors which are commonly used in object/scene classification. These comparative experiments can be carried out in terms of the behaviour of the transformation and the classification results.

4.1 Experimental Setup

In this section, three diverse databases are used in our experiments to evaluate the performance of six SIFT descriptors (mentioned previously): a SIMPLICITY image database [5], an object database composed of ten categories (nine categories were collected from the website[†] and the other one was used by R. Fergus et al. [21]) and our own scene database with data collected from internet. Since gray-scale images are included in the object database and the scene database, we transformed a single channel of each gray-scale image into three channels of color images. The values of three channels are the same ($x_R = x_G = x_B$). The classification process is summarized in Fig. 1. By using the ICA method, an adaptive transformation matrix is first learned from training datasets of each category. Then, the transformation matrix is utilized to transform the original color component (x_R, x_G, x_B) into three color independent components for all images in the database. We detect the key points by a regular grid. In each independent component image, 32×32 pixels overlapping grid is used as a key point. Each key point in each channel is represented by a 128-dimension SIFT descriptor. We combine the SIFT descriptors of the corresponding key point in each channel into a 384 dimension feature vector (128×3).

After SIFT descriptors have been calculated, the orderless bag-of-feature [6] is used as a feature representation. We build a visual vocabulary by clustering the fea-

ture vectors (descriptors) from the training set and then represent each image in the data set as a histogram of visual words drawn from the vocabulary. In our experiments, k-means [23] is used as a clustering method and the number of clusters (visual words) is 100 for the SIMPLICITY database and 300 for the object database and the scene database. We employed SVMs as a classifier, which has been widely used and shown to be efficient for object/scene classification [7]. In our experiments, LIBSVM package has been employed [11]. For fair comparison, we assign the optimum parameter (the value of gamma is set to 0.3–0.8) of radial basis function kernel in SVMs for each SIFT descriptor.

All the experiments are implemented on a standard PC (Intel(R) Core (TM) 2, 1.86 GHZ and 3.00 GB RAM), and the learning phase for getting the transformation matrix is carried out with Matlab 2007.

What we should mention is that the test images need to be transformed by all learned transformation matrices since we do not really know which category they belong to in advance. Therefore, for each test image (A), we can obtain N^2 classification probabilities (N categories \times N transformation matrices). The test image (A) is allocated to the category (C) that has the highest value of probabilities, as shown in Eq. (10):

$$C(A) = \arg \max_i (P_{W_1}(A, i), P_{W_2}(A, i), \dots, P_{W_N}(A, i)) \quad (10)$$

where $i = 1, 2, \dots, N$ is the category label; W_j ($j = 1, 2, \dots, N$) are transformation matrices learned from each category; $P(x)$ denotes classification probabilities of a test image allocated into each different categories.

4.2 Experimental Results

4.2.1 Color Space Transformation

Compared with the conventional color transformations, which are uniquely defined, ICA can learn adaptive transformation (color coding) from samples statistically.

[†]<http://www.robots.ox.ac.uk/~vggg/data/data-cats.html>

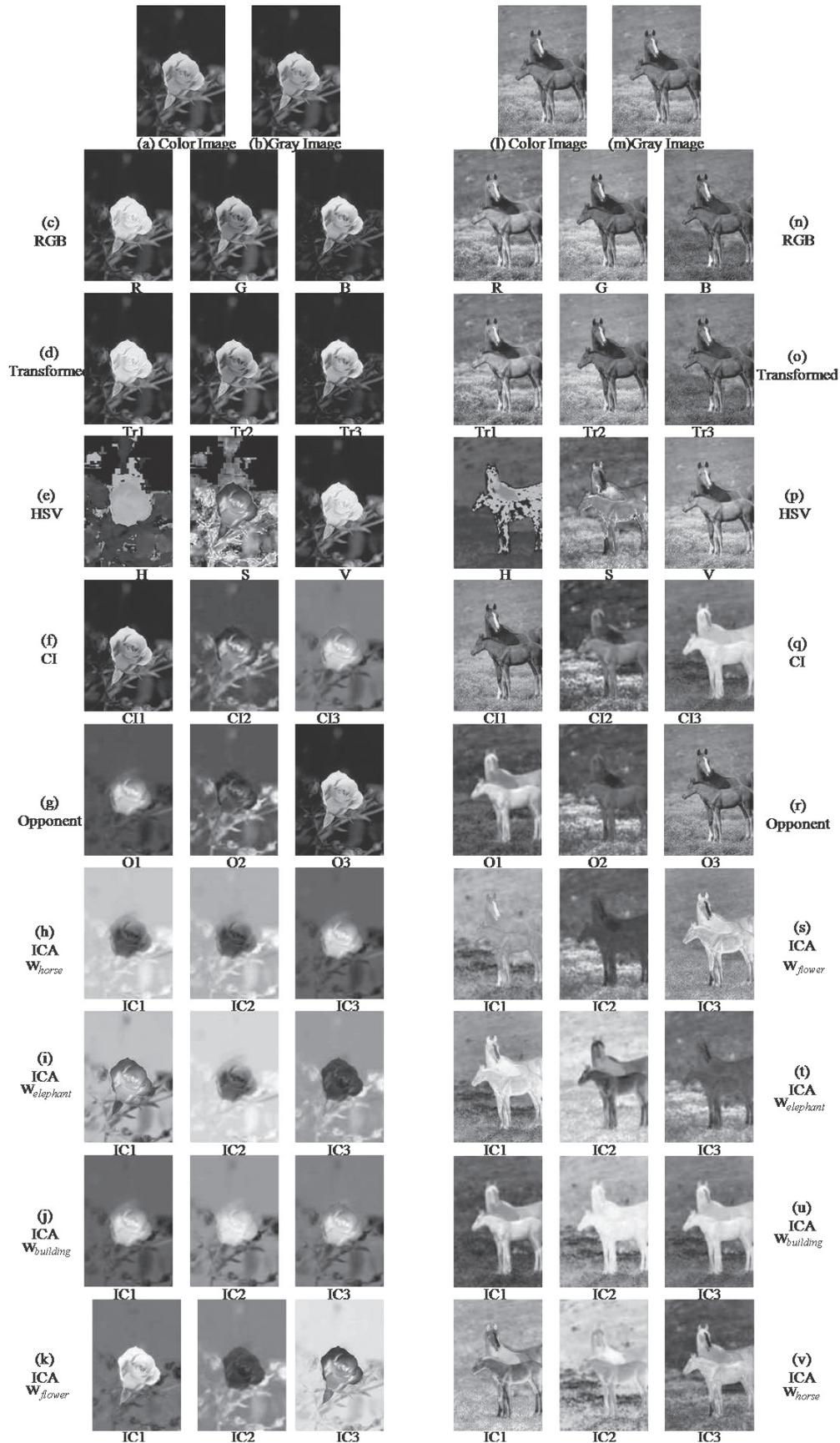


Fig. 2 Examples of the flower and horse categories and their transformed images.

In order to discuss the effectiveness of ICA-based color transformation, we transformed original images into seven different color spaces (Gray space, RGB color space, Transformed color space, HSV color space, Color Invariant space, Opponent color space and ICA-based color space).

Two sample images in the flower category and the horse category are shown in Fig. 2, which includes seven different transformed images. As we can see from Fig. 2 (c) and Fig. 2 (n), both object (flower/horse) and background (leaves/prairie) appear in original R, G and B channels clearly. They seem to be similar compared to those obtained by transformed color transformation (Fig. 2 (d) and Fig. 2 (o)) which is normalized from RGB color space (Eq. (2)). The object of the hue-channel in HSV color space is difficult to distinguish and the saturation and value channel do not significantly enhance the contrast (Fig. 2 (e) and Fig. 2 (p)). We can not distinguish the object from the background in Color Invariant space clearly (Fig. 2 (f) and Fig. 2 (q)), which is an approximation for the human vision system [19]. In opponent color space, the summation component O3 that corresponds to illumination can not enhance the contrast between the object and the background, while the contrast can be enhanced in opponent components O1 and O2 (Fig. 2 (g) and Fig. 2 (r)). However, since the transformation matrix, which is defined uniquely, is not a suitable one for the flower or horse, the boundary between the object and the background is not clear.

Images transformed by ICA-based transformation matrices are shown in Fig. 2 (h), Fig. 2 (s), Fig. 2 (k) and Fig. 2 (v). For the flower and horse categories, the transformation matrix in Eq. (9) can be denoted using $\mathbf{W}_{ICA_{flower}}$ and $\mathbf{W}_{ICA_{horse}}$, respectively. They are shown in Eq. (11) and Eq. (12):

$$\mathbf{W}_{ICA_{flower}} = \begin{bmatrix} -0.0446 & 0.8608 & -0.7229 \\ 0.0214 & -1.1898 & 0.9565 \\ -0.2962 & -0.4821 & 1.3432 \end{bmatrix} \quad (11)$$

$$\mathbf{W}_{ICA_{horse}} = \begin{bmatrix} -0.0250 & 0.0086 & -0.1445 \\ -0.1057 & -0.1061 & 0.1843 \\ 0.4650 & -0.3021 & -0.1404 \end{bmatrix} \quad (12)$$

Two rows of the transformation reflect an opposing-color model by combining R, G and B into opposing color components. This is similar to the idea in many color transformations. The interesting point is that there is no summation component, which usually appears in many transformations.

Our basic idea is to enhance the contrast between the object and background by extracting independent components. Toshiharu Nakai has shown that independent component analysis (ICA) can be applied to images for enhancing the contrast of different objects [26]. The keypoint of ICA is that we need use similar images (the flower), which have similar statistical properties, as training samples to learn the transformation matrix (the basis). While if we use different category images, such as horse images, to extract the basis, we just obtain the adaptive transformation matrix of horse

images rather than that of flower images. Therefore, when we project the flower images on such basis, we could not get exact independent components for the flower images. Consequently, the boundary is clear and distinct when projecting the flower image on the base of the flower (shown in Fig. 2 (k)) but blurry when projecting the flower image on the base of the horse (shown in Fig. 2 (h)). We have tested the projection of the flower with other category spaces, such as the elephant and building, and shown the results in the Fig. 2 (i) and Fig. 2 (j), respectively. It is obvious that the boundaries are blurred, or the objects can not be enhanced or specifically distinguished. To state our points more clearly, we have shown another example of the horse image. We can see that the boundary is clear and the object can be specifically distinguished when projecting the horse image on the base of the horse (Fig. 2 (v)), while the boundary is blurry or the object can not be distinguished when projecting it on the base of the flower, elephant or building (Fig. 2 (s), Fig. 2 (t) and Fig. 2 (u)).

4.2.2 Classification Results

In this subsection, the advantage of ICA as a preprocessing tool for extracting SIFT descriptors was evaluated by comparison of the overall categorization accuracy and average classification rates among six different color SIFT and standard SIFT descriptors.

(1) The SIMPLICITY Image Database

In the SIMPLICITY database (see Fig. 3), 1000 object and scene images including ten different categories (people, beaches, buildings, buses, dinosaurs, elephants, flowers, horses, mountains and food) are divided into two parts: 600 color images for training (60 for each category) and 400 color images for testing (40 for each category).

As shown in Fig. 6, the results for the people, bus, flower and mountain categories using CIC-SIFT descriptors are the best ones among those using other SIFT descriptors. Furthermore, we can see obviously from Fig. 4 that the average classification rate of our proposed CIC-SIFT descriptor is estimated as 82%, which is the highest among all other SIFT descriptors. It denotes that CIC-SIFT descriptors can

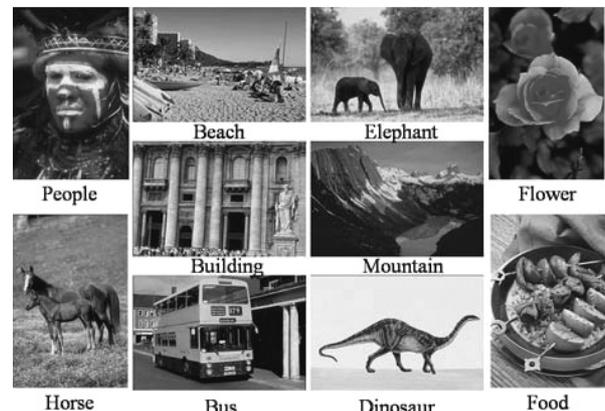


Fig. 3 The SIMPLICITY image database.

improve object/scene classification.

(2) Object Database

The object database (see Fig. 5) is composed of ten object categories summarized in Table 1. Since the image sizes vary from 16,000 up to 530,432 pixels, we resized all images referring to aspect ratio to about 120,000 pixels.

Figure 7 shows the results of overall categories' classification experiments using 40 images per category for training and the rest for testing. The highest categorization rate (97.7%) is obtained in car_brad with CIC-SIFT descriptors. Meanwhile, the results using CIC-SIFT descriptors for leopards, cars-brad, faces, house and leaves categories outperform those using other SIFT descriptors. As shown in Fig. 8, our proposed CIC-SIFT based average classification rate is

62.3%, which is the highest one among all other SIFT descriptors. Thus, we can see that CIC-SIFT descriptors are effective and discriminative for object/scene classification.

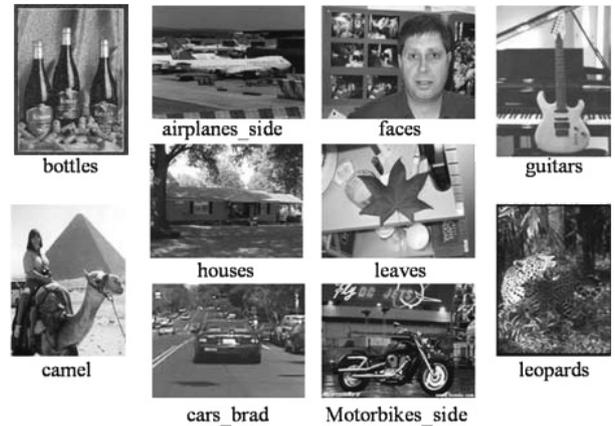


Fig. 5 The object database.

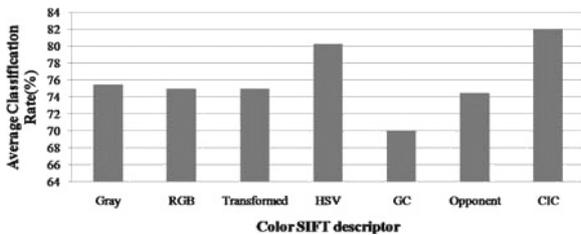


Fig. 4 Average classification rate of the SIMPLICITY using seven SIFT descriptors.

Table 1 Image number of each category.

Category	leopards	airplanes	bottles	camel	cars_brad
Number	200	1074	247	356	1155
Category	faces	guitars	house	leaves	motorbikes
Number	450	1030	1000	186	826

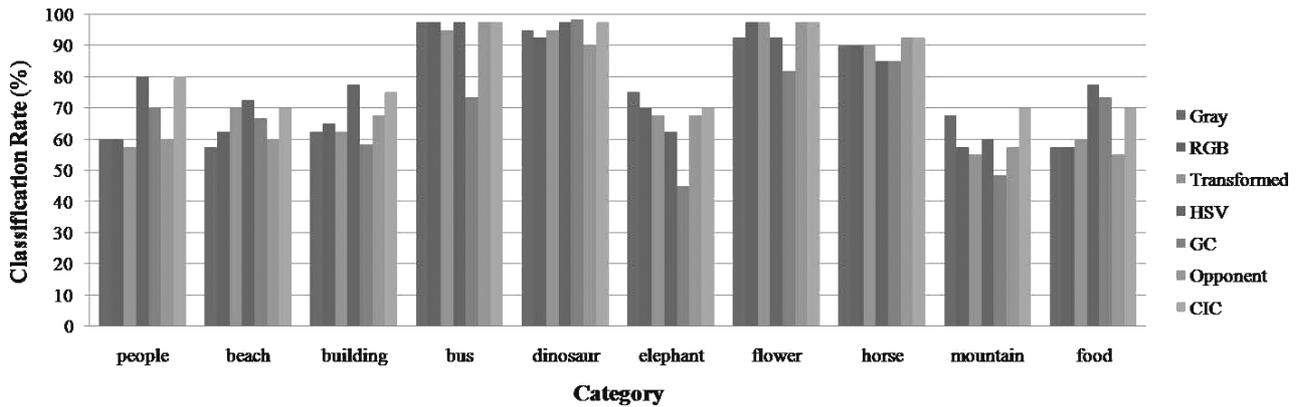


Fig. 6 Classification rate of ten categories in the SIMPLICITY database.

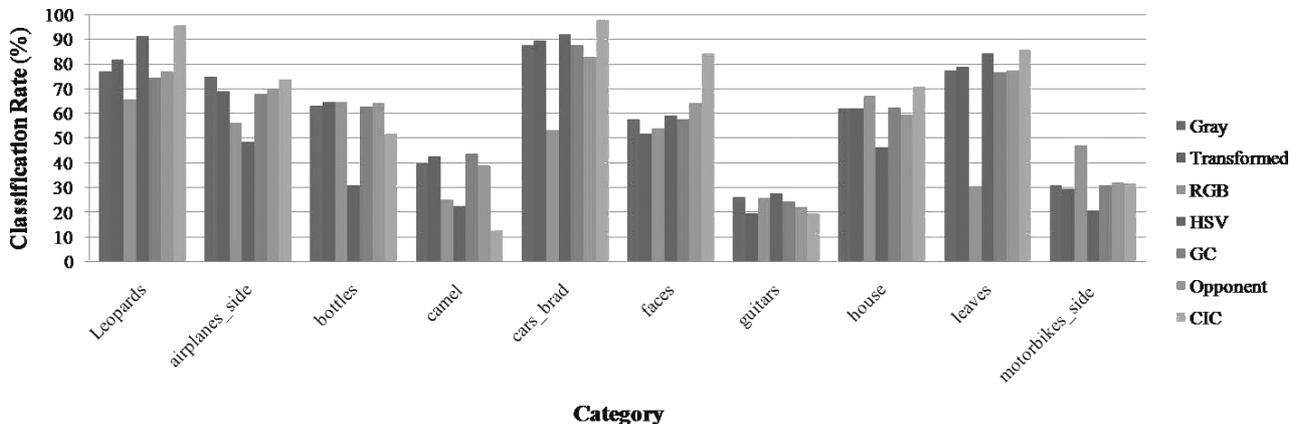


Fig. 7 Classification rate of ten categories in the object database.

(3) Scene Database

The final experiment was implemented with the Scene database, which is composed of eight scene categories (see Fig. 9). Each category has 400 images, and resized images are 500×375 pixels. 40 images per category are used as training data, and the rest are test data.

Figure 10 displays the categorization results over all eight scene categories constituting the entire database. Most categories, such as beach, closeup_flower, cooking, party

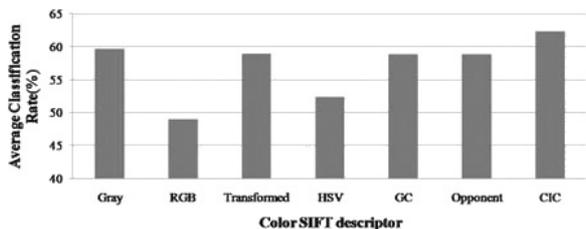


Fig. 8 Average classification rate of the object database using seven SIFT descriptors.

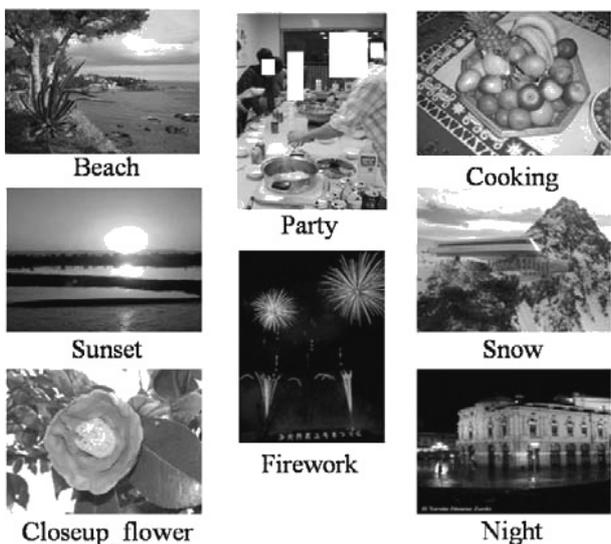


Fig. 9 The scene database.

and sunset, can get the best categorization rates using CIC-SIFT descriptors. We now take a closer look at the classification results from the confusion matrices of Opponent-SIFT (Table 2) and CIC-SIFT (Table 3).

The confusion matrices show that the precision of the CIC-SIFT descriptors is higher than those of the Opponent-SIFT descriptors. As shown in Table 2, beach, cooking and snow can obtain the top three classification rates. In comparison with Opponent-SIFT, CIC-SIFT utilizes both specific color information and a clearer boundary for classification, which can extract more adaptive and discriminative features from each category. Take beach category for example, major confusion appears between beach and sunset, when the

Table 2 Confusion matrix of the Opponent-SIFT based classification.

Average (62.1%)	Classification rates in (%)							
	be	cl	co	fi	ni	pa	sn	su
True class								
beach	77.5	0.0	0.0	3.3	0.0	0.8	8.1	10.3
flower	0.3	43.1	42.2	3.1	0.8	6.4	3.9	0.3
cooking	3.3	7.8	66.4	3.6	0.3	1.4	16.4	0.8
firework	2.5	0.0	0.6	66.7	22.2	0.8	4.2	3.1
night	3.9	0.6	0.3	30.0	50.0	7.8	3.1	4.4
party	0.4	6.7	25.6	2.1	5.5	58.8	0.8	0.0
snow	14.7	0.0	10.6	3.1	0.8	0.8	63.1	6.9
sunset	20.0	0.0	0.0	1.7	2.2	0.8	4.7	70.6
Precision	63.3	77.1	48.5	59.1	62.5	67.3	60.7	73.2

Table 3 Confusion matrix of the CIC-SIFT based classification.

Average (69.1%)	Classification rates in (%)							
	be	cl	co	fi	ni	pa	sn	su
True class								
beach	84.4	0.0	0.3	0.8	0.3	0.6	8.6	5.0
flower	0.0	56.4	35.3	1.9	0.3	4.7	1.4	0.0
cooking	3.9	2.8	82.8	1.1	1.4	2.2	5.3	0.6
firework	0.6	0.6	4.2	63.1	19.7	0.8	6.7	4.4
night	5.0	0.0	1.9	20.6	56.7	9.2	4.2	2.5
party	0.0	2.1	24.4	2.9	4.2	65.1	1.3	0.0
snow	8.6	0.0	3.6	1.1	1.4	0.8	79.2	5.3
sunset	20.6	0.0	1.1	3.3	2.8	1.4	5.6	65.3
Precision	68.6	92.3	57.0	67.2	66.4	68.6	70.9	78.6

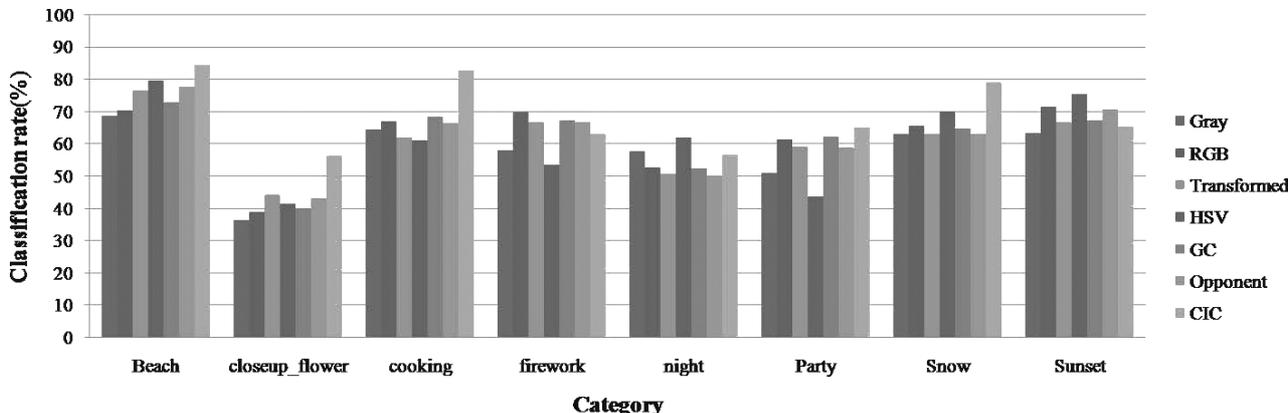


Fig. 10 Classification rate of eight categories in the scene database.

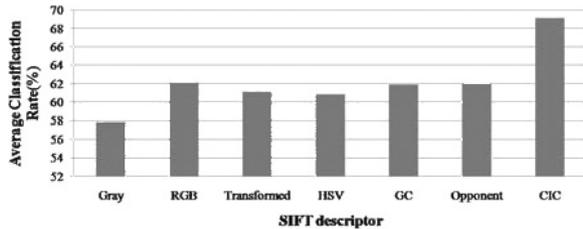


Fig. 11 Average classification rate of the scene database using seven SIFT descriptors.

Opponent-SIFT is used. This results mainly from the fact that except for a darker quality, the sunset images are similar to beach images. Since our proposed CIC-SIFT learns transformation matrix from each category, unlike Opponent-SIFT, specific color information can be extracted. Therefore, the negative classification rate to sunset is reduced by using CIC-SIFT. Another major confusion appears between cooking and snow when the Opponent-SIFT is used. However, most of boundaries between different objects are blurry. Consequently, by using CIC-SIFT, the negative classification rates to cooking and snow are both reduced remarkably.

As shown in Fig. 11, we can see clearly that the average classification rate is 69.1% using CIC-SIFT as descriptors, which is much higher than other best results of 62.1%, achieved using Opponent-SIFT.

Thus, we can conclude that our proposed CIC-SIFT descriptors can not only increase the classification rate of object-based images, but also improve that of scene images.

Since only SIFT descriptors are utilized as the local feature of images, the classification rate is not high enough. However, the motivation of our work in this paper is to prove the efficiency of the CIC-SIFT descriptor. The experimental results show that our proposed CIC-SIFT is superior to other SIFT descriptors in object/scene classification.

5. Conclusion

In this paper, we proposed a novel color independent components based SIFT descriptor (CIC-SIFT) for object/scene classification. The proposed color transformation based on independent component analysis (ICA) is adaptive and suitable to each category, which can be used for enhancing the contrast between the object and its background. The proposed CIC-SIFT can extract more effective and discriminative local features for object/scene classification. The experimental results show that our proposed CIC-SIFT is superior to other conventional color SIFT descriptors for both object and scene images.

References

- [1] K.E.A. van de Sande, T. Gevers, and C.G.M. Snoek, "Evaluation of color descriptors for object and scene recognition," IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2008.
- [2] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol.60, no.2, pp.91–110, 2004.
- [3] A. Bosch, A. Zisserman, and X. Muoz, "Scene classification using a hybrid generative/discriminative approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.30, no.4, pp.712–727, 2008.
- [4] P. Comon, "Independent component analysis—a new concept?," *Signal Process.*, vol.36, pp.287–314, 1994.
- [5] J.Z. Wang, J. Li, and G. Wiederhold, "SIMPLiCity: Semantics-sensitive integrated matching for picture libraries," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.23, no.9, pp.947–963, 2001.
- [6] J.G. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid, "Local features and kernels for classification of texture and object categories: A comprehensive study," *Int. J. Comput. Vis.*, vol.73, no.2, pp.213–238, 2007.
- [7] B. Scholkopf and A. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond*, MIT Press, Cambridge, MA, 2002.
- [8] R. Datta, D. Joshi, J. Li, and J.Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Comput. Surv.*, vol.40, no.2, pp.1–60, 2008.
- [9] J. van deWeijer, T. Gevers, and A. Bagdanov, "Boosting color saliency in image feature detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.28, no.1, pp.150–156, 2006.
- [10] X.Y. Zeng, Y.W. Chen, and Z. Nakao, "Classification of remotely sensed images using independent component analysis and spatial consistency," *J. Advanced Computational Intelligence and Intelligent Informatics*, vol.8, no.2, pp.216–222, 2004.
- [11] C.C. Chang and C.J. Lin, *LIBSVM: A Library for Support Vector Machines*, Software available at: <http://www.wisie.ntu.edu.tw>, 2001.
- [12] A.J. Bell and T.J. Sejnowski, "An information maximization approach to blind separation and blind deconvolution," *Neural Comput.*, pp.1129–1159, 1995.
- [13] G. Buchsbaum and A. Gottschalk, "Trichromacy opponent colours coding and optimum colour information transmission in the retina," *Proc. Royal Society London, B*, vol.220, pp.89–113, 1983.
- [14] D.L. Ruderman, "Statistics of cone responses to natural images: Implications for visual coding," *J. Opt. Soc. Am.*, vol.15, no.8, pp.2036–2045, 1998.
- [15] G. Deco and D. Obradovic, *Independent Component Analysis: General Formulation and Linear Case, An Information-theoretic Approach to Neural Computing*, Springer, 1996.
- [16] H. Barlow, *Sensory Communication*, MIT press, pp.217–234, 1961.
- [17] S. Muraki, T. Nakai, Y. Kita, and K. Tsuda, "Basic research for coloring multichannel MRI data," *IEEE Trans. Vis. Comput. Graphics*, vol.7, no.3, pp.265–273, 2001.
- [18] A.E. Abdel-Hakim and A.A. Farag, "CSIFT: A SIFT descriptor with color invariant characteristics," *Computer Vision and Pattern Recognition (CVPR'06)*, vol.2, pp.1978–1983, 2006.
- [19] J.M. Geusebroek, R. van den Boomgaard, A.W.M. Smeulders, and H. Geerts, "Color invariance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.23, no.12, pp.1338–1350, 2001.
- [20] F. Mindru, T. Tuytelaars, L. Van Gool, and T. Moons, "Moment invariants for recognition under changing viewpoint and illumination," *Computer Vision and Image Understanding*, vol.94, no.1-3, pp.3–27, 2004.
- [21] R. Fergus, P. Perona, and A. Zisserman, "A sparse object category model for efficient learning and exhaustive recognition," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2005.
- [22] L. Fei-Fei, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories," *IEEE. CVPR 2004, Workshop on Generative-Model Based Vision*, 2004.
- [23] J.B. MacQueen, "Some methods for classification and analysis of multivariate observations," *Proc. 5th Berkeley Symposium on Mathematical Statistics and Probability*, pp.281–297, University of California Press, 1967.
- [24] D. Han, W. Li, T. Wang, L. Liu, and Y. Wang, "Independent components analysis for representation interest point descriptors," *Lect. Notes Comput. Sci.*, vol.4113, pp.1219–1223, 2006.

- [25] M. Yamazaki and S. Fels, "Local image descriptors using supervised kernel ICA," Pacific-Rim Symposium on Image and Video Technology (PSIVT), pp.94–105, 2009.
- [26] T. Nakai, S. Muraki, E. Bagarinao, Y. Miki, Y. Takehara, K. Matsuo, C. Kato, H. Sakahara, and H. Isoda, "Application of independent component analysis to magnetic resonance imaging for enhancing the contrast of gray and white matter," Neuroimage, vol.21, no.1, pp.251–260, Jan. 2004.



Dan-ni Ai received a B.E. degree in 2005 from Xi'an Jiaotong University, China, an M.E. degree in 2008 from Xi'an Jiaotong University, China. She is currently a Ph.D. candidate at the graduate School of Engineering and Science, Ritsumeikan University. Her current research interests include image processing and analysis, feature extraction and image annotation.



Xian-hua Han received a B.E. degree in 1999 from ChongQing University, China, an M.E. degree in 2002 from ShanDong University, JiNan, China, and a D.E. degree in 2005 from University of the Ryukyus, Okinawa, Japan. From 2006 to 2007, she was a lecture with Electronics and Information Engineering School Central South University of Forestry and Technology, China. She is currently a post-doctoral fellow at college of Information Science and Engineering, Ritsumeikan University. Her current

research interests include image processing and analysis, feature extraction, image annotation and pattern recognition.



Xiang Ruan received the B.S. degree in Shanghai Jiaotong University in 1997 and the M.S. and the Ph.D. degrees in Osaka City University in 2001 and 2004. He is currently a researcher of electronic components company of OMRON Corp. His current research interests include computer vision, pattern recognition and image processing. He is a member of the IEEE and the IEEE Computer Society.



Yen-wei Chen received a B.E. degree in 1985 from Kobe Univ., Kobe, Japan, a M.E. degree in 1987, and a D.E. degree in 1990, both from Osaka Univ., Osaka, Japan. From 1991 to 1994, he was a research fellow with the Institute of Laser Technology, Osaka. From Oct. 1994 to Mar. 2004, he was an associate Professor and a professor with the Department of Electrical and Electronic Engineering, Univ. of the Ryukyus, Okinawa, Japan. He is currently a professor with the college of Information Science and Engineering, Ritsumeikan University, Kyoto, Japan. He was a visiting professor with the Oxford University, Oxford, UK in 2003. He is an Overseas Assessor of Chinese Academy of Science and Technology. He is an associate Editor of International Journal of Image and Graphics (IJIG) and editorial board members of the International Journal of Knowledge-based and Intelligent Engineering Systems and the International Journal of Information. His research interests include pattern recognition, image processing and machine learning. He has published more than 200 research papers in these fields.

ence and Engineering, Ritsumeikan University, Kyoto, Japan. He was a visiting professor with the Oxford University, Oxford, UK in 2003. He is an Overseas Assessor of Chinese Academy of Science and Technology. He is an associate Editor of International Journal of Image and Graphics (IJIG) and editorial board members of the International Journal of Knowledge-based and Intelligent Engineering Systems and the International Journal of Information. His research interests include pattern recognition, image processing and machine learning. He has published more than 200 research papers in these fields.