Color-Structured Image Search

Jingdong Wang, Xian-Sheng Hua and Yinghai Zhao

MSR-TR-2009-82

July 7, 2009

Abstract

In this paper, we propose a novel interactive image ranking technique, color-structured image search, by exploiting color spatial relation information, and demonstrate it in image research results refinement. This technique enables users to draw a few color strokes, called target color structure, to indicate the intent. Beyond the conventional image retrieval methods based on query by image content that simply view the target color structure as an image and compute the similarity merely according to the spatial correspondence criterion, the proposed method represents the target color structure using a set of spatially relational colors, and evaluates the similarities between target color structure and image color structure by exploiting their spatial correspondence information, and an extra novel criterion, the consistency of spatial relation of colors implied in target color structure, and even offering a scheme to handle the uncertainty of target color structure. This similarity evaluation process can be accomplished efficiently. Moreover, a convenient interactive interface is presented to allow users to specify target color structure flexibly. Experimental results demonstrate the effectiveness and efficiency of the proposed approach.

Index Terms

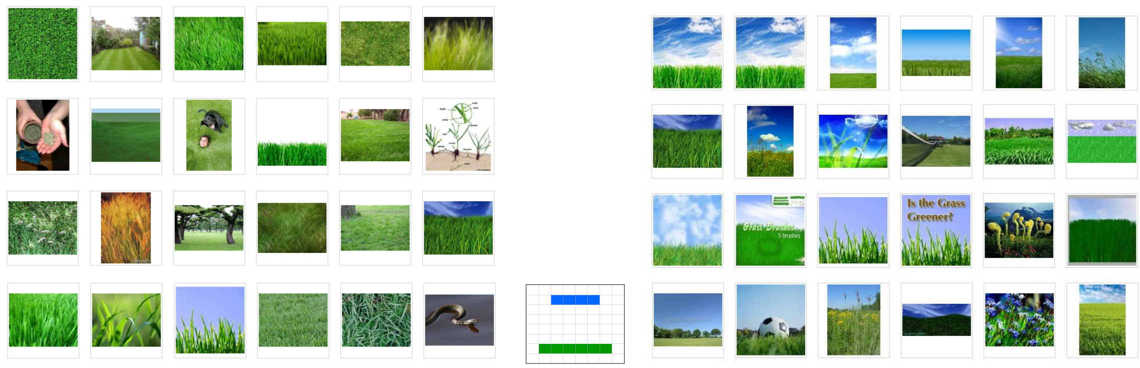
Color structure, image search

I. Introduction

With the rapid development of commercial image search engines, people can easily get a large amount of images by inputting a text query. However, existing search engines return search results only using the relevance of text information (e.g., the surrounding text) of images with the text query but without utilizing the rich and useful image visual information. Therefore, many efforts have been made to refine search results by leveraging image content.

J. Wang and X-S. Hua are with Media Computing Group, Microsoft Research Asia. E-mail: {jingdw, xshua}@microsoft.com

Y. Zhao was an intern with Microsoft Research Asia.



(a) (b) (c)

Fig. 1. Color-structured image search. (a) shows the original search results, (b) shows the target color structure, and (c) shows the refined results.

Existing methods for content based image search result refinement can be roughly divided into two types: visual reranking [7], [9] and implicit intent based refinement [3]. Visual reranking aims to make use of the visual information to reorder the search results without users' intervention such that visually similar images are ordered consistently. IntentSearch [3], an implicit intent based refinement, provides a mechanism to allow users to select a few images of interest, and automatically mine the users' intent to reorder image search results. In practice, it is essentially very difficult to mine the intent from the selected images.

It is promising to provide an effective and efficient scheme to allow users to explicitly express the intent and refine the search results. It is nontrivial to deliver a general way to allow users to depict the intent, but it is very practical to offer an easy manner for users to present some specific intents. This will bring double-side advantages. On the user side, users can explicitly express the goal, and on the search engine side, the intent ambiguity is reduced, and the intent is clearly grasped for reordering the images. This scheme is referred as the third type of image search result refinement, called explicit intent based refinement.

In this paper, we investigate explicit intent based refinement allowing users to explicitly express the intent in an intuitive manner. The goal of searching images is usually to utilize the returned images to help their work, e.g., used as an illustration in the presentation slides or poster design. Intention can be described in different ways. Color spatial arrangement is viewed as a straightforward way to describe a desired image. For example, given image search results shown in Fig. 1(a), users may be interested in images with blue sky on the top and green grasses on the bottom. But it is difficult to only use a text query to meet such requirement; it is time-costly to scroll search results to manually find such images; it is also difficult to merely use example images to indicate the requirement because the intent from example images is ambiguous and not easily automatically mined. In this paper, therefore, we propose a novel technique, color structured image search, which provides a convenient mechanism to allow users to scribble a few color strokes to indicate interests of both color and spatial structure and then promotes the images that are more consistent with the scribbled target color structure. The proposed new method is beyond the conventional image retrieval methods based on query by image content, and exploits the consistency of spatial relation of colors implied in the target color structure and even offering a scheme to handle the uncertainty of target color structure besides their spatial correspondence information investigated in the conventional methods. Using the proposed technique, we can have reordered top 24 results as shown in Fig. 1(c) according to the target color structure specified by some user, e.g., as shown in Fig. 1(b).

A. Related Work

There are many efforts that have been made to refine image search results. As aforementioned, there exist two types of refinement methods, including visual reranking [7], [9], implicit intent based search [3]. Visual reranking is to reorder images only according to image visual content, but it does not present a way to introduce users' intent into refinement. Intentsearch tries to make use of users' intent, but the intent is given in an implicit way by selecting several images, and has to be carefully mined, which is very difficult in practice. Differently, the method proposed in this paper aims to deliver an explicit way to allow users to express the intent, specifically, color structure, and reorder the images according to the consistency of the colors and their spatial structure between target color structure and images.

In the conventional query by image content (QBIC) image retrieval methods, color content has been investigated. Here, we will review the most relevant techniques that specify the query by providing color strokes. Almost all these methods represent the whole image with little ability of distinguishing the spatial relation of different colors. Those methods roughly consist of two types: global color feature and color spatial correspondence, roughly reviewed in the following.

The simple technique is to allow users to merely select colors of interest or input a color image and return the images according to the global color feature [15]. Some online applications based on this technique have been delivered, e.g., multicolor search lab [1] and Color Fields Experimental Colr Pickr [2]. However, this technique only makes use of the frequency feature (histogram) of interest colors to rank the images, without using any spatial information.

Color based retrieval by sketch, instead, allows users to scribble a color image, which provides the color-spatial information. This kind of method is usually to view the sketched image as an example image and directly evaluates the similarity by matching the most pronounced slabs of colors, i.e., spatial correspondence. More attention on this topic has been paid to pursuit fast retrieval techniques, including fast wavelet based matching [8], [18], sequenced multi-attribute tree [14], filtering out some irrelevant images [2], multi-precision querying [11], and etc. There are some web application examples, such as Retrievr[3] that is implemented using the fast multiresolution technique [8], and hermitage web site[4] that is based on IBM's query by image content [5]. Those methods only take into account spatial correspondence of individual colors across images, but without exploring spatial relation among different colors as a factor for similarity evaluation.

Besides, VisualSEEk [17], another color content based image retrieval method, tries to utilize spatial relation between colors to help image retrieval. But there are several differences between it and our approach. 1) They have different capability to make use of spatial relation. The capability of VisualSEEk is limited in that it adopts 2D string to approximate the spatial relation, while ours is direct to exploit the spatial relation using a set of spatially-relational colors. 2) They are applied to different scenarios. Our approach is novel for refining image search results using color structure. 3) They have different structure representation. VisualSEEk's spatial relation representation is a little complicated and is not suitable for real applications, while our representation is more compact, more efficient and effective for ranking, and hence more practical. 4) Our approach has the extra ability to deal with the uncertainty of target color structure, which reduces the user's labor of specifying it.

There are some other relevant techniques. The sketch-based image retrieval system [1], [12], [13] also provides a mechanism for users to sketch an image, but differently the shape consistency

[1] http://labs.ideeinc.com/multicolour/

[2] http://www.krazydad.com/colrpickr/

[3] http://labs.systemone.at/retrievr/

[4] http://www.hermitagemuseum.org/fcgi-bin/db2www/qbicLayout.mac/qbic?selLang=English

between sketched image and images in the database is adopted to search the images. The structure information among objects is also exploited for image retrieval [6], [4], but the structure is represented by a complicated graph model, leading to the difficulty of evaluating the similarity between the structures in both the accuracy and the computation.

B. Contributions

In this paper, we propose a novel technique, color-structured image search. This method offers a novel scheme to represent the color structure with a set of spatially-relational colors. This representation looks straightforward, but it is very sufficient to represent the color structure, particularly for the target color structure which is only a few color strokes specified by the user. Some powerful image representations, e.g., spatial pyramid matching [10], multi-precision histogram [11], and so on, are essentially unsuitable to represent the target color structure since there are only a few color strokes. Moreover, the proposed color structure representation makes the similarity evaluation more powerful. The proposed novel similarity evaluation scheme takes into consideration color spatial correspondence across images, which is exploited in the conventional image retrieval techniques based on query by image content, and additionally the spatial relation of different colors within an image is even investigated to compare the images, which is less investigated in the conventional methods. Furthermore, the proposed method does not need the user to specify the target color structure accurately, and our method offers a scheme to deal with the rough color structure. Besides, a convenient interactive interface is presented to allow users to specify target color layout flexibly.

C. Organization

The remainder of this paper is organized as follows. Section II presents the color structure extraction method. Section III gives the similarity evaluation scheme. Section IV introduces the interest color structure specification mechanism. Section V reports experimental results to demonstrate the proposed approach. Finally, Section VI concludes this paper.

II. Color Structure Extraction

The proposed color structure extraction process, consists of two steps: extract representative colors for each cell and concatenate them into spatial structures with each corresponding to the same quantized color.

A. Representative Colors

We divide the image uniformly into $g \times g$ grids. For each grid (x, y), we aim to find dominant colors, $\mathcal{C}_{xy} = \{\mathbf{c}_{xy}^i\}$, to represent the colors in this grid. The dominant colors are found as follows. We calculate the frequencies for all the quantized colors appearing in this grid, and sort them according to their frequencies. We select the colors corresponding to the first k largest frequencies such that $f_{i-1} < 2f_i$, $\forall i \leqslant k$ and $i \neq 1$. In our experiment, we find that $g = 8$ works well. After extracting the colors in each grid, we find the representative colors for the whole image,

$$\mathcal{C} = \cup_{xy} \mathcal{C}_{xy}.$$

As preprocessing, we first transform the RGB (red, green, blue) color space into the HSV (hue, saturation, value) space, which is then quantized into $n_h \times n_s \times n_v$. In this paper, $n_h = 12$ $n_s = 4$, and $n_v = 4$, and then there are totally 192 quantized colors. The finer quantization for the hue channel is reasonable and works well because it is observed that human is perceptually more sensitive to hue variance. For each grid, in addition, we preform a median filtering to smooth the images on the HSV color space.

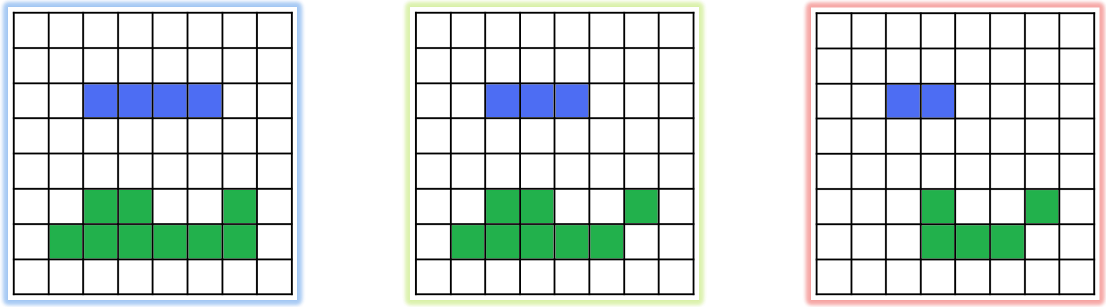
B. Color Structure

For each color $\mathbf{c} \in \mathcal{C}$, we compute its spatial structure as a binary vector $\mathbf{l}$ of dimension $g \times g$ such that

$$l_k = \begin{cases} 1 & \text{if } \mathbf{c}_{xy}^i = \mathbf{c},\ \exists \mathbf{c}_{xy}^i \in \mathcal{C}_{xy}, \\ 0 & \text{else}, \end{cases}$$

where k is one dimensional index corresponding to two dimensional indices (x, y). Intuitively, if color $\mathbf{c}$ appears in grid (x, y), the corresponding entry l_k is set to be 1, and 0 otherwise. Then in summary, the color structure of the image is represented by $\mathcal{I} = \{(\mathbf{c}_t, \mathbf{l}_t)\}_{\mathbf{c}_t \in \mathcal{C}}$. This whole algorithm for color structure extraction is easily implemented, and the computational cost is also very cheap, which will be justified in the experiment part. Moreover, the extracted color structure feature is an effective feature and is proved very valuable with the combination of the proposed similarity evaluation scheme that will be presented in Section III.

Since the spatial structure, $\mathbf{l}$, is a binary vector, we can represent it in a compact manner by several byte variables, i.e., $[b_1 \ \cdots \ b_n]^T$ with $n = \lceil \frac{g^2}{8} \rceil$, which will bring fast similarity evaluation by using a precomputed lookup table. For convenience, we still use the binary vector $\mathbf{l}$ in the following presentation.



(a) (b) (c)

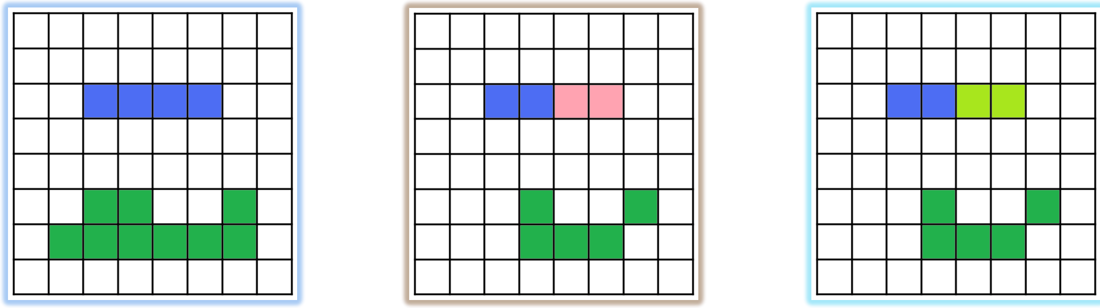
Fig. 2. Illustration of color spatial correspondence compatibility. (a) target color structure, (b) color image 1, and (c) color image 2.

III. Color Structure Similarity Evaluation

In this section, we present the similarity evaluation scheme between target color structure and image color structure. Suppose image color structure is denoted as $\mathcal{I} = \{(\mathbf{c}_k, \mathbf{l}_k)\}$, and target color structure is denoted as $\mathcal{Q} = \{(\mathbf{c}_q, \mathbf{l}_q)\}$. We made an investigation from user experience on the latent meaning for matching between target color structure and image color structure, and analyzed that it includes two aspects, spatial correspondence compatibility and spatial relation consistency. Therefore, we define the similarity from those two points: 1) Color spatial correspondence compatibility. It takes into account the similarity of individual colors within the associated spatial areas between target color structure and image color structure; 2) Color spatial relation consistency. It exploits the consistency of spatial relations of different colors between target color structure and image color structure. In addition, to tolerate the roughness of target color structure, we define another similarity: 3) Contextual color structure similarity. It aims to complement the roughness of target color structure, which makes users specify target color structure quite conveniently.

A. Spatial Correspondence Compatibility

This subsection defines the similarity, called color spatial correspondence compatibility, to measure the accordance of each individual color appearance on the associated areas between target color structure and image color structure. Let's see an illustrative example in Fig. 2 to justify this compatibility measure. Figs. 2(a), 2(b) and 2(c) show the target color structure, color image 1, and color image 2. It is implied that color image 1 in Fig. 2(b) is more compatible



(a) (b) (c)

Fig. 3. Illustration of color spatial relation consistency. (a) target color structure, (b) color image 1, and (c) color image 2.

to the target color structure Fig. 2(a) as the overlapped area for interest colors (blue and green) between target color structure and color image 1 is larger than the one between target color structure and color image 2.

Following this observation, we define such a color spatial correspondence compatibility measure. For each individual color in target color structure, we compare its spatial structure and the spatial structures of all the colors in the image, aggregate their similarities with the color similarity as the weight, and get the compatibility for this specific individual color structure between it and the image. Then, we can obtain the overall compatibility by adding the compatibilities together. Mathematically, this process is formulated as

$$\begin{aligned} s_a(\mathcal{Q}, \mathcal{I}) &= \sum_{q|\mathbf{c}_q \in \mathcal{C}_q} \frac{1}{Z_q} s_a((\mathbf{c}_q, \mathbf{l}_q), \mathcal{I}) \\ &= \sum_{q|\mathbf{c}_q \in \mathcal{C}_q} \frac{1}{Z_q} \sum_{k|\mathbf{c}_k \in \mathcal{C}_i} \mathrm{sim}(\mathbf{c}_q, \mathbf{c}_k)\,\mathrm{sim}(\mathbf{l}_q, \mathbf{l}_k), \end{aligned} \tag{1}$$

where $\mathcal{C}_q$ and $\mathcal{C}_i$ are the representative color sets for target color structure and image color structure, Z_q, a normalization variable, is equal to one norm of $\mathbf{l}_q$, $\|\mathbf{l}_q\|_1$, $s_a((\mathbf{c}_q, \mathbf{l}_q), \mathcal{I})$ is to evaluate the compatibility between the individual color structure, $(\mathbf{c}_q, \mathbf{l}_q)$, and the whole image color structure $\mathcal{I}$, $\mathrm{sim}(\mathbf{c}_q, \mathbf{c}_k)$ is the color similarity between $\mathbf{c}_q$ and $\mathbf{c}_k$, $\mathrm{sim}(\mathbf{l}_q, \mathbf{l}_k)$ is the spatial occurrence consistency between $\mathbf{l}_q$ and $\mathbf{l}_k$. The details of these two similarities will be discussed later.

B. Spatial Relation Consistency

This section defines another similarity to take into consideration the spatial relation of different colors between target color structure and image color structure. Let's consider the example in Fig. 3 to demonstrate the necessity of the color spatial relation consistency, in which color image 1 in Fig. 3(b) is different from color image 2 in Fig. 3(c) only in two cells. Merely using the compatibility evaluation in Eqn. (1), it is not easy to differentiate the two images as this similarity measure may lead to the similar (even same) scores. However, common sense tells us that color image 1 is more similar to target color structure because the aqua color in color image 2, appearing in the grids corresponding to the grids with target blue color in target color structure, is more similar to the green color, which is one target color in the bottom scribbled grids of target color structure. This case is called here the spatial relation inconsistency of different colors.

Therefore, we introduce a color spatial relation consistency measure. Intuitively, it is preferred that the color, appearing in the image region corresponding to one target color region, should be less similar to other target colors, otherwise it will be in conflict with the user intent in color spatial relation. Hence, we check the spatial structure associated with each target color to compute the spatial consistency between it and image color structure, and accumulate them together to obtain the overall consistency score. Mathematically, this spatial relation consistency is defined as

$$\begin{aligned} s_s(\mathcal{Q},\mathcal{I}) &= \sum_{q|\mathbf{c}_q \in \mathcal{C}_q} \frac{1}{Z_q} s_s((\mathbf{c}_q,\mathbf{l}_q),\mathcal{I}) \\ &= \sum_{q|\mathbf{c}_q \in \mathcal{C}_q} \frac{1}{Z_q} \sum_{k|\mathbf{c}_k \in \mathcal{C}_i} \mathrm{asim}(\mathbf{c}_q,\mathbf{c}_k,\mathcal{C}_q)\,\mathrm{sim}(\mathbf{l}_k,\mathbf{l}_q) \\ &= \sum_{q|\mathbf{c}_q \in \mathcal{C}_q} \frac{1}{Z_q} \sum_{k|\mathbf{c}_k \in \mathcal{C}_i} [\min_{q'|\mathbf{c}_{q'} \in \mathcal{C}_q, q' \neq q} (\mathrm{sim}(\mathbf{c}_q,\mathbf{c}_k) - \mathrm{sim}(\mathbf{c}_{q'},\mathbf{c}_k))\,\mathrm{sim}(\mathbf{l}_k,\mathbf{l}_q)], \end{aligned} \tag{2}$$

where $Z_q = \|\mathbf{l}_q\|_1$, $\mathrm{asim}(\mathbf{c}_q,\mathbf{c}_k,\mathcal{C}_q)$ is the similarity between target color $\mathbf{c}_q$ and color $\mathbf{c}_k$, adjusted by all the target colors $\mathcal{C}_q$. Combining it with the spatial occurrence consistency $\mathrm{sim}(\mathbf{l}_k,\mathbf{l}_q)$, the measure will reflect the consistency degree between the spatial relations of target colors in target color layout and the ones of colors in the image.

C. Contextual Color Structure Similarity

Practice shows that users often present a rough target color structure. Thus it is insufficient for color structure similarity evaluation to only involve the scribbled regions in target color structure. Let's see an example in Figs. 4(b) and 4(c). It is clear that their similarities using the above two measures are the same if target color structure is given as shown in Fig. 3(a). However, common sense will tell us that color image 2 in Fig. 4(c) is more similar than color image 1 in Fig. 4(b).

Therefore, we propose a scheme to “generalize” target color structure. We propagate target color structure by viewing the scribbled grids as seeds. Specifically, we adopt the front propagation algorithm [16] to propagate the seeded colors. To differentiate the propagated color structure from the original color structure, we assign different weights to the propagated grids according to their distances from the seed grids. For computational efficiency, we assign only two different weights for the propagated structures, i.e., set the weight as 0.25 for the near structures, and 0.125 for the far structure. Fig. 4(a) shows a propagation example, in which the weights are depicted by the color saturation. Let’s denote the propagated weighted color structure as $\bar{\mathcal{Q}} = \{(w_q, \mathbf{c}_q, \bar{\mathbf{l}}_q)\}$ with excluding the original scribbled grids. For example, in the propagated target color structure in Fig. 4(a), the propagated blue grids is represented by two types of components, with different weights and spatial structures for the same color. The similarity, w.r.t. the propagated color structure, is defined as

$$s_c(\mathcal{Q}, \mathcal{I}) = s'_a(\bar{\mathcal{Q}}, \mathcal{I}) + s'_s(\bar{\mathcal{Q}}, \mathcal{I}), \tag{3}$$

where $s'_a(\bar{\mathcal{Q}}, \mathcal{I})$ is contextual color spatial correspondence compatibility, and $s'_s(\bar{\mathcal{Q}}, \mathcal{I})$ is contextual color spatial relation consistency. They are defined as follows,

$$s'_a(\bar{\mathcal{Q}}, \mathcal{I}) = \sum\nolimits_{q|\mathbf{c}_q \in \bar{\mathcal{C}}_q} \frac{w_q}{Z_q} s_a((\mathbf{c}_q, \bar{\mathbf{l}}_q), \mathcal{I}), \tag{4}$$

$$s'_s(\bar{\mathcal{Q}}, \mathcal{I}) = \sum\nolimits_{q|\mathbf{c}_q \in \bar{\mathcal{C}}_q} \frac{w_q}{Z_q} s_s((\mathbf{c}_q, \bar{\mathbf{l}}_q), \mathcal{I}), \tag{5}$$

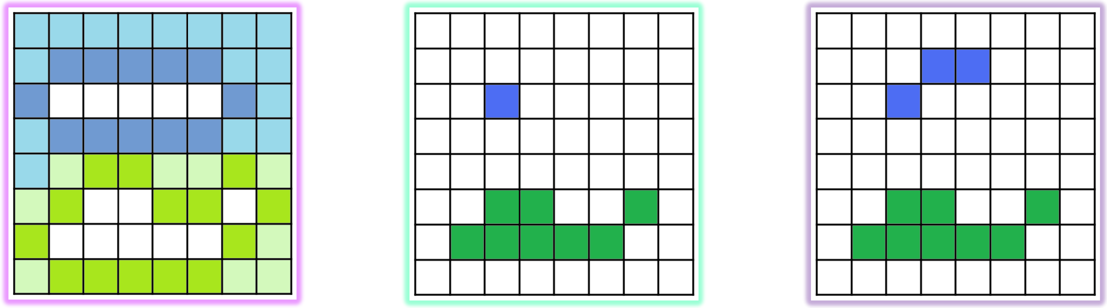
where Z_q, $s_a((\mathbf{c}_q, \bar{\mathbf{l}}_q), \mathcal{I})$, and $s_s((\mathbf{c}_q, \bar{\mathbf{l}}_q), \mathcal{I})$ are same defined as the previous two subsections, and $\bar{\mathcal{C}}_q$ is the weighted representative color set for the propagated target color structure.

Overall similarity measure

After obtaining all the above three measures, the whole similarity between $\mathcal{I}$ and $\mathcal{Q}$ is formulated as follows

$$s(\mathcal{Q}, \mathcal{I}) = \alpha_a s_a(\mathcal{Q}, \mathcal{I}) + \alpha_s s_s(\mathcal{Q}, \mathcal{I}) + \alpha_c s_c(\mathcal{Q}, \mathcal{I}), \tag{6}$$

where α_a, α_s and α_c are the weights for different similarity measures. We assign a less weight to $s_c(\cdot, \cdot)$ since it is a generalized one. In our experiments, we find it performs well when $\alpha_a = \alpha_s = 2 \times \alpha_c$.



(a) (b) (c)

Fig. 4. Illustration of contextual color structure similarity. (a) propagated target color structure, (b) color image 1, and (c) color image 2.

D. Color Similarity and Spatial Occurrence Consistency

Color similarity

The color similarity is evaluated in the HSV color space. According to the physical meaning of the HSV colors, the distance between $\mathbf{c}_i$ and $\mathbf{c}_j$, $d(\mathbf{c}_i, \mathbf{c}_j)$, is defined as

$$\sqrt{(s_i \cos h_i - s_j \cos h_j)^2 + (s_i \sin h_i - s_j \sin h_j)^2 + (v_i - v_j)^2},$$

where $[h_i \ s_i \ v_i]^T = \mathbf{c}_i$ and $[h_j \ s_j \ v_j]^T = \mathbf{c}_j$. Then, we can define the color similarity as

$$s(\mathbf{c}_i, \mathbf{c}_j) = (1 - d(\mathbf{c}_i, \mathbf{c}_j)/d_{\max})^{\beta}, \tag{7}$$

where $d_{\max} = \max_{mn} d(\mathbf{c}_m, \mathbf{c}_n)$ and $\beta = 2$ in this paper.

Experiences show that human is more sensitive to the H (hue) color channel. Hence we should pay more attention on this H channel, and define the similarity over the H channel,

$$s(h_i, h_j) = 1/\sigma\sqrt{2\pi}\exp(-(h_i - h_j)^2/(2\sigma^2)), \tag{8}$$

where $\sigma = \frac{\pi}{3}$ with h's range being $[0, 2\pi]$. This assignment of σ makes the similarity between different colors with larger difference of H channel much smaller. Then, the overall color similarity is as the following

$$\mathrm{sim}(\mathbf{c}_i, \mathbf{c}_j) = s(h_i, h_j) \times s(\mathbf{c}_i, \mathbf{c}_j). \tag{9}$$

As mentioned before, we have quantized the HSV space into $12 \times 4 \times 4$ colors, i.e., totally 192 colors. We precompute their similarities according to Eqn. (9) and store them in a lookup table of size 192×192. In this manner, we reduce the computation cost by only looking up this table instead of expensive online computation.

Spatial occurrence consistency

The spatial occurrence consistency between spatial structures is very straightforward and evaluated by the inner product, $\mathrm{sim}(\mathbf{l}_q, \mathbf{l}_k) = \mathbf{l}_q^T \mathbf{l}_k$. As mentioned before, $\mathbf{l}$ is a binary vector with entries equal to 0 or 1, and represented by a byte variable vector $[b_1 \ \cdots \ b_n]^T$. Hence the inner product can be fast evaluated by the *and* operation as the following

$$\mathbf{l}_q^T \mathbf{l}_k = \sum\nolimits_{i=1}^{n} \mathrm{count}(b_{qi} \wedge b_{ki}), \tag{10}$$

where the basic operation $\wedge$ is *and* operation on byte variables, and $\mathrm{count}(\cdot)$ is the number of set bits. There are total 256 different values for a byte variable. To avoid online computation, we precompute a lookup table of size 256×256, in which each entry is the number of set bits of the *and* result for two byte variables.

IV. Target Color Structure Specification

The proposed interface for target color structure specification consists of two modules: target color selection and target structure specification.

Target color selection

The target color selection scheme provides color palettes and allows the user to pick up a target color. Our main work on this color palette focuses on how to generate a color palette. We provide three color selection schemes.

1) Standard color palette. We organize the colors without using the information from image search results.
2) Color palette from image search results. We mine image search results, and present the most-frequent colors to form a color palette.
3) Color palette from an example image. We generate a color palette directly using an example image selected from search results. Specifically, the user can drag an image as the color palette, and then pick up the colors from the image.

Target structure specification

After selecting the target color, the user then can specify the target structure. We provide three specification schemes.

1) Free-scribbling. The user can freely draw strokes to indicate the target spatial structure.
2) Template structure. The user can select an target structure from a set of templates.

3) Structure from an example image. The user can select an example image from search results, and draw on this image to indicate the target color structure.

V. Experiments

To evaluate the proposed approach, we crawled images from the photo sharing engine, flickr, and image search engines (e.g., google, live, yahoo!). For each engine, we crawled top 1000 images from the returned images for a specific text query. There are totally 51 queries, which are selected from top popular queries such that those cover different types of queries, such as object, scene, portrait and etc.

We collected target color structures from the users, and asked users to help label the ground truth. To differentiate different relevance degrees, we adopted a graded relevance scale, and use four levels from level 0 to level 3. Level 3 corresponds to the most relevant, and level 0 the least relevant. We asked 5 users to manually label the relevance level for each interest color layout, and then selected the most frequent level as the final level for each image. To avoid any bias on the labeling, those users were selected such that they have no special knowledge on image search and are initially unknown about the proposed technique.

To evaluate the performance, we use the normalized discounted cumulative gain (nDCG) measure. DCG measures the usefulness, or gain, of a document based on its position in the result list. The gain is accumulated cumulatively from the top of the result list to the bottom with the gain of each result discounted at lower ranks. Two assumptions of DCG measure are that highly relevant documents are more useful when appearing earlier in a search engine result list (have higher ranks) and that highly relevant documents are more useful than marginally relevant documents, which are in turn more useful than irrelevant documents. Comparing a search engine performance from one query to the next cannot be consistently achieved using DCG alone, so the cumulative gain at each position (p) should be normalized across queries. This is done by sorting documents of a result list by the ground truth, producing an ideal DCG at position p. Mathematically, nDCG at position p is calculated as

$$\mathrm{nDCG}(p) = \mathrm{DCG}(p) / \mathrm{IDCG}(p), \tag{11}$$

$$\mathrm{DCG}(p) = \sum\nolimits_{i=1}^{p} r_i / \log_2(i+1), \tag{12}$$

where r_i is the graded relevance of the result at position i, and calculated as $r_i = 2^{c_i} - 1$, with c_i the ground truth level of the image at position i, $\mathrm{IDCG}(p)$ is an ideal DCG at position p. The nDCG values for all queries (all target color structures in our experiments) can be averaged to obtain a measure of the average performance for many queries.

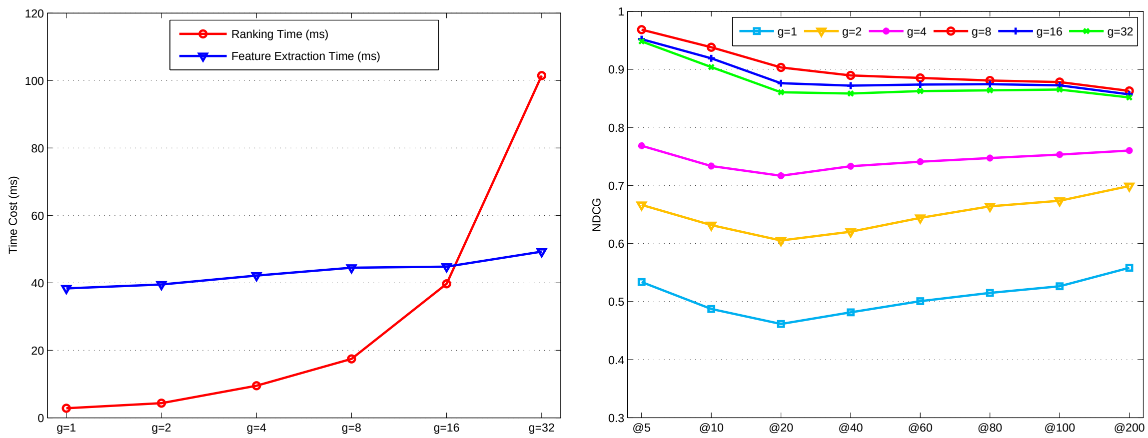
A. Sensitivity Analysis

In this subsection, we conducted experiments to show the performance variance with the grid dimension varying. We report and analyze the effects of the grid dimension on several aspects: computational cost for color structure feature extraction, computational cost for similarity evaluation and ranking performance.

For the computational cost of feature extraction, we average all the computational time over all the images. Ranking cost herein is the whole time cost from the time when the ranking process starts to the time when the ranked results are obtained. In other words, the whole process includes the analysis of target color structure, similarity evaluation between target color structure and images, and sorting procedure. The computational costs for feature extraction and image ranking are depicted in Fig. 5(a). From this figure, it can be seen that the extraction time cost is almost stable and that time cost increase for image ranking is stable when grid dimension is from 1 to 8 and time const increase is much bigger when grid dimension is changed from 8 to 32.

We also show nDCG performance variance with grid dimension changing, which is shown in Fig. 5(b). This figure shows the performance when rank positions $\{5, 10, 20 \cdots 200\}$ are considered. From this figure, it can be observed that the performance is improved when the grid dimension increases till 8 and that the performance is deteriorated a little when the grid dimension is chosen as 16 or 32. This deterioration is because in this case the color structure representation is too detailed and makes the distinctive structure characteristics ambiguous.

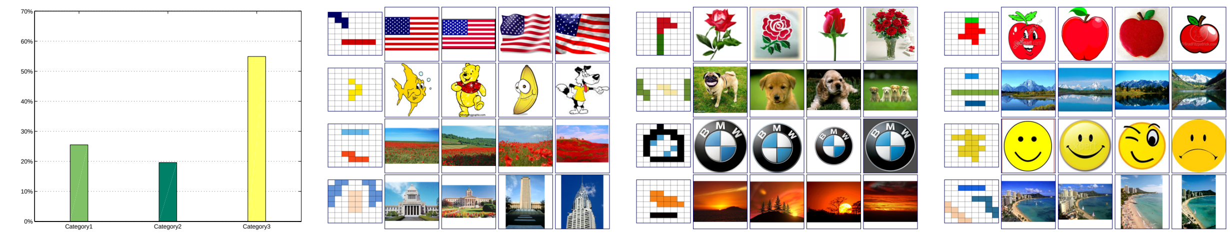
In our application, feature extraction is offline, and ranking is online. From the perspective of a real application, a proper grid dimension should be carefully selected so that the following properties are satisfied 1) the online ranking is conducted on the fly, 2) the performance is satisfactory, and 3) the offline feature extraction is not expensive. According to Fig. 5 and the above analysis, we conclude that it is better that grid dimension is selected to be 8. Then, the average computational costs for color structure extraction and image ranking are about 44 ms and 17 ms, respectively. In the following , all the experimental results are reported using grid dimension 8.



(a) (b)

Fig. 5. Sensitivity to grid dimension. (a) shows the extraction cost and online ranking cost with different grid dimensions, and (b) shows the nDCG performance with different grid dimensions.

We move to the last issue, storage cost for the color layout feature. It is better that the storage cost is cheap in that color layout feature extraction is performed when the image is crawled and is stored as metadata. In our implementation, we store the color layout feature color by color. Specifically, we store the color index, followed by the spatial layout feature, for each color. The layout feature can be directly saved as 8 bytes. However, from the statistics in our experiment, we observed that on average each spatial layout (binary vector of length 64) only has 6.3 ones. Therefore, we use a byte variable to indicate the index for the entry valued as one, which will reduce the storage cost from 8 bytes to 6.3 bytes. Moreover, the largest value for the grid index is 63, which leaves the top two bits of a byte variable unused. We can use those bits to separate the layout feature of different colors. In other words, we store the color index (one byte), followed by the corresponding spatial layout feature (several bytes), and particularly, for the last layout feature of each color we store it after increasing it by 64, which will be used as a flag to separate different colors. By comparison, this color-by-color scheme costs lower storage than grid-by-grid scheme that may need to store the color number and color indices for each grid. In summary, our scheme only needs to store the color indices $\{\mathbf{c}_i\}_{i=1}^n$ and the layout features$\{\mathcal{F}_i\}_{i=1}^n$ with $\mathcal{F}_i = \{f_j\}_{j=1}^{m_i}$ and m_i the number of ones, so the total storage cost is $n + \sum_{i=1}^n m_i$. From the statistics, $n = 15.67$, $\sum_{i=1}^n m_i = 98.65$, and hence the total cost is 114.32 bytes.



(a) (b) (c) (d)

Fig. 6. Target color structure categorization. (a) shows the statistics of different categories. (b), (c) and (d) show target structures of three categories and their corresponding images of interest.

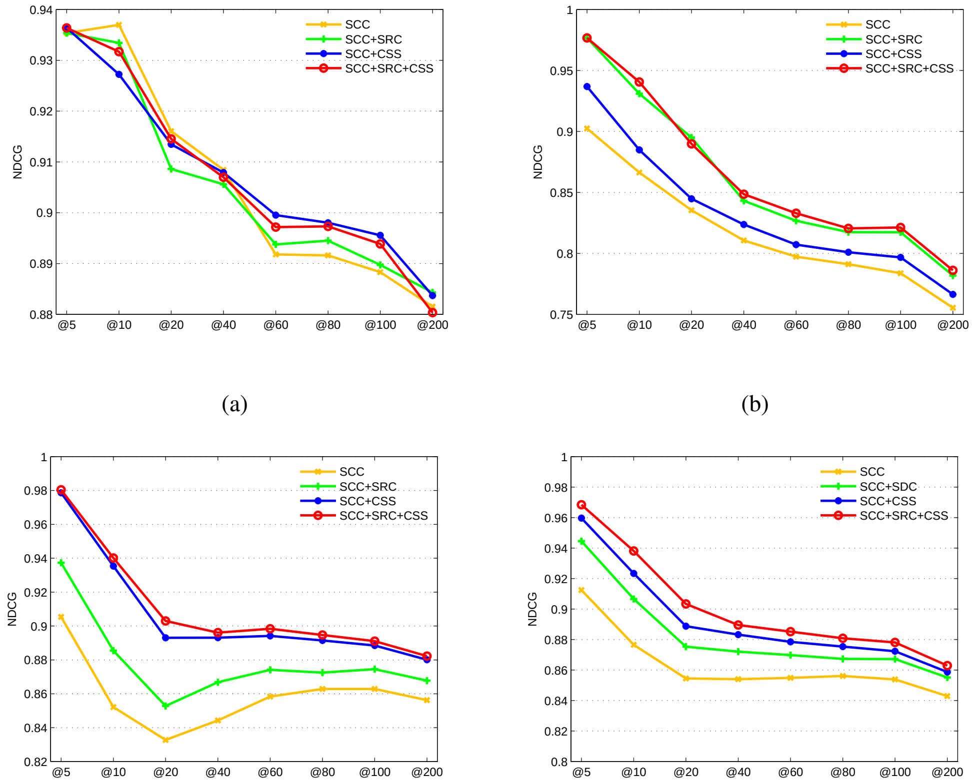
B. *Quantitative Evaluation*

In this subsection, we would like to present quantitative evaluation of our approach, by investigating the effects of different components in the similarity evaluation formulation and comparing it with several related methods.

Effects of three components

The color structure similarity measure is composed of three components: spatial correspondence compatibility, spatial relation consistency, and contextual color structure similarity. To investigate their effects on the performance clearly, we divide target color structures into several categorizes, which makes it clear where and how each component works. It should be noted that the target color structures are collected from the users that do not know anything about three components, i.e., they scribble the target color structures for those image search results without any special hints that favor the three components. Hence, the ratios of different categorizes reflect the practice, without specially designed for our approach. The target color structure categorization is shown in Fig. 6. From Fig. 6(a), it can be observed that realistic target structures indeed cover those three categories. Figs. 6(b), 6(c) and 6(d) visually illustrate different categories.

To justify those three components, we perform a serial of experiments using different combinations of the three components over four types of data sets: three categories and the whole data set. For convenience, we denote color spatial correspondence compatibility as SCC, color spatial relation consistency as SRC, and contextual color structure similarity as CSS. We consider four combinations, SCC, SCC + SRC, SCC + CSS, and SCC + SRC + CSS. The first combination inspects the effect of the basic color structure similarity measure, the second and third combinations aim to examine the effects of SRC and CSS, respectively, and the final combination is



(a) (b)

(c) (d)

Fig. 7. Illustration of effects of different components. (a), (b) and (c) show the performance over three categories of data sets. (d) shows the performance over the whole data set

actually the proposed approach. Such experimental results are shown in Fig. 7, from which we can have the following observations:

1) Fig. 7(a) shows the performance over the first category of target color structures. It can be observed that in such a case the basic component SCC is sufficient since the performances of all the methods are almost the same.
2) Fig. 7(b) shows the performance over the second category. This figure discloses that in some case the second component, SRC, makes effects.
3) Fig. 7(c) shows the performance over the third category. The comparison that SCC + CSS has the similar performance with SCC + SRC +CSS shows that CSS actually makes the key role of performance improvement.
4) Fig. 7(d) shows the performance over the whole data set. SCC + SRC + CSS, i.e., the proposed approach, gets the best performance.

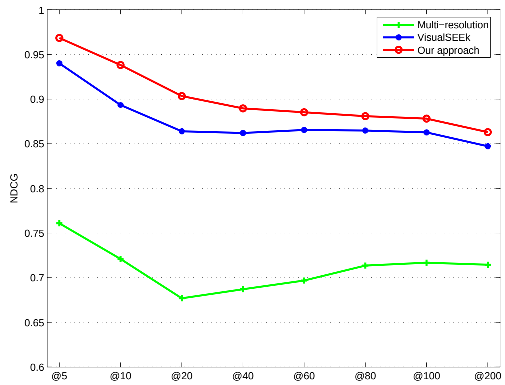


Fig. 8. Quantitative comparison with two representative related techniques.

	Our approach	VisualSEEk	Multi-resolution
spatial correspondence	Y	Y	Y
spatial relation	Y	Y	N
robustness to roughness	Y	N	N

TABLE I

COMPARISON OF OUR APPROACH WITH TWO REPRESENTATIVE RELATED TECHNIQUES.

Comparison

It is novel to refine image search results using color structure sensitive feature. But some existing techniques using color information in content based image retrieval, can be applied to image search result refinement. To justify the superiority of the proposed approach, we perform a quantitative comparison with two representative techniques, visualSEEk [17] and the multi-resolution based method [8].

The comparison over the whole data set is shown in Fig. 8. It can be observed that our approach is the best and that the multi-resolution based method gets the worst performance. Some analysis is given as follows. Our approach is superior over VisualSEEk at least in two aspects: 1) The color structure representation has larger ability to represent the spatial relation; and 2) Our similarity measure is more capable to discriminate the spatial relation consistency and hence is more compatible with users' intent. Compared with the multi-resolution based method, our approach exploits both spatial correspondence and spatial relation, while the multi-resolution based method only considered the color spatial correspondence. Tab. I summarizes their differences.

Fig. 9. Visual comparison between our approach and the spatial correspondence based approach. (a) shows the original top 10 images. (b) and (c) show two target color structures with the same colors and different spatial structures. (d) and (f) show the top 10 images of our approach, corresponding to the two target color structures. (e) and (g) show the top 10 images of the spatial correspondence based approach. It can be easily observed that the results of our approach meet with the target color structure and are better.

C. Visual Comparison

In this subsection, we present the visual results to compare the proposed approach and the conventional method that only considers the spatial correspondence.

In Fig. 9, we present a visual comparison to show that our approach has the ability to distinguish different color spatial relations. Fig. 9(a) shows the original top ten images corresponding to the query "beach". Figs. 9(b) and 9(c) show two target color structures from the users. It should be noted that these two target color structures have the same colors but different spatial relation. For the target color structure in Fig. 9(b), the user hopes that the yellow sand appears in the bottom left and the blue sky appears in the top right. The top ten images of our approach and the approach only exploiting spatial correspondence are shown in Fig. 9(d) and Fig. 9(e), respectively. We can see that the images in Fig. 9(d) meet with the user's intent, i.e., the yellow sand appearing in the bottom left and the blue sky appearing in the top right, while the ninth and tenth images in Fig. 9(e) are not consistent with the requirements. For the target color structure in 9(c), our results shown in Fig. 9(f) are very satisfactory, and obviously are better than the results in Fig. 9(g). In the results based on spatial correspondence shown in Fig. 9(g), the third, fourth, and fifth images are not consistent with the user's requirements. This comparison shows that our approach is superior over the spatial correspondence based method and can distinguish different spatial relations even with the same colors.

We present two additional visual comparisons shown in Fig. 10 to demonstrate that our approach can be applied to an interesting application, dominant color query. By specifying a single color structure, the user intends to find the images that the whole image has such color tones. Figs. 10(a) and 10(d) show the target color structures from the users over two image search results corresponding to "mountain" and "grass". Figs. 10(b) and 10(e) show the results of our approach, and Figs. 10(c) and 10(f). It can be observed that our results are better. The superiority benefits from that contextual color structure similarity which in some sense handle the roughness of the target color structure. It should be noted that some other method without exploiting the spatial information, e.g., color histogram over the whole image, can obtain the similar results. But with this experiment, we want to show that among the methods exploiting the spatial information, our approach can also perform better over such cases than other methods.

In addition, we present an experiment with the target color structure from an example image, as shown in Fig. 11. Figs. 11(a) and 11(b) show the selected example image and the specified color structure. Fig. 11(c) show the original and refined results, respectively.

D. User Study

In this subsection, we present the user study for the target color structure specification mechanism. The mechanism consists of two main modules: color selection and structure specification. For color selection, we present three schemes: 1) stand color palette, 2) color palette from image search results, and 3) color palette from an example image. We collect the feedbacks from 20 users. Those users are unknown of our system at the beginning and then are trained to use our system. After two days of using our system, the users are asked to give scores for each scheme

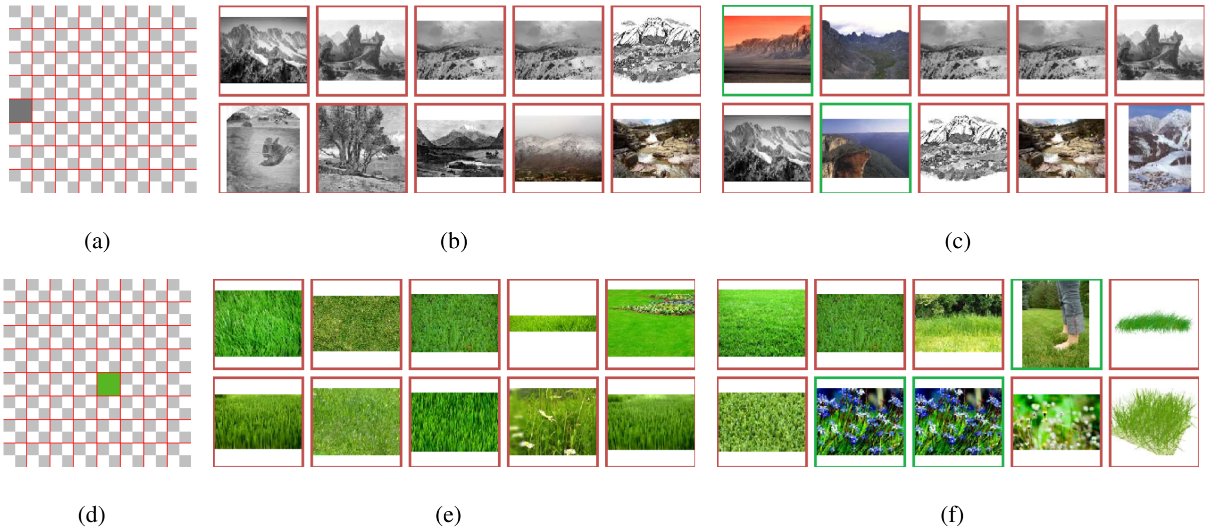


Fig. 10. Visual comparisons for single color query.

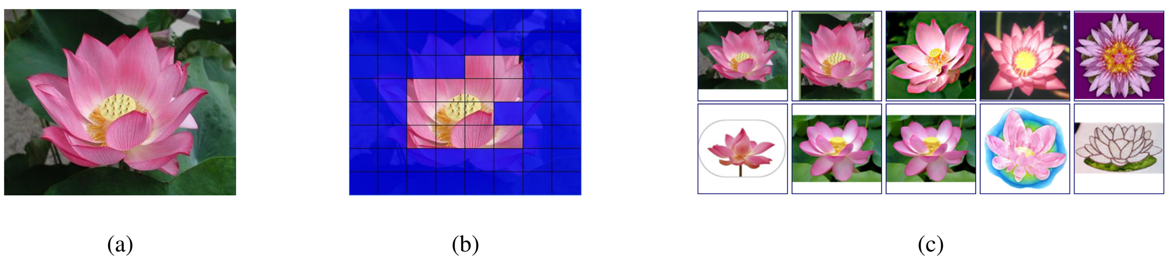
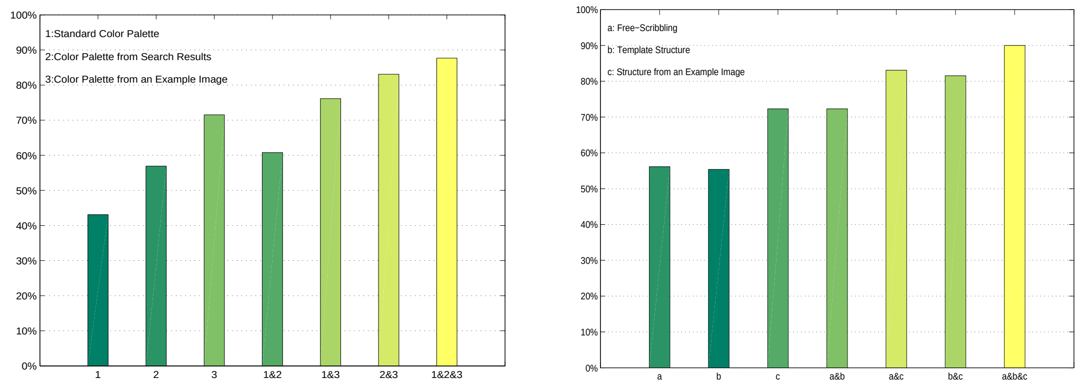


Fig. 11. Visual results with the target color structure from an example image. (a) shows the selected image, (b) shows the specified color structure, in which the center region without superposing blue color corresponds to target color structure, (c) shows the refined results.

and the combinations of schemes according to the convenience and frequencies of using them. The score is from level 1 to level 5, and level 5 is the best score. The average scores of all the combinations of schemes is shown Fig. 12(a). From this figure, it can be observed that the combination of three schemes are most useful and that color palette from an example image is most attractive among the tree individual schemes. From this sense, the color selection scheme is successful. For target color structure specification, we deliver three schemes: a) free-scribbling, b) template structures and c) structure from images. We perform a similar user study, and the result is shown Fig. 12(b). The overall scores show that the structure specification interface is also very convenient.



(a) (b)

Fig. 12. User study for color selection and structure specification interfaces. (a) shows average scores for different combinations of color selection schemes, and (b) shows average scores for different combinations of structure specification schemes.

VI. Conclusion

In this paper, we propose a novel interactive image search technique to meet with users' intent on the colors and their spatial structure. The proposed approach is superior over the conventional color based image retrieval method that only utilizes the spatial correspondence information because our approach additionally exploits the spatial relation of colors within an image to comparing and is robust to the roughness of target color structure specified by the user. The experimental results over real image search results demonstrate the efficiency and effectiveness.

References

[1] A. D. Bimbo and P. Pala. Visual Image Retrieval by Elastic Matching of User Sketches. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(2):121–132, 1997.

[2] T.-S. Chua, K.-L. Tan, and B. C. Ooi. Fast Signature-Based Color-Spatial Image Retrieval. In *ICMCS*, pages 362–369, 1997.

[3] J. Cui, F. Wen, and X. Tang. IntentSearch: Interactive On-line Image Search Re-ranking. In *ACM Multimedia*, 2008.

[4] J. Fauqueur and N. Boujemaa. Mental image search by boolean composition of region categories. *Multimedia Tools Appl.*, 31(1):95–117, 2006.

[5] M. Flickner, H. S. Sawhney, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker. Query by Image and Video Content: The QBIC System. *IEEE Computer*, 28(9):23–32, 1995.

[6] C.-S. Fuh, S.-W. Cho, and K. Essig. Hierarchical Color Image Region Segmentation for Content-Based Image Retrieval System. *IEEE Transactions on Image Processing*, 9(1):156–162, 2000.

[7] W. H. Hsu, L. S. Kennedy, and S.-F. Chang. Reranking Methods for Visual Search. *IEEE MultiMedia*, 14(3):14–22, 2007.

[8] C. E. Jacobs, A. Finkelstein, and D. Salesin. Fast Multiresolution Image Querying. In *SIGGRAPH*, pages 277–286, 1995.

[9] Y. Jing and S. Baluja. Pagerank for Product Image Search. In *WWW*, pages 307–316, 2008.

[10] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *CVPR (2)*, pages 2169–2178, 2006.

[11] S. Lin, M. T. Özsu, V. Oria, and R. T. Ng. An extendible hash for multi-precision similarity querying of image databases. In *VLDB*, pages 221–230, 2001.

[12] S. Matusiak, M. Daoudi, T. Blu, and O. Avaro. Sketch-Based Images Database Retrieval. In *Multimedia Information Systems*, pages 185–191, 1998.

[13] B. M. Mehtre, M. S. Kankanhalli, and W. F. Lee. Shape Measures for Content Based Image Retrieval: A Comparison. *Inf. Process. Manage.*, 33(3):319–337, 1997.

[14] B. C. Ooi, K.-L. Tan, T.-S. Chua, and W. Hsu. Fast Image Retrieval Using Color-Spatial Information. *VLDB J.*, 7(2):115–128, 1998.

[15] N. Sebe and M. S. Lew. Color-Based Retrieval. *Pattern Recognition Letters*, 22(2):223–230, 2001.

[16] J. Sethian. *Level Sets Methods and Fast Marching Methods*. Cambridge University Press, 2nd edition, 1999.

[17] J. R. Smith and S.-F. Chang. VisualSEEk: A Fully Automated Content-Based Image Query System. In *ACM Multimedia*, pages 87–98, 1996.

[18] J. Z. Wang, G. Wiederhold, O. Firschein, and S. X. Wei. Content-Based Image Indexing and Searching Using Daubechies' Wavelets. *Int. J. on Digital Libraries*, 1(4):311–328, 1997.