Front. Comput. Sci., 2017, 11(6): 1023–1035
DOI 10.1007/s11704-016-5538-y

RESEARCH ARTICLE

Color space quantization-based clustering for image retrieval

Le DONG (✉), Wenpu DONG, Ning FENG, Mengdie MAO, Long CHEN, Gaipeng KONG

School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

Abstract Color descriptors of an image are the most widely used visual features in content-based image retrieval systems. In this study, we present a novel color-based image retrieval framework by integrating color space quantization and feature coding. Although color features have advantages such as robustness and simple extraction, direct processing of the abundant amount of color information in an RGB image is a challenging task. To overcome this problem, a color space clustering quantization algorithm is proposed to obtain the clustering color space (CCS) by clustering the CIE1976L*a*b* space into 256 distinct colors, which adequately accommodate human visual perception. In addition, a new feature coding method called feature-to-character coding (FCC) is proposed to encode the block-based main color features into character codes. In this method, images are represented by character codes that contribute to efficiently building an inverted index by using color features and by utilizing text-based search engines. Benefiting from its high-efficiency computation, the proposed framework can also be applied to large-scale web image retrieval. The experimental results demonstrate that the proposed system can produce a significant augmentation in performance when compared to block-based main color image retrieval systems that utilize the traditional HSV(Hue, Saturation, Value) quantization method.

Keywords content-based image retrieval, color space quantization, feature coding, inverted index

1 Introduction

In recent years, explosive growth in digital image libraries has occurred due to the popularization of web cameras, digital cameras, and mobile phones equipped with such devices. Thus, the tremendous scale and diversity of images inevitably raises the need to develop an efficient technique that can automatically search through desired images in an image database. Typically, there are two main automatic image retrieval methods [1–3]: text-based image retrieval (TBIR) and content-based image retrieval (CBIR) [3,4]. TBIR is based on keywords and can often achieve the highest retrieval accuracy when images are well annotated. However, there are many problems limiting the further development of TBIR, such as the cost of manual annotation of keywords, and differences in annotator perceptions and interpretations. In addition, a picture is worth a thousand words [5]; in other words, the keywords of an image cannot describe all of the associated information. In the early 1990s, Kato [6] first proposed the concept of content-based image retrieval. Instead of taking keywords as input, CBIR directly takes an image as the query.

CBIR utilizes visual features to represent and index images, and returns a ranked list of its nearest neighbors from an image database. Feature extraction is one of the key steps of CBIR. Color, texture, shape, spatial relations, and other characteristics of images are widely used visual features in image retrieval, as they reveal the fundamental information of the content and do not require special knowledge or context to understand. In recent years, researchers have launched more extensive studies on CBIR that have included image indexing

Received December 13, 2015; accepted December 5, 2016

E-mail: ledong@uestc.edu.cn

[7–9], semantic relationships [10–12], visual feature extraction, selection [13–18], re-ranking [19,20], and user feedback [5,21]. As a result, a number of successful image retrieval systems [18,22–24] have been implemented.

The visual features used in CBIR can be further classified as general features such as color, texture, shape, and domain-specific features such as human faces and fingerprints. Among all features, color is obviously a fundamental and intuitive characteristic of visual content, and is deemed to be the most common visual feature, as mentioned in [25–27]. The popularity of color features is also owed to their robustness to the image scale and orientation, and efficiency in computation. Color features are extracted to represent image color distributions and are often presented as color histograms [25]. Such common color feature extraction methods first rigidly partition the underlying color spaces into a fixed number of bins, then place pixels into the bin that most closely matches their color. The value of a bin denotes the percentage of pixels in an image belonging to that bin. Therefore, a color histogram can be regarded as a quantized color distribution of an image.

In general, human vision cannot identify every true color, or more specifically, it can only distinguish approximately 150 different colors. Thus, it undoubtedly causes excess space and time consumption when color features are directly computed in the original color space. This excess overhead is unnecessary and avoidable. In order to decrease the complexity of the color feature extraction process, [28,29] addressed this problem through color quantization in the HSV color space. Although these methods achieved satisfactory efficiency, they reduced the descriptive capacity of the color space. Furthermore, several studies demonstrated that the CIE1976L*a*b* color space generally outperforms RGB and HSV [14]. Inspired by this information, we propose a color space clustering quantization method by employing the K-means algorithm with almost no reduction in the descriptive capacity of the color space, or effect on human visual perception. First, we divide the colors in the CIE1976L*a*b* color space into two parts: gray and non-gray. Then, we employ K-means on the two parts to obtain the 256 distinct colors that adequately accommodate human visual perception ability. The 256 distinct colors comprise the clustering color space (CCS) in which each distinct color corresponds to a color code. In order to achieve efficient color feature extraction, we then map the images to the CCS to obtain color-coded images.

In [30], Zhang et al. analyzed the robustness of several histogram techniques by examining image characteristics. They verified that block color histograms demonstrate superior performance than common color histograms. Thus, in this study, we employ block color histograms as the initial descriptors of color information. Then, the distinct color code of the largest value bin of each block color histogram will be extracted to serve as the color feature of the corresponding image block. To efficiently retrieve similar images from the database by utilizing the color features, we study the techniques of the inverted index [31] and text-based commercial search engines (e.g., Google, Bing, Yahoo, etc.). An inverted index is a data structure that stores a mapping of content to its location in a database file or in a set of documents. Specifically, it contributes to the reduction of the overhead of storing and comparing high-dimensional color features, and avoids the comparison between the query image and all the dataset images. The inverted index and text-based search engines have been thoroughly established and achieved substantial success in text retrieval. To exploit the advantages of these techniques in CBIR, we propose a novel feature coding method called feature-to-character coding (FCC). This method involves dividing each main color code into the unit digit part and the other digit part. As the number of distinct colors in CCS is 256, the range of the unit digit part is [0–9] and the range of the other digit part is [0–25], both of which are contained in the corresponding set of alphabetic characters. Thus, FCC can implement efficient coding by mapping the color features to characters.

The main contribution of this study is the proposed color space clustering quantization method that operates by aggregating similar colors and taking the human color vision system into account. The method successfully avoids inducing substantial time and memory consumption when directly extracting the color features from RGB images, and at the same time improves retrieval accuracy based on the traditional HSV quantization algorithm. In addition, we devise an efficient feature coding method to encode color features to character codes and explore the superiority of content-based image retrieval. We achieve this by combining the inverted index and a text-based search engine. To evaluate the proposed framework, we employ a subset of a Corel image collection containing 10 categories, each of which contains 100 images. Extensive experiments were performed and the results are presented to demonstrate the superiority of our framework.

The remainder of this study is structured as follows. Section 2 introduces the related works, the proposed framework is elaborated in Section 3, the experiments are described in detail in Section 4, and Section 5 concludes the study and discusses the direction of future work.

2 Related work

To achieve good image retrieval performance, numerous works have been proposed by applying the bag-of-words (BoW) model [9,31–34], some of which have been successful. In this approach, an image is represented as a high-dimensional and sparse histogram of visual word counts. The standard pipeline based on the BoW model is initiated by setting up the standard framework [9,32,33]. This involves using term frequency and inverse document frequency to rank similar images. However, as mentioned in [35], this pipeline may result in feature drop-out, noisy descriptors, inappropriate metrics for descriptor comparison, or loss of accuracy due to descriptor quantization.

Different from the BoW model, several other studies calculate the similarities of visual contents by measuring the image similarity based on color, texture, shape, or spatial relationships [13–16,18,21]. The basic idea of CBIR is proposed in 1990 [25]. To improve the retrieval performance, numerous methods combine multiple features to describe the visual contents of an image. For instance, block-based visual features and user query, concept-based, semantic features were integrated in [16]. In [18], a novel hierarchical manifold subgraph ranking system for CBIR was proposed. The system achieved more meaningful structures for the image space by incorporating both visual and semantic similarities. In [19], a re-ranking stage used a compression-based distance measurement for the top-ranked results returned by the traditional CBIR process. The visual features that refined the candidate annotation generated from pages and relevant user feedback were combined to classify the visual features into three categories [23]. These techniques have made obvious improvements in terms of the accuracy of image retrieval performance. However, as the number of features increases, the difficulty of extracting features and measuring the similarity also increases, especially when the number of images increases to a much larger scale (millions). In this study, multiple feature fusion is not addressed.

Among the visual contents, color features seem to be more robust to image scale and are easier to be extracted. Color histograms were used as image representations in color-based image retrieval [25]. Several later works continue to improve color-based image retrieval. For instance, in [12], three novel morphological color descriptors are proposed: one utilizes granulometries that are independently computed for each subquantized color, and the other two descriptors employ the principle of multiresolution histograms to describe color. They utilize morphological leveling and watersheds, respectively.

Based on the binary quaternion-moment- preserving thresholding technique, an adaptive color feature extraction scheme [13] is proposed by preserving the color distribution of an image. These methods achieved an improvement in retrieval performance, but they led to an increase in computational complexity. Therefore, it is a better choice to improve performance through simple and effective designs. By analyzing the invariance properties and the distinctiveness of color descriptors in a structured manner, it was noted that color histograms have invariance to scale and rotation, and the color histogram in HSV color space is better than RGB histogram and color moments [26]. In [30], the color histogram based on representative color, and the color histogram based on representative color and partitioning, is demonstrated to have better retrieval results with a small computational overhead. Here, we follow this approach and aim to take full advantage of the block-based color histogram in the HSV color space to achieve good image retrieval performance.

The process of color space quantization is a technique that reduces the number of colors used in the color space. Usually, all colors within a specific range in a color space are mapped to a single histogram bin and treated as the same color. Color space quantization is the basis of color histogram extraction and has an influence on the accuracy of a color histogram. Most color space quantization methods are used for the HSV color space. For example, it presented a new thresholding method to separate gray pixels from non-gray pixels, and quantized the HSV color space into 24 non-uniform bins based on HSV soft decision [28]. In [29], a color quantization approach in a conical HSV color space was proposed, and it generated a 3D color histogram from the 2D SV plane while having the same number of bins. It is well known that the color space selection will influence the color feature extraction and description. In general, it was proved that CIE1976L*a*b* outperformed RGB, HSV, and LSH in the context of color description [14, 36]. Thus, we propose a new color space quantization method by using K-means on CIE1976L*a*b*. By analyzing properties of the CIE1976L*a*b* color space, we divide the color space into a gray section and a non-gray section, and then we employ K-means on the both sections. This yield 8 gray colors and 248 other distinct colors, which preserves the descriptive ability of the color space as much as possible.

Based on an analysis of the similarity of feature vectors, we propose a novel FCC method that aims to convert color

features into character codes. The main color code of each image block is divided into two parts, a unit digit part and other digit part, and then each part is encoded to characters according to the mapped rules between the value of each part and characters. Thus, images are represented by the character codes. The method utilizes the well-developed inverted index technology and highly-efficient text-based search engine to perform image retrieval.

3 Methodology of the framework

3.1 System overview

In this section, we first present the different parts of our image retrieval system, and then we introduce the proposed color space clustering quantization and FCC methods.

Figure 1 illustrates the proposed retrieval framework. First, it quantizes the CIE1976L*a*b* color space into 256 distinct colors by applying the classical *K*-means clustering algorithm on the gray and non-gray sections. The 256 colors are used to construct the CCS. Specifically, each color in the CCS corresponds to a color cluster containing every color within a specific range in the CIE1976L*a*b* color space. The 256 distinct colors are then sorted and the corresponding serial numbers are regarded as color codes. We combine distinct colors and their corresponding color codes to form the color code table that represents the CCS. Given an image typically represented in the RGB color space, we transform the image pixels into the CIE1976L*a*b* color space, then map the CIE1976L*a*b* pixel values into the CCS by comparing them to 256 distinct colors one at a time. This procedure must comply with the principle of minimum chromatic aberration (see Section 3.2.1). Secondly, the mapped images, referred to as the color-coded images, will be divided into sub blocks to compute the block color histograms. The color code coming from the largest histogram bin of each block will be selected as the main color code of this block. The collection of the main color codes is called the scalar color code and is used to serve as the color feature of an image. Thirdly, according to the proposed FCC method, the scalar color codes are encoded to the corresponding character codes. Each character code consists of two or four characters. These character codes are used to build the inverted index, and then a text-based search engine is employed to accomplish the image retrieval task with the character codes representing the given image.

3.2 Color space clustering quantization

3.2.1 The color space conversation

True color images typically use 24 bits per pixel, which results in an overall gamut of 224; i.e., more than 16 million different colors. However, human eyes can distinguish approximately 150 different colors. Hence, it is feasible to quantize all colors in the color space to a smaller number of distinct colors without losing major descriptive power in the

Fig. 1 System framework

histograms. In this study, we choose the CIE1976L*a*b* color space because of its high similarity to human vision and the perceptual uniformity. The color space denotes that the same geometrical distances in different positions and directions correspond to the equal chromatic aberration CIE1931XYZ, which is derived from the liner transformation of CIE1931RGB (abbreviated as RGB). The CIE1976L*a*b* color space is converted from CIE1931XYZ [34] and the conversion formula is defined as Eq. (1):

$$\begin{cases} L^* = 116f(Y/Y_0) - 16, \\ a^* = 500[f(X/X_0) - f(Y/Y_0)], \\ b^* = 200[f(Y/Y_0) - f(Z/Z_0)], \end{cases} \tag{1}$$

where L^* denotes the color brightness, and a^* and b^* jointly determine the color position in the color space. Specifically, the positive values of a^* and b^* represent red and yellow, while the negative values indicate green and blue. In Eq. (1), X, Y, and Z are the tristimulus values of the object, while X_0, Y_0, and Z_0 denote the CIE XYZ tristimulus values of the reference white point. Here, $f(\cdot)$ is a function and is defined by Eq. (2):

$$\begin{cases} f(t) = t^{1/3}, & t > \left(\dfrac{6}{29}\right)^3, \\ f(t) = \dfrac{1}{3}\left(\dfrac{29}{6}\right)^2 t + \dfrac{16}{116}, & \text{otherwise.} \end{cases} \tag{2}$$

Note that although the color space quantization can serve as an effective method for reducing the scale of the color space, it demands that the mapped image must visually resemble the original image as closely as possible. To this end, we introduce the principle of minimum chromatic aberration, which acts as an important criterion in the determination of whether two colors are similar.

Chromatic aberration identifies the difference between two colors from the perspective of human vision. Given two colors represented by L_1^*, a_1^*, b_1^*, and L_2^*, a_2^*, b_2^*, the total chromatic aberration, ΔE_{ab}^*, and each individual chromatic aberration (lightness difference ΔL^*, chroma difference Δa^* and Δb^*) can be calculated according to the following formulas:

$$\Delta L^* = L_1^* - L_2^*, \tag{3}$$

$$\Delta a^* = a_1^* - a_2^*, \tag{4}$$

$$\Delta b^* = b_1^* - b_2^*, \tag{5}$$

$$\Delta E_{ab}^* = \sqrt{(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2}. \tag{6}$$

A smaller ΔE_{ab}^* value indicates that the two colors are more visually similar. Equation (6) presents an equal form of the Euclidean distance metric formula, which demonstrates that we can apply the clustering methods based on Euclidean distance to quantize the CIE1976L*a*b* color space. In our work, we utilize K-means to achieve the quantization purpose.

3.2.2 The color space clustering quantization

Color quantization uses a color palette that contains only a small number of colors to represent true color images. Because K-means is easy to implement with high efficiency, it is commonly used to perform clustering analysis. The proposed color space clustering quantization is based on K-means. The main intention of K-means is to find k clusters $[c_1, c_2, \ldots, c_k]$ from the vector set that has n D-dimensional feature vectors $[x_1, x_2, \ldots, x_n]$ by minimizing the within-cluster sum of the squares of the Euclidean distance from each vector to the nearest clustering center [37,38]. The square of the distance is called the objective function and is marked as W_n:

$$W_n = \sum_{i=1}^{n} \sum_{x_i \in c_j} \|x_j - u_j\|^2, \tag{7}$$

where x_i denotes the ith D-dimensional vector and c_j and u_j denote the jth cluster and its cluster center. Despite its simplicity and high efficiency, K-means undesirably possesses an inherent shortcoming—it has difficulty determining the value of K. Selecting an incorrect K value may have a negative influence on clustering performance. However, in this research, K can be easily chosen based on the application. There are two reasons why the value of K is set to 256. One reason is that the color-coded images cost 8 bits per pixel. Thus, the range of K is usually between 8 and 256, and here, we choose the maximum value of K. The other reason is that the number of colors that the human eye can distinguish is approximately 150 with a small fluctuation. Taking into consideration that different people have a different ability to distinguish colors, we set K as 256 to capture the maximum range of human vision. Two hundred and fifty-six distinct colors comprise the CCS and Fig. 2 illustrates the visualization of the CCS. As depicted, Fig. 2 is divided into two parts by a vertical bar. Specifically, each color bar on the left denotes one of 256 distinct colors, while each color bar on the right shows all the colors belonging to the corresponding color cluster of CIE1976L*a*b*. We can observe that the colors in a color cluster present few visual differences; this demonstrates that it is feasible and effective to utilize a distinct color to represent all the colors belonging to the corresponding color cluster. Meanwhile, all the distinct colors can be clearly distinguished from each other; this suggests that the gained CCS

successfully meets the requirement of low scale and obvious distinction. Figure 3 provides a comparison of the original images and the color-coded images. Visually, there are almost no distinct differences between the two color images in Figs. 3(a)–3(d), while the lighter images, Figs. 3(e) and 3(f), have obvious differences in the background. The deficiency is mainly because the gray color accounts for a very small proportion of the entire CIE1976L*a*b* color space. Thus, it is difficult to maintain the gray color information in the process of clustering.

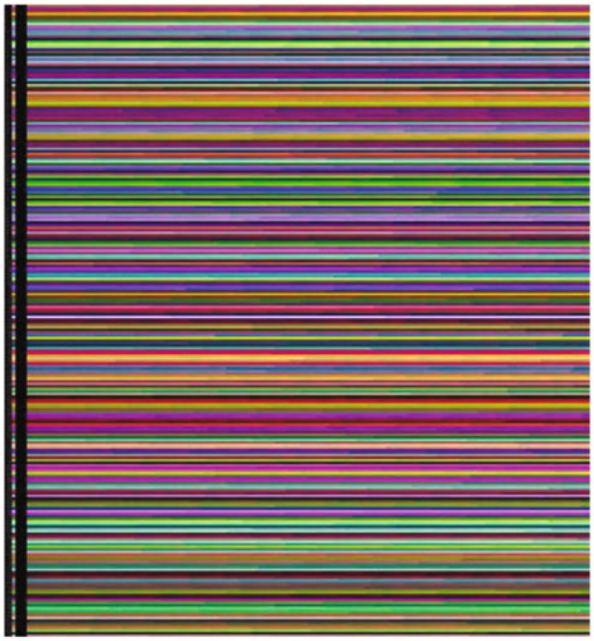


Fig. 2 The visualization of CCS

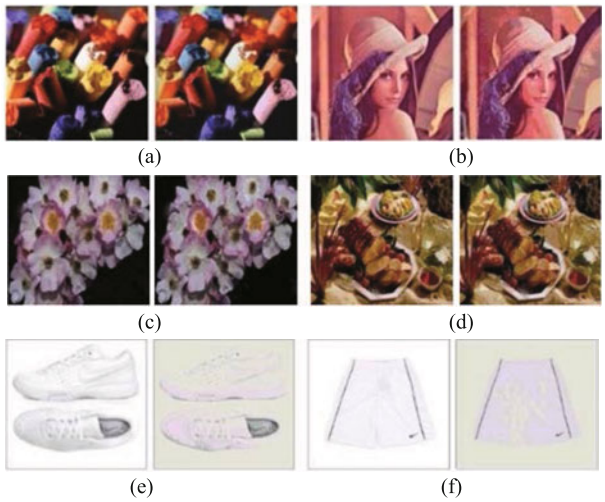


Fig. 3 Comparison of the original images and color-coded images (Left: original images; Right: color-coded images)

To overcome the aforementioned deficiency and compensate for the lost gray level information, the gray colors in the CIE1976L*a*b* color space are distinguished from other colors and clustered separately. The entire color space is divided into gray and non-gray sections. We apply K-means to the gray section to obtain 8 distinct grey colors, and we apply K-means to the non-gray section to acquire 248 other distinct colors. From Fig. 4, it is clear that the results of lighter images from mapping of the optimized CCS are much better than the results without optimization.

Fig. 4 Comparison between results of CCS with optimization and results without optimization

Given the 256 distinct colors, we sort them to gain the serial numbers as the color codes. Each distinct color is represented by a color code, and the distinct colors together with the corresponding color codes comprise the color code table. Figure 5 provides the color code table where each individual vertical table presents a distinct color and its corresponding color code. For instance, the first column with vectors (58.5, 9.96, 31.6, 0), 58.5, 9.96, 31.6 represent the values of L^*, a^*, b^* of a distinct color, respectively, and 0 denotes the color code.

58.5	43.23	...	38.67	...	72.35	...	100
9.96	72.35		−12.44		−0.83		0.00
31.6	−67.7		−16.18		4.93		−0.01
0	1		33		51		255

Fig. 5 Color code table

3.2.3 Scalar color code

Color features have played important roles in CBIR owing to their simplicity and efficient extraction. In particular, color histograms serve as a commonly used color feature. While color histograms provide a global description of the color information, they fail to present detailed color information within an image. This limitation motivates us to employ a

color histogram based on representative color and partitioning to present the color distribution of color blocks in an image.

Given an RGB image whose size is $w \times h$, we divide it into $p \times q$ blocks, each of which shares the same size of $(w/p) \times (h/q)$. Specifically, the RGB color value of each pixel can be converted into the color value of the CIE1976L*a*b* color space, and then the L^*, a^*, and b^* values are mapped to the CCS to obtain the color codes that comply with the principle of minimum chromatic aberration. After obtaining the color codes, we then compute the block color histograms and choose the color codes from the largest bins of the block histograms to be the main color codes of the corresponding image blocks. The main color codes from all the image blocks are collected to construct scalar color codes. Thus, the given image can be represented by the scalar color codes containing $p \times q$ main color codes.

3.3 Feature to character coding

The index and retrieval algorithms play an important role in CBIR insofar as they affect the retrieval performance in terms of accuracy, efficiency, and memory usage. An inverted index is one of the most widely used index structures. Using an inverted file index not only avoids storing and comparing high-dimensional visual features, it also reduces the number of target images. This is because only the images that share the same visual features with the query image need to be matched. Text-based commercial search engines are more thoroughly developed compared to content-based image search engines, and have achieved great success in text retrieval. To directly utilize the inverted index and text-based search engine, we propose the simple FCC coding method to convert the scalar color codes into character codes.

The main idea of the FCC method is to map the feature vector to alphabetic characters. Specifically, if the coding operation involves color spatial distribution information, we call it feature-to-character coding with spatial information (FCCS); otherwise, it is simply referred to as FCC. For each image block, the coordinate information can represent the spatial location of block color histograms and serve as an identification label. Therefore, we record the coordinate information as the spatial distribution information. Each main color code is encoded to a character code that contains two characters by FCC, or four characters by FCCS. Here, we take the FCCS as an example to introduce the detailed coding procedure.

1) The coordinate (i, j) of the main color code corresponds to the ith and jth characters of the alphabet, where $i \in [0, p-1]$ and $j \in [0, q-1]$. Let A represent the alphabet and (i, j) is coded as (A_i, A_j). For example, the coordinate $(3, 2)$ is encoded as (D, C).

2) Each main color code is partitioned into two parts. The tens and hundreds digits are regarded as the first part and the unit digit is the second part. Thus, the main color is represented by (m, n) for $m = 0, 1, \ldots, 25$ and $n = 0, 1, \ldots, 9$. Likewise, the coordinate, (m, n) is coded as (A_m, A_n). For instance, 252 is represented as $(25, 2)$ and encoded as (Y, C).

Figure 6 shows the detailed coding procedure. As shown in the figure, the image is represented by the 8×8 scalar color codes and $(8, 8)$ main color code 83 is encoded as HHID according to FCCS. The first two letters indicate the coordinate information of the block histogram, and the last two letters denote the main color code. When all the main color codes are encoded as characters, the image will eventually be represented by the character codes, which consist of $8 \times 8 \times 4$ characters for FCCS. Hence, the image retrieval system can employ the inverted index and a text-based search engine for image retrieval.

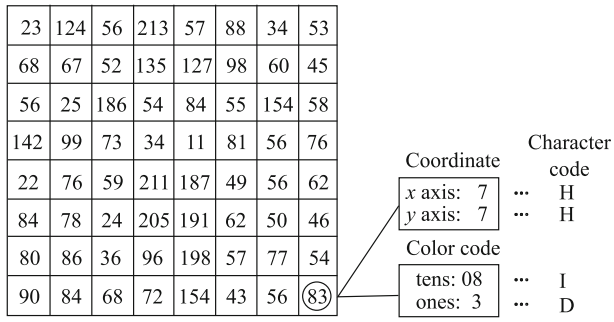


Fig. 6 The procedure of FCCS

4 Experiment and results

In the experiments, we mapped the dataset images to the CCS and extracted the scalar color codes, and then converted the color feature codes into character codes according to FCC and FCCS. Next, we built indexes for these character codes by utilizing inverted index technology, which is critical for achieving fast retrieval. Here, Apache Solr [39], an efficient full-text searching server developed by JAVA and based on Lucene, is adapted in our work to serve as a high-efficiency search engine. Uploading the character codes of all images to the Solr search engine server, we can then obtain the corresponding index information. Because the proposed image retrieval framework is based on the color space quantization and the color FCC method, we consider the method proposed in [30] as our baseline. This method also utilizes the block

color histogram to implement an effective image retrieval system in which the HSV color space is employed and no character coding is used as an optimization. Here, for the sake of simplicity, the approach proposed in [30] is referred to as HSV, and we follow the same experimental setup used in [30] in the experiments. In general, our proposed method is not only compared to HSV, but is also compared to the best performance of four methods based on the HSV color space in [30]. We conduct experiments on ten categories of images and also on cases with different values of N. If not specified explicitly, the following experiments are conducted using a PC with a Pentium(R) Dual-Core 2.93 GHz CPU and 4 GB RAM running the Windows-7 OS.

4.1 Image database and criterion

In order to evaluate the suitability of the proposed image retrieval framework, we performed tests on a subset of a Corel image collection. This dataset consists of ten image classes (people, beaches, buildings, buses, flowers, horses, elephants, dinosaurs, foods, mountains) and each image class includes 100 Corel images. The typical image within each class is shown in Fig. 7. The images in one class have the same perceptual meaning and similar color information. Specifically, all the images are both indexed as database images and used as queries. In the experiments, we resize each image to 200×200 and then divide it into 8×8 blocks. The equal-size grid has advantages when the image is big enough. The performance of the proposed image retrieval framework is measured in terms of the average retrieval precision and recall. In particular, precision indicates the proportion of the retrieved images that are relevant to the query, while recall denotes the proportion of relevant images that are retrieved in the database in response to a query. Each time a query image is selected from the database, it attempts to retrieve N best matched images from the database. The average retrieval precision and recall are defined as P_p and P_R, respectively, as follows:

$$P_p = \frac{\sum_{i=1}^{M} n_i}{\sum_{i=1}^{M} N}, \tag{8}$$

$$P_R = \frac{\sum_{i=1}^{M} n_i}{\sum_{i=1}^{M} K}, \tag{9}$$

where M is the total number of images in the database, and K denotes the number of images belonging to the same category as the query image in the dataset. In our experiments, $K = 100$ and n_i is the number of returned images belonging to the correct image class of the query image indexed by i.

4.2 Experimental results

4.2.1 Comparison between FCC and FCCS

In order to demonstrate the effect of the spatial information of scalar color codes on retrieval performance, we built a set of experiments based on the FCC and FCCS method. Tables 1 and 2 provide the retrieval results of the FCC and FCCS methods in terms of precision and recall.

In Table 1, the precision decreases as N grows to both FCC and FCCS, and the precision achieves its best value when N is set to 5. In contrast to precision, in Table 2, recall increases as N increases. This is because increasingly similar images in the retrieval results are returned when N increases. From Tables 1 and 2, we can also observe that no matter the value of N, the average precision of all the categories based on FCC is higher than that based on FCCS. By analyzing the retrieval performance of each image category, it is clear that the retrieval performance of the FCC method is better than that of the FCCS method in most categories, except for the dinosaur and flower categories.

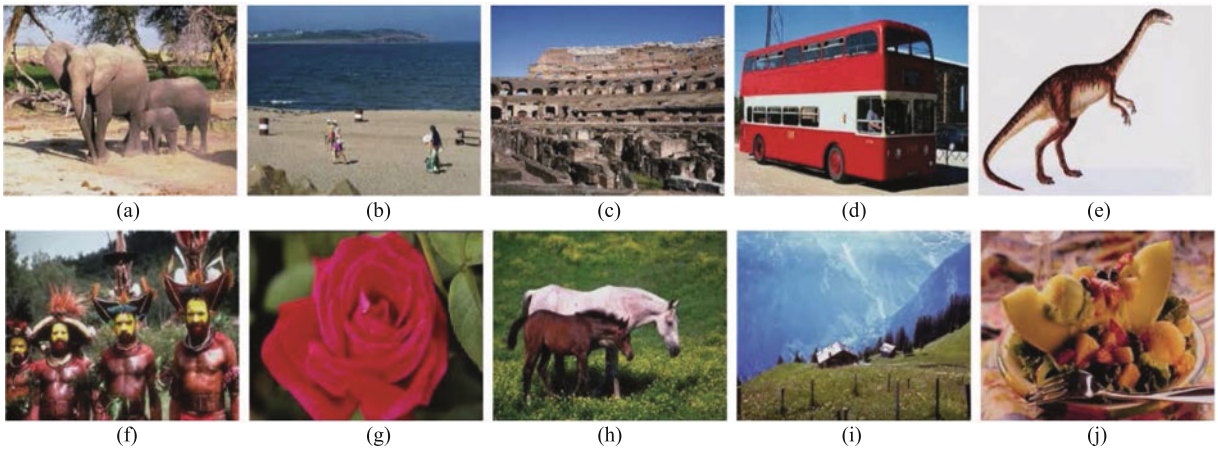


Fig. 7 Typical image in each class

Table 1 Comparison of the top N precision between FCCS and FCC

N	Method	Image category										Average
		1	2	3	4	5	6	7	8	9	10	
5	FCCS	0.558	0.558	0.546	0.600	1	0.686	0.874	0.954	0.536	0.442	0.678
	FCC	0.834	0.610	0.650	0.642	1	0.798	0.836	0.980	0.586	0.774	0.771
10	FCCS	0.457	0.494	0.430	0.495	1	0.570	0.800	0.923	0.444	0.325	0.594
	FCC	0.782	0.524	0.543	0.573	0.995	0.668	0.717	0.945	0.486	0.701	0.693
15	FCCS	0.425	0.449	0.391	0.463	1	0.496	0.769	0.887	0.403	0.297	0.558
	FCC	0.748	0.478	0.499	0.550	0.985	0.578	0.635	0.918	0.451	0.645	0.649
20	FCCS	0.407	0.422	0.358	0.435	1	0.452	0.729	0.867	0.383	0.276	0.533
	FCC	0.721	0.455	0.462	0.525	0.984	0.516	0.686	0.899	0.431	0.612	0.629
25	FCCS	0.385	0.405	0.330	0.428	1	0.416	0.708	0.836	0.344	0.265	0.512
	FCC	0.708	0.440	0.432	0.503	0.982	0.475	0.532	0.879	0.416	0.589	0.596

Table 2 Comparison of the top N recall between FCCS and FCC

N	Method	Image category										Average
		1	2	3	4	5	6	7	8	9	10	
5	FCCS	0.028	0.029	0.027	0.030	0.050	0.034	0.044	0.048	0.027	0.022	0.034
	FCC	0.042	0.031	0.033	0.032	0.050	0.040	0.042	0.049	0.029	0.039	0.039
10	FCCS	0.046	0.049	0.043	0.050	0.100	0.057	0.080	0.092	0.044	0.033	0.059
	FCC	0.078	0.052	0.054	0.057	0.100	0.067	0.072	0.095	0.049	0.070	0.069
15	FCCS	0.064	0.067	0.059	0.069	0.150	0.074	0.115	0.133	0.061	0.045	0.084
	FCC	0.112	0.072	0.075	0.083	0.148	0.088	0.095	0.138	0.068	0.097	0.098
20	FCCS	0.082	0.085	0.072	0.087	0.200	0.091	0.146	0.173	0.077	0.055	0.107
	FCC	0.142	0.091	0.092	0.105	0.197	0.103	0.117	0.179	0.086	0.123	0.124
25	FCCS	0.096	0.101	0.083	0.105	0.250	0.104	0.177	0.209	0.091	0.066	0.128
	FCC	0.177	0.110	0.108	0.126	0.246	0.119	0.133	0.219	0.104	0.147	0.149

We attribute the difference in retrieval performance to the distinct design of FCC and FCCS. FCCS defines two images that have similar color codes in the same position as similar, whereas FCC describes two images as similar as long as they have similar color codes, regardless of the location. Thus, the similarity between two images defined by FCCS is stricter than that by FCC. When images include the objects such as people, buses, and mountains, which are usually located in different positions, the retrieval precision of FCCS is undoubtedly lower than that of FCC. In the dinosaur and flower categories, these objects are often located in the center of the images, which means similar color information and coordinate information are represented. The leaf or grass image blocks in the flower images often mislead the FCC method into recognizing the flower images as horses or mountains categories as the most similarscalar color codes, while the FCCS method experiences no confusion in this regard. FCCS considers the similarity between the flower, horse, and mountain images to be very small due to the difference in coordinate information. Therefore, FCCS outperforms FCC in the dinosaur and flower categories. FCCS even maintains 100% precision with the dinosaur images when N increases.

Because the coordinate information provides extra retrieval restriction in the FCCS method, it also limits the application range of the FCCS. From the analysis, we can draw the conclusion that FCCS is suitable for image categories that have similar objects in similar locations, while FCC is more effective for the categories where objects have varying positions.

4.2.2 Results comparison with HSV

In this section, in order to show the improvement introduced by the proposed system, we compare the results of the performed experiment with those of the baseline method. Figure 8 presents the top 16 retrieved images using three methods: HSV, FCC, and FCCS. Every three adjacent rows serve as a comparison group, where the top row corresponds to the HSV results, the middle row corresponds to the FCCS results, and the bottom row corresponds to the FCC results. The first image in each row is the query image and first return images. In addition, the returned images not belonging to the corresponding category are outlined to provide a clear and quick comparison. As shown in Fig. 8, the results of the HSV color space are obviously worse than those of the FCC and FCCS methods. Specifically, when retrieving flower images, all returned images from the FCC and FCCS methods belonged to the flower category, i.e., FCC and FCCS both achieved 100%

precision, while HSV returned two images belonging to the people category. Similarly, the FCCS and FCC methods successfully returned all of the top 16 dinosaur images, whereas the results obtained from the HSV undesirably contained a beach image. These images featured color information that is extremely similar to that seen in the dinosaur images.

Furthermore, for the beach category of images, the HSV results only contained six beach images, while the FCCS and FCC returned 14 and 13 correct images, respectively. These comparisons demonstrate that the proposed color space clustering quantization outperforms the traditional quantization method on the HSV color space, and the CCS has better ability in color distinguishing than the quantized HSV color space. All these examples demonstrate the superiority of the proposed system, which combines color space clustering quantization and color FCC.

Figure 9 presents the statistics of the results, which demonstrates the average retrieval precision and recall among the

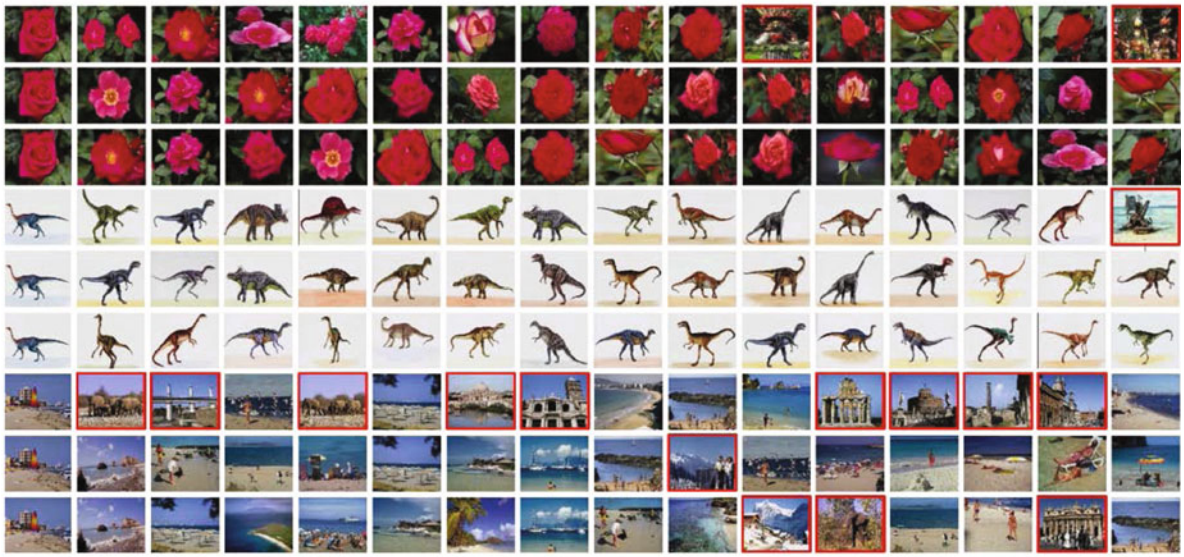


Fig. 8 Retrieval result examples of HSV, FCCS and FCC

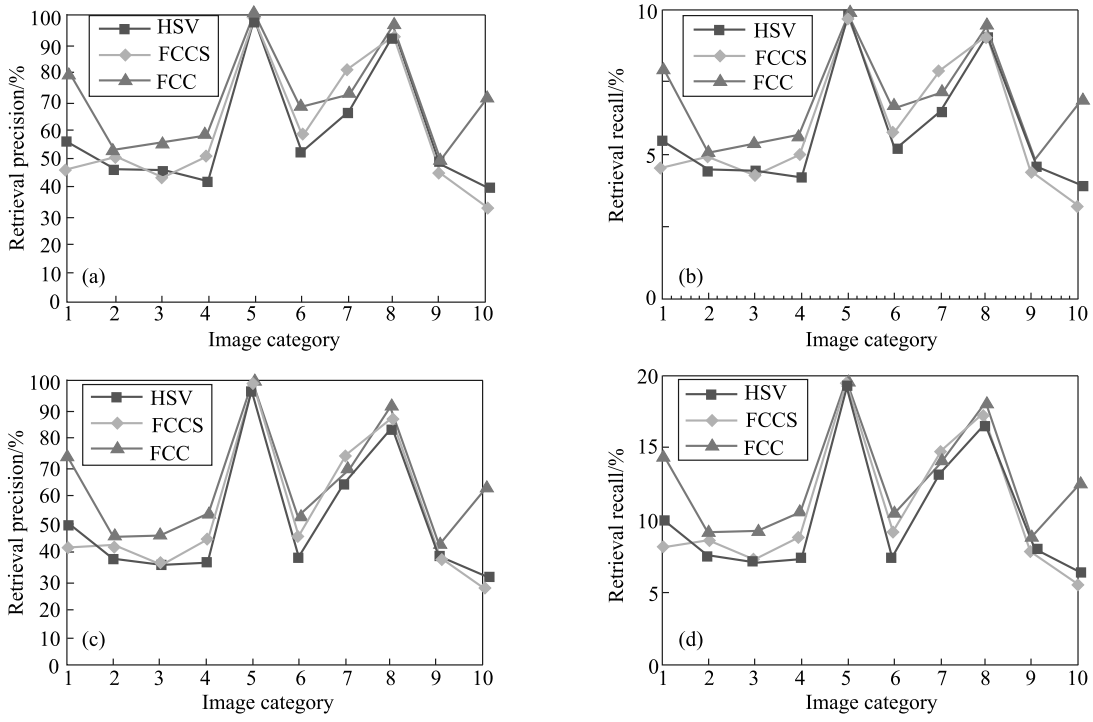


Fig. 9 The comparison of retrieval performance among HSV, FCC, and FCCS. (a) Retrieval precision (N = 10); (b) retrieval recall (N = 10); (c) retrieval precision (N = 20); (d) retrieval recall (N = 20)

HSV, FCCS, and FCC methods. As each category consists of 100 images of the same subject, when N is set to 10 and 20, N is proportional to K. According to definitions of the average retrieval precision and recall in Eqs. (8) and (9), the recall is also in proportion to precision. Thus, recall and precision have the same distribution. We can observe from Fig. 9 that for each image category, both the retrieval precision and recall of the FCC method outperform that of HSV to different degrees. For instance, when retrieving the beach, bus, and elephant images, both the retrieval precision and recall of the FCC method are close to those of HSV. Moreover, the superiority of FCC is apparent for people and food images, mainly due to the various colors presented in these two image categories. The retrieval precision and recall of FCCS are lower than those of HSV when retrieving some individual categories, such as people and food. This is because FCCS requires a similar color code and similar coordinate information, but the positions of the objects in the people and food images are different. Although FCCS shows poor performance for the specific image category, it is also obvious that the FCCS method performs better than the HSV overall. Over the whole dataset, when $N = 10$, FCCS improves the average precision of HSV from 57.9% to 59.4%, and FCC achieves 69.3% average precision, which is 11.4% higher than that of HSV. When $N = 20$, FCCS improves the precision of HSV from 50.9% to 53.3%, and FCC increases average precision by 12.2%. The results demonstrate that the CCS is more effective compared to HSV when utilized for the images employed in this study, and the proposed system has successfully improved the retrieval performance.

5 Conclusion

In this study, we proposed a new color-based image retrieval framework that consists of two correlated algorithms, color space clustering quantization and FCC. Color space clustering quantization clusters the CIE1976L*a*b* color space to obtain the clustering color space representing true color images in 256 distinct colors. The proposed quantization method reduces the scale of the color space while keeping the mapped image visually similar to the original RGB image. In addition, FCC and FCCS are proposed to encode the scalar color codes to character codes, in order to utilize the advantages of the inverted index to collaborate with the text-based search engine in CBIR. Due to its efficient calculation, the proposed system can be easily expanded to large-scale image retrieval. Experiments were conducted on a subset of Corel image datasets. We first evaluated the effect of spatial information on the retrieval performance, and then drew the following conclusion: FCCS worked well for the categories with similar objects in the same location, and FCC was suitable for the images that contained objects in various positions. Then, we compared the retrieval performance of our proposed method with that of the approach presented in [30], and the experimental results demonstrated that both FCCS and FCC outperformed the approach proposed in [30]. The color-based image retrieval framework proposed in the study is based on global image features that are not related to the object contained in the image. As a result, semantic errors are inevitable. In the following work, we may consider the combination of image retrieval and object detection methods.

Acknowledgements This work was supported in part by the National Natural Science Foundation of China (Grant No. 61370149), in part by the Fundamental Research Funds for the Central Universities (ZYGX2013J083), and in part by the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry.

References

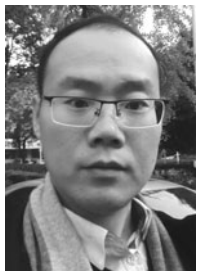
1. Datta R, Joshi D, Li J, Wang J Z. Image retrieval: ideas, influences, and trends of the new age. ACM Computing Surveys, 2008, 30(2): 5
2. Liu Y, Zhang D S, Lu G J, Ma W Y. A survey of content-based image retrieval with high-level semantics. Pattern Recognition, 2007, 40(1): 262–282
3. Smeulders A W M, Worring M, Santini S, Gupta A, Jain R. Content-based image retrieval at the end of the early years. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(12): 1349–1380
4. Priya R, Shanmugam T N. A comprehensive review of significant researches on content based indexing and retrieval of visual information. Frontiers of Computer Science, 2013, 7(5): 782–799
5. Bian W, Tao D C. Biased Discriminant Euclidean Embedding for Content-Based Image Retrieval. IEEE Transactions on Image Processing, 2010, 19(2): 545–554
6. Kato T. Database architecture for content-based image retrieval. Proceedings of SPIE: The International Society for Optical Engineering, 1992, 1662: 112–123
7. Tak Y S, Hwang E. Tertiary hash tree: indexing structure for content-based image retrieval. In: Proceedings of the 20th International Conference on Pattern Recognition. 2010
8. Liu Z, Li H Q, Zhang L Y, Zhou W G, Tian Q. Cross-indexing of binary SIFT codes for large-scale image search. IEEE Transactions on Image Processing, 2014, 23(5): 2047–2057
9. Kong G P, Dong L, Dong W P, Zheng L, Tian Q. Coarse2Fine: two-layer fusion for image retrieval. 2016, arXiv:1607.00719
10. Chang R, Xiao Z M, Wong K S, Qi X J. Learning a weighted semantic manifold for content-based image retrieval. In: Proceedings of the 19th IEEE International Conference on Image Processing. 2012
11. Zhao F, Huang Y Z, Wang L, Tan T N. Deep semantic ranking based

hashing for multi-label image retrieval. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2015, 1556–1564

12. Ma H, Zhu J K, Lyu M R T, King I. Bridging the semantic gap between image contents and tags. IEEE Transactions on Multimedia, 2010, 12(5): 462–473
13. Barrett S, Chang R, Qi X J. A fuzzy combined learning approach to content-based image retrieval. In: Proceedings of IEEE International Conference on Multimedia and Expo. 2009, 838–841
14. Liang Y, Dong L, Xie S S, Lv N, Xu Z Y. Compact feature based clustering for large-scale image retrieval. In: Proceedings of IEEE International Conference on Multimedia and Expo Workshops. 2014, 1–6
15. Chen W T, Liu W C, Chen M S. Adaptive color feature extraction based on image color distributions. IEEE Transactions on Image Processing, 2010, 19(8): 2005–2016
16. Xiao Z M, Qi X J. Block-based long-term content-based image retrieval using multiple features. In: Proceedings of IEEE International Conference on Multimedia and Expo. 2011
17. Xie B J, Liu Y, Zhang H, Yu J. Efficient image representation for object recognition via pivots selection. Frontiers of Computer Science, 2015, 9(2): 383–391
18. Chang R, Qi X J. A hierarchical manifold subgraph ranking system for content-based image retrieval. In: Proceedings of IEEE International Conference on Multimedia and Expo. 2013
19. Chai L S, Qin Z, Zhang H G, Guo J, Shelton C R. Re-ranking using compression-based distance measure for content-based commercial product image retrieval. In: Proceedings of the 19th IEEE International Conference on Image Processing. 2012, 1941–1944
20. Xu C, Li Y X, Zhou C, Xu C. Learning to rerank images with enhanced spatial verification. In: Proceedings of IEEE International Conference on Image Processing. 2012, 1933–1936
21. Zhang L N, Shum H P H, Shao L. Discriminative Semantic Subspace Analysis for Relevance Feedback. IEEE Transactions on Image Processing, 2016, 25(3): 1275–1287
22. Lin X F, Gokturk B, Sumengen B, Diem V. Visual search engine for product images. In: Proceedings of SPIE, Multimedia Content Access: Algorithms and Systems II. 2008
23. Xu W G, Zhang Y F, Lu J J, Li R, Xie Z H. A framework of Web image search engine. In: Proceedings of IEEE International Joint Conference on Artificial Intelligence. 2009, 522–525
24. Jiang F, Hu H M, Zheng J. A hierarchal BoW for image retrieval by enhancing feature salience. Neurocomputing, 2016, 175: 146–154
25. Michael J S, Dana H B. Color indexing. International Journal of Computer Vision, 1991, 7(1): 11–32
26. Ke V D S, Gevers T, Snoek C G. Evaluating color descriptors for object and scene recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(9): 1582–1596
27. Luo X Y, Zhang J, Dai Q H. Hybrid fusion and interpolation algorithm with near-infrared image. Frontiers of Computer Science, 2015, 9(3): 375–382
28. Zhang Y G, Gao L J, Gao W, Liu J. Combining color quantization with curvelet transform for image retrieval. In: Proceedings of International Conference on Artificial Intelligence and Computational Intelligence. 2010, 474–479
29. Pun C M, Wong C F. Image retrieval using a novel color quantization approach. In: Proceedings of the 9th IEEE International Conference on Signal Processing. 2008, 773–776
30. Zhang H, Hu R M, Chang J, Leng Q M, Chen Y. Research of image retrieval algorithms based on color. In: Proceedings of International Conference on Artificial Intelligence and Computational Intelligence. 2011, 516–522
31. Zheng L, Wang S J, Liu Z Q, Tian Q. Packing and padding: coupled multi-index for accurate image retrieval. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2014, 1947–1954
32. Jégou H, Douze M, Schmid C. Hamming embedding and weak geometric consistency for large scale image search. In: Proceedings of European Conference on Computer Vision. 2008, 304–317
33. Dong L, Liang Y, Kong G, Zhang Q N, Cao X C, Izquierdo E. Holons visual representation for image retrieval. IEEE Transactions on Multimedia, 2016, 18(4): 714–725
34. Jégou H, Douze M, Schmid C. Improving bag-of-features for large scale image search. International Journal of Computer Vision, 2010, 87(3): 316–336
35. Arandjelović R, Zisserman A.. Three things everyone should know to improve object retrieval. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2012, 2911–2918
36. Chen Y, Hao P W. Optimal transform in perceptually uniform color space and its application in image retrieval. In: Proceedings of the 7th IEEE International Conference on Signal Processing. 2004, 1107–1110
37. McQueen J. Some methods for classification and analysis of multivariate observations. In: Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability. 1967, 281–297
38. Chen T W, Chen Y L, Chen S Y. Fast image segmentation based on *K*-Means clustering with histograms in HSV color space. In: Proceedings of the 10th IEEE Workshop on Multimedia Signal Processing. 2008, 322–325
39. Jin S. The design and research of personalized search engine based on Solr. Dissertation for the Master Degree. Beijing: Beijing University of Chemical Technology, 2011

Le Dong received the PhD degree in electronic engineering and computer science from Queen Mary, University of London, UK in 2009. She is an associate professor in University of Electronic Science and Technology of China, China. Her research interests include computer vision, big data analysis, and biologically inspired system.

Wenpu Dong is an undergraduate majoring in computer science and technology at University of Electronic Science and Technology of China, China. His research interest is mainly on image retrieval.



Ning Feng is a PhD student majoring in computer science and technology at University of Electronic Science and Technology of China, China. His research interests are computer vision and image segmentation.

Long Chen received his master degree in University of Electronic Science and Technology of China, China in 2013. He is now working at Chengdu FunMi Technology Company. His research interests are computer vision and image retrieval.

Mengdie Mao is an undergraduate majoring in computer science and technology at University of Electronic Science and Technology of China, China. Her research interest is mainly on image retrieval and deep learning.

Gaipeng Kong is an undergraduate majoring in computer science and technology at University of Electronic Science and Technology of China, China. Her research interest is mainly on image retrieval.