Interactive Image Search by Color Map

Jingdong Wang
Microsoft Research Asia
and
Xian-Sheng Hua
Microsoft Research Asia

The availability of large scale images from the Internet has made the research on image search attract a lot of attention. Text-based image search engines, e.g., Google/Microsoft Bing/Yahoo! image search engines using the surrounding text, have been developed and widely used. However, they suffer from an inability to search image content. In this paper, we present an interactive image search system, image search by color map, which can be applied to, but not limited to, enhance text-based image search. This system enables users to indicate how the colors are spatially distributed in the desired images, by scribbling a few color strokes, or dragging an image and highlighting a few regions of interest in an intuitive way. In contrast to the conventional sketch based image retrieval techniques, our system searches images based on colors rather than shapes, and we, technically, propose a simple but effective scheme to mine the latent search intention from the user's input, and exploit the dominant color filter strategy to make our system more efficient. We integrate our system to existing Web image search engines to demonstrate its superior performance over text-based image search. The user study shows that our system can indeed help users conveniently find desired images.

Categories and Subject Descriptors: H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*Retrieval models*

General Terms: Algorithm, Experimentation, Human Factors

Additional Key Words and Phrases: Image search, color map, intention map

1. INTRODUCTION

With the rapid development of multimedia on the Internet, users can easily access a very large amount of images. But these images are essentially orderless and unorganized, which makes browsing them quite difficult. This poses a big challenge for designing efficient indexing, search, and visualization methods to manipulate such large scale images. Recently, research efforts have been made on designing satisfactory image search schemes to help consumers find images.

Most existing image search engines rely on the tags or texts associated with images to index and search images. For example, Flickr image search allows users

Jingdong Wang and Xian-Sheng Hua are with the Media Computing Group, Microsoft Research Asia, Beijing, P.R. China.
E-mail: {jingdw, xshua}@microsoft.com

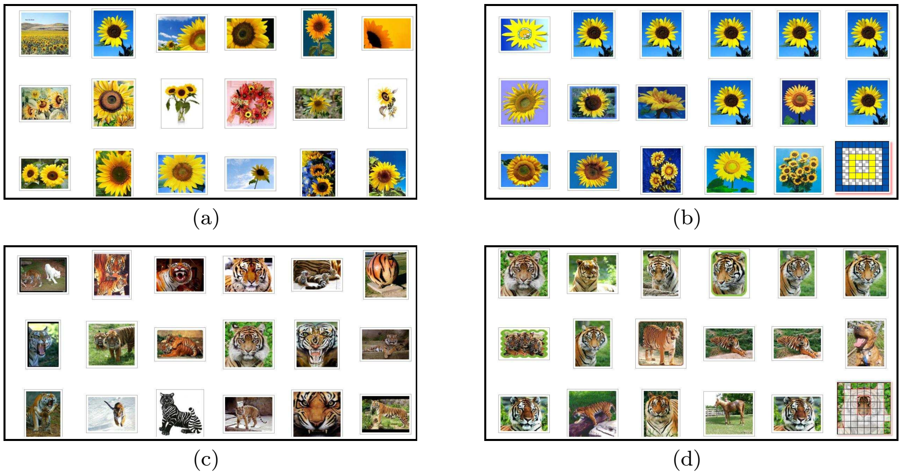


Fig. 1. Illustrative examples. The top row corresponds to the stroke scribbling case with the text query "sunflower". The bottom row corresponds to the image dragging case with the text query "tiger". (a) and (c) show the original search results. (b) and (d) show the reordered search results with the target color map shown in the bottom right.

to type a text query, and then matches it with images in the database, in which images are organized through human assigned tags. Google/Microsoft Bing/Yahoo! image search engines alternatively explore the texts surrounding the images to index them. However, such image search engines suffer from an inability to search image contents because the tags/texts are not sufficient or exact enough to describe the contents.

In the last century, many content based image retrieval (CBIR) techniques [Datta et al. 2008; Lew et al. 2006; Rui et al. 1999; Smeulders et al. 2000; Wang et al. 1997] were investigated to search according to image contents. Typically, those techniques require users to submit a visual query (e.g., an example image or a sketch), and rank all images in the database according to the visual similarities between the query and images. Several CBIR systems have been developed. For example, a sketch based image retrieval system[1] was implemented over some Flickr images. However, it did not perform well, as the visual content in such forms has limited capability to describe the search goal. Also, such techniques cannot scale up, and they may take second-order computation time to rank a thousand images.

Recently, techniques that exploit both tags/texts and visual content have been designed and applied to commercial image search engines. Bing image search and Google image search have a show-similar-image feature, which enables users to select an image and reorder the returned images according to the similarities between the selected image and the remaining images. It looks interesting, but it is still not convenient for users to clearly indicate their intent from an example image.

In this paper, we propose a novel image search system, image search by color map. It enables end-users to intuitively specify how the colors are spatially distributed

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

in the desired images using a color map. The proposed system is motivated by intuitive observations. For text-based image search results with the query “flower”, a color map, with red surrounded by green, can describe an image showing a red flower surrounded by green leaves. For text-based image search results with the query “Brooke Hogan” (an American singer, actress, model, and television personality), a color map, with red on the bottom, can show Brooke Hogan walking on a red carpet.

To use our system, end-users merely scribble a few color strokes in a blank canvas, or drag an image to the blank canvas and mask the color regions of interest. Our system then mines and formulates the search intent from the user's input, and reranks the image search results to help users find desired images. We show two typical examples to give a quick overview of the proposed technique. A user wants to find images in which a sunflower is under the sky. The user may submit a target color map by drawing a few color strokes, shown in the bottom right image of Fig. 1(b), and obtains the reordered results of “sunflower”, as shown in Fig. 1(b). As another interaction mode, the user may drag an image and highlight a few cells of interest. For example, a user aims to find images with a tiger in green grass. The user then drags an image into the canvas from the original image search results shown in Fig. 1(c), and highlights a few cells to indicate the intent as shown in the bottom right image in Fig. 1(d), and finally gets the results shown in Fig. 1(d).

The proposed image search system by color map can be applied to enhance existing text-based image search engines, e.g., tag-based Flickr image search and text-based Google/Microsoft Bing/Yahoo! image search, by providing a way to enable consumers to indicate their visual expectation in the desired images. The main contribution of this paper is summarized as follows. First, an intuitive interface is presented to enable end-users to specify a color map to show how the colors are spatially distributed on desired images in a convenient manner. Using the color map is quite straightforward, only requiring users to pick colors, draw strokes, or drag an image. Second, in contrast to the conventional sketch based image retrieval techniques, the proposed system searches images based on colors rather than shapes, and we, technically, propose a simple but effective scheme to mine the latent search intention from the users input, Finally, we design a two-step reranking procedure, multiple dominant color filtering that filters out a lot of undesired images, and top 1000 image reordering. The experimental results show our search-by-color-map system indeed helps users more conveniently find desired images.

2. RELATED WORK

In recent years, there has been some research undertaken on interactive image search [Cui et al. 2008; Fogarty et al. 2008; Luo et al. 2008]. Bing image search, Google image search and other online systems, e.g., Tiltomo [2], and GazoPa [3], have a show-similar-image feature, which enables users to select an image and reorder the returned images according to the similarities between the selected image and the remaining images. It looks interesting, but it is still not entirely convenient

[2] http://www.tiltomo.com/

[3] http://www.gazopa.com/

as it is not clear which part of the selected image the user is interested in. The CueFlik system [Fogarty et al. 2008] allows end-users to provide examples of images to quickly create their own rules for reranking the images. However, it is not straightforward for users to clearly specify their intent, for the same reason as above, that it is unclear what part of the selected image the user is interested in. In our system, we provide an interface to enable users to specify desired colors directly. A multi-class query based image search system was proposed in [Luo et al. 2008] to utilize information across similar classes to fine-tune the similarity measures. The semantic sketch based techniques are introduced in sketch2photo [Chen et al. 2009] and search by concept map [Xu et al. 2010b; 2010a]. The key differences from ours lie in that they require users to input the textual keywords in a blank canvas while our approach enables users to draw color strokes on the canvas; and that sketch2photo aims to synthesize new images.

In the multimedia community, the closest work is content-based image search (e.g., Google image search's "Similar" and Microsoft Bing image search's "Similar images"), which requires users to submit a visual query (e.g., an example image or a sketch) and reorder the images according to the similarity scores between the query and images. In the following, we compare their differences from the perspectives of user interfaces and algorithms.

2.1 User Interface

Color map vs. color frequency There is a color based feature in Google image search, called dominant color filter. This filter allows users to select a desired color from 12 colors and returns the images in which the selected color is one of the dominant colors. Some other works have investigated techniques using the color frequency (e.g., [Sebe and Lew 2001]) for image search. There are a few online demos about it, e.g., multicolor search lab [4] and Color Fields Experimental Colr Pickr [5]. However, this technique only makes use of the frequency feature of interested colors to rank the images. In contrast, our color map based image search is beyond dominant color search, and particularly provides users with the ability to indicate the spatial distribution intention of different colors.

Color map vs. query by sketch Query-by-sketch in image retrieval [Bimbo and Pala 1997; Colombo and Bimbo 1999; Matusiak et al. 1998; Mehtre et al. 1997] is an interface to allow users to draw strokes to indicate the target. The conventional goals of query-by-sketch include two aspects: Shape indicator and spatial relationship of objects. By comparison, our interface enables users to indicate the spatial distribution of colors under certain image search results. The spatial relationship of objects or colors from query-by-sketch, e.g., the system in [Colombo and Bimbo 1999] and VisualSEEk [Smith and Chang 1996], is very similar to ours. However, our scheme does not require users to describe the query exactly, but provides a powerful scheme to mine the intention map associated with each color from the input. Moreover, our interface can provide a unique feature that enables users to indicate which colors do not appear in the desired images.

Color map vs. boolean composition of regions A mental image search by

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

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

boolean composition of region categories [Fauqueur and Boujemaa 2006] presents an interface to enable users to select the types of regions which should and should not appear in the desired image, where the regions are generated from the unsupervised clustering of salient regions over the image database. A drawback of this technique is that there is a lack of the spatial relation between these selected regions, in contrast with our color map, which has no such issue.

2.2 Algorithm

Color and intention map vs. Color spatial layout The color spatial layout based retrieval techniques [Smith and Chang 1996; Yee et al. 1998] mainly focused on effective representation of the color spatial information of the images in the image database and efficient indexing of those features [Chua et al. 1997; Cieplinski 2001; Deselaers et al. 2008; Jacobs et al. 1995; Lin et al. 2001; Ooi et al. 1998; Stehling et al. 2000].

There are some web application examples, such as Retrievr[6], which is implemented using the fast multiresolution technique [Jacobs et al. 1995], and hermitage web site[7], which is based on IBM's query by image content [Flickner et al. 1995]. These methods may work well when the query is *an example image* (or viewed as an example image). In our case, the input is *a color map* that may be *rough* and *sparse*, which makes the color spatial techniques [Yee et al. 1998] unsuitable for our application. In our scheme, we extract *color maps* for the images in the database, which seems a little similar to the spatial decomposition based representations [Yee et al. 1998; Lin et al. 2001] (e.g., partition-based histogram, multi-precision histogram), but our color map is more compact and in average costs only about 80 bytes to represent an image, which makes it very suitable for web-scale image search. The color signature based representation [Chua et al. 1997], the color layout feature [Cieplinski 2001] and the grid based approach [Stehling et al. 2000] are very similar to ours, but different at least in two aspects. 1) In our color map representation, each cell may have one or two colors, while each cell can only have one color in [Chua et al. 1997]. 2) Its utilization [Chua et al. 1997; Smith and Chang 1996] is very different from ours in that we mine out *an intention map* from the rough and sparse color map to effectively interpret the query color map.

Intention map based similarity vs. image similarity Different from VisualSEEk and [Colombo and Bimbo 1999], we do not directly compare the color map and an image. Instead, we mine the latent intention map to help evaluate the similarity score between the intention map and an image. Different from multiresolution [Jacobs et al. 1995], multi-precision histogram [Lin et al. 2001], and spatial pyramid matching [Lazebnik et al. 2006], we handle a sparse color map, which is easily specified by the users. With our intention map, the similarity measure is partially invariant to translation and scaling because the color map may be rough. Our similarity measure distinguishes rotation, since in our system different rotations can be easily discriminated by the users and easily specified.

[6] `http://labs.systemone.at/retrievr/`

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

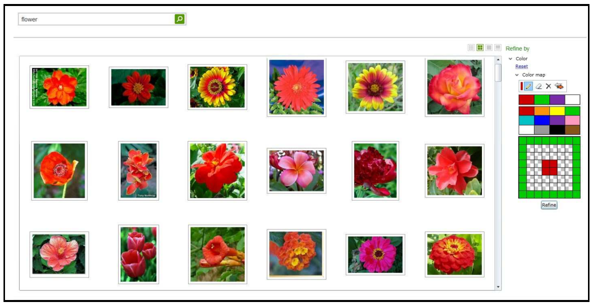


Fig. 2. A snapshot of the interface.

3. SYSTEM OVERVIEW

The process of searching images using the image search-by-color system is described as follows. A user submits a target color map after getting text-based search results, by color stroke scribbling or image dragging. After that, the back-end algorithm will compare the submitted color map and the images, and return the ordered images to users. Then, the users inspect the results and perform an *optional* operation to edit the target color map to get a new ordered result. In the following, we first introduce the user interface components, then present two types of user interactions, and finally highlight potential applications of our system.

3.1 User Interface

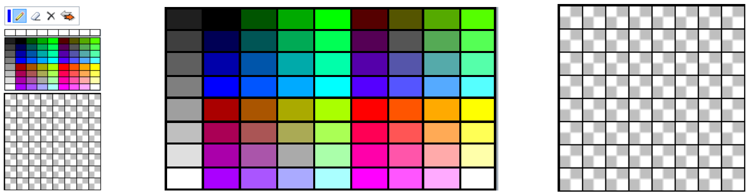
A snapshot of the interface is shown in Fig. 2, which consists of the text query submission box, the result display area, and the interface for color map specification. Here, we describe the interface for color map specification in a little more detail, which is special for the search-by-color-map system. As shown in Fig. 3, it consists of three main modules: Operation bar, color palette, and canvas.

Operation bar In the operation bar, there are five buttons. The leftmost cell is used to show the color in use. The following four buttons are *pen*, *erase*, *clear* and *switch*.

Recent color bar The recent color bar records recently used colors to help users see their color picking history. These colors are displayed in the most recently used order.

Color palette In our system, there are two color palettes: A coarse color palette and a fine color palette. By default, the system only displays the coarse color palette, as shown in Fig. 2, and if the user requires a more accurate color, he may switch to the fine color palette by clicking the *switch* button.

The coarse color palette consists of 12 typical colors, which are selected such that they are representative colors over the whole color space and can be easily differentiated by the users. The fine color palette with more colors consists of the gray table on the left column and the color table on the right. The gray colors are uniformly quantized into 8 levels, which are shown from the top to bottom with the color varying from black to white. For the colors, we uniformly divide a 3-dimensional RGB cube into 4 × 4 × 4 cells, split these cells into four blue-green planes along the red color direction, and finally tile four blue-green planes into a



(a) (b) (c)

Fig. 3. The user interface for specifying the color map.

2 × 2 table to show those quantized colors in the color palette.

Canvas The canvas is adopted for users to draw the target color map. For convenience, we divide the canvas into 8 × 8 cells, with the checkerboard being the background. The canvas is initially covered by the checkerboard to indicate that no color map is input. In the image dragging case, the cells being overlaid by white indicates that the corresponding areas are not of interest, otherwise, it indicates that the colors in the cells are intended to appear in the desired color map.

3.2 User Interaction

A user may submit a target color map, by color stroke scribbling or image dragging. Color stroke scribbling enables users to draw a few color strokes on the blank canvas. Image dragging enables users to drag an image from the search results to the canvas and highlight the regions of interest.

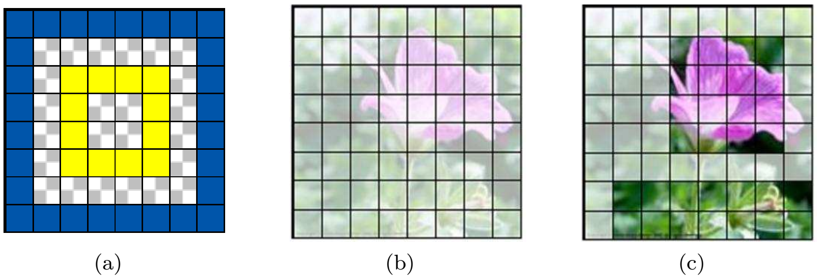
3.2.1 *Stroke Scribbling.* In color stroke scribbling mode, a user can perform the following operations. 1) The user may select the *pen* mode, pick a target color from the color palette, and draw strokes on the canvas by moving the mouse to get a target color map (an example is shown in Fig. 4(a)). During the process of inputting the color map, the user may reselect the *pen* mode to draw another color on blank cells or overlay the original color. 2) The user may select the *erase* mode to revise the color map by erasing the color in some cells. 3) The user may select the *clear* mode to clear the current color map in order to redraw it.

3.2.2 *Image Dragging.* When a user drags an image from the search results to the canvas, the interaction mode is automatically switched to image dragging mode. After the image is dragged into the canvas, it is overlaid with a white mask, as shown in Fig. 4(b), to indicate that no cells are highlighted. Then, in the *pen* mode, the user can highlight some cells by moving the mouse. An example masked image is shown in Fig. 4(c). In the *erase* mode, the user can erase some highlighted cells. In the *clear* mode, the user may clear the highlighted cells and the selected image.

3.3 A Tour of Our System

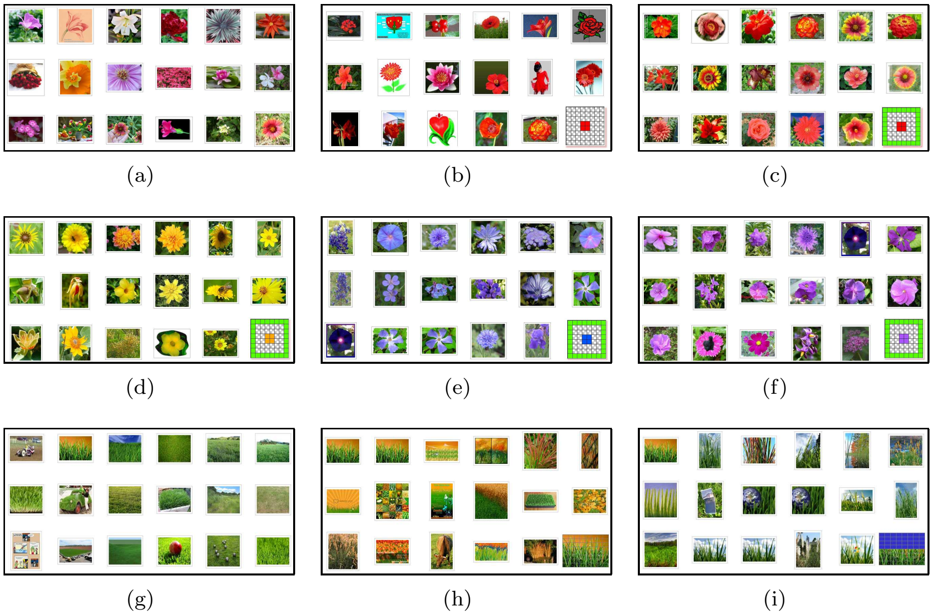
We present three examples to illustrate how to use our system to find desired images.

3.3.1 *Stroke Scribbling Case.* In this tour, we give a series of example interactions to illustrate the stroke scribbling case. In this example, the user is supposed to



(a) (b) (c)

Fig. 4. Illustration of different interactions. (a) shows an example color map from color stroke scribbling. (b) shows the initial status when an image is dragged into the canvas. (c) shows the partially-masked image.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Fig. 5. A tour of the stroke scribbling case is shown in (a) - (f). (a) shows the original search results. (b) shows the search results, with the target color map of red in the center. (c), (d), (e), and (f) show the results with different target color maps, shown in the bottom right. A tour of playing with the system using the two modes is shown in (g) - (i), image dragging and color stroke scribbling . (g) shows the original image search results of "grass". (h) shows the refined results using the target color map in the bottom right from image dragging. (i) shows the results refined using the target color map in the bottom right from stroke scribbling.

find flower images, so the user submits a textual query "flower" and then performs a series of operations to get flower images with different colors.

The user first selected red and scribbled the colors into the middle cells, in order to search *red flowers*. Fig. 5(b) shows the corresponding target color map (in the

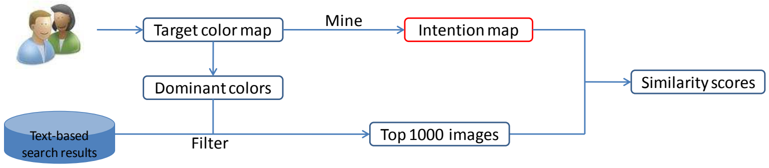


Fig. 6. The flow chart of our approach.

bottom right), and the results. Next, the user had a more specific requirement, hoping that in the desired image the red flower is surrounded by green leaves. Hence the user draws a few green strokes. The corresponding results are shown in Fig. 5(c). After that, the user may want to find images with yellow, blue, and purple flowers, and re-draws the colors to get the desired images. The results are shown in Fig. 5(d), Fig. 5(e), and Fig. 5(f).

3.3.2 *Image Dragging Case.* We present an example to show the dragging image mode to help find desired images. The user aims to find a tiger in green grass. To do so, the user first submitted the text query "tiger" to get the images. She found that there exists one image that satisfies her requirement, but wanted to find more images to be able to select a better one. So, she dragged that image into the canvas, and highlighted a few cells on the tiger and a few cells on the grass, then performed the refinement. This process is illustrated in Figs. 1(c) and 1(d).

3.3.3 *Hybrid Case.* We present an example that uses the two interaction modes together to help users conveniently find desired images. An example is shown in Figs. 5(g), 5(h) and 5(i). Fig. 5(g) shows the original search results with the query "grass". Fig. 5(h) shows the results with the dragged image shown in the bottom-right. The user may like such style of grass in the dragged, but with blue sky, and thus she draw blue strokes on the top. Such a hybrid color map is shown in the bottom-right in Fig. 5(i), and the corresponding search results are shown in Fig. 5(i).

4. APPROACH

This section presents the approach for image representation and similarity evaluation. The flow chart of our approach is shown in Fig. 6. First, we process the target color map to extract the dominant colors and mine the intention map. Then, the dominant colors are used to filter the image search results associated with the text query. Next, we compute the similarity scores between the mined intention map and the top 1000 images that remain after dominant color filtering, and finally reorder those images.

4.1 Dominant Color Filtering

For each image in the image database, we extract the dominant colors and assign them as the tags of the image. In our implementation, we find the dominant colors from the candidate set consisting of 11 colors: BLACK, BLUE, BROWN, GRAY, GREEN, ORANGE, PINK, PURPLE, RED, WHITE, and YELLOW. The selection of the dominant colors is based on human perception and the color representativeness. The colors are selected so that the users can easily differentiate them, and at the same time they are the most representative in natural images. On the other hand, the number of colors should not be so large in order to keep the query process efficient and the storage cost small. For each image, we compute the frequency of each color, and set the colors as dominant colors if their corresponding frequencies are larger than 10%. Specifically, we build a look-up table to store the contribution of each RGB color to the 11 colors, and the frequency is obtained by summing up the contributions of all the pixels to the corresponding dominant colors.

4.2 Image Representation

The proposed color map process consists of two steps: Extract representative colors for each grid and concatenate them into spatial maps with each corresponding to the same quantized color.

4.2.1 *Representative Colors.* We divide the image uniformly into $g \times g$ grids. For each grid (x, y), we aim to find dominant colors, $\mathcal{C}_{xy} = \{\mathbf{c}^i_{xy}\}$, 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 t largest frequencies such that $f_{i-1} < 2f_i$, $\forall i \leqslant t$ 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,

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

In the preprocessing stage, 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 thus there are a total of 192 quantized colors. The finer quantization for the hue channel is reasonable and works well because it is observed that humans are perceptually more sensitive to hue variance. For each grid, in addition, we perform a median filtering to smooth the images on the HSV color space.

4.2.2 *Color Map.* For each color $\mathbf{c} \in \mathcal{C}$, we compute its spatial map as a binary vector $\mathbf{m}$ of dimension $g \times g$ such that

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

where p is one dimensional index corresponding to two dimensional indices (x, y). Intuitively, if color $\mathbf{c}$ appears in grid (x, y), the corresponding entry m_p is set to 1, and 0 otherwise. In summary, the color map of the image is represented by $\mathcal{I} = \{(\mathbf{c}_t, \mathbf{m}_t)\}_{\mathbf{c}_t \in \mathcal{C}}$. This whole algorithm for color map extraction can be easily implemented, and the computational cost is also very low, at about 50 ms per image.

4.3 Target Color Map and Intention Map

The target color map from the user can be computed in a way similar to that described above. We denote the target map as $\mathcal{Q} = \{(\mathbf{c}_i, \mathbf{m}_i)\}$. A straightforward

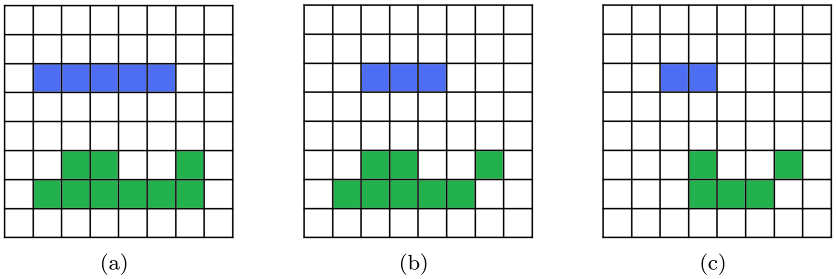


Fig. 7. Consistency intention. (a) the target color map (white represents unpainted), (b) image 1, and (c) image 2.

approach to evaluate the similarity between target color map and image color map is to directly compare the colors for each cell, mathematically written as,

$$\mathrm{sim}(\mathcal{Q},\mathcal{I}) = \sum\nolimits_{\mathbf{c}_q \in \mathcal{C}_q} \sum\nolimits_{\mathbf{c}_i \in \mathcal{C}_I} \mathrm{sim}(\mathbf{c}_q, \mathbf{c}_i)\mathbf{m}_q^T \mathbf{m}_i, \tag{1}$$

where $\mathrm{sim}(\mathbf{c}_q, \mathbf{c}_k)$ is the color similarity between $\mathbf{c}_q$ and $\mathbf{c}_k$, computed in the HSV color space.

However, we notice that the target color map is sparse and rough. The straightforward evaluation approach in Eqn. (1) essentially considers the target color map to be an exact representation, which would require users to describe their intention exactly. This requirement means that the users have to carefully paint their intention, not roughly but exactly, which makes the system inconvenient and means that users would be disinclined to use it. On the other hand, the target color map is sparse, with some cells unpainted. The evaluation in Eqn. (1) leaves the unpainted cells aside, which will lose information. We observe that users actually have some latent intention with regards to the unpainted cells, which is implicitly expressed from the painted cells. Considering the drawbacks of Eqn. (1), we therefore propose a new similarity evaluation approach, in which the key idea is to estimate the intention map from the target color map.

During the design of the similarity evaluation scheme, we communicated with users in order to perform a lot of investigations on the intentions of their painted target color maps. We analyzed our investigation and summarize below three common implicit intentions.

Consistency intention Consistency intention means that the color in the corresponding cells of the image should be similar to the color in the target color map. Consider an example in Fig. 7. Figs. 7(a), 7(b) and 7(c) show the target color map, color image 1, and color image 2. It is implied that color image 1 in Fig. 7(b) is more compatible to the target color map Fig. 7(a) as the overlapped area for the colors of interest (blue and green) between target color map and color image 1 is larger than the one between target color map and color image 2.

Relation intention Let's consider the example in Fig. 8, in which color image 1 in Fig. 8(b) is different from color image 2 in Fig. 8(c) in only two grids. If we only consider consistency intention, it would not be easy to differentiate the two images as this similarity measure may lead to similar (even the same) scores. However,

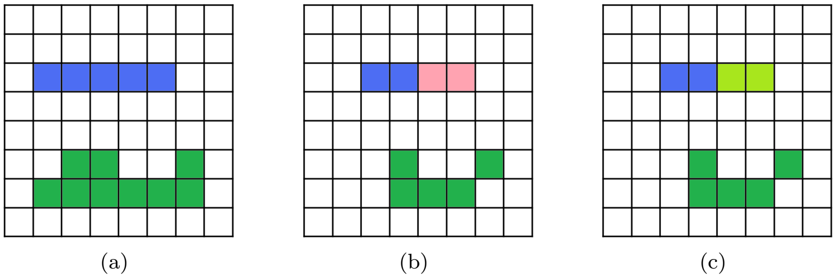


Fig. 8. Relation intention. (a) the target color map (white represents unpainted), (b) image 1, and (c) image 2.

common sense tells us that color image 1 is more similar to target color map because the color aqua in color image 2, which appears in the grids that correspond to blue in target color map, is too similar to green.

Intention propagation Practice shows that users often present a rough target color map. Thus it is insufficient for color map similarity evaluation to only involve the scribbled regions in target color map. Let's see an example in Figs. 9(b) and 9(c). It is clear that their similarities using the above two measures are the same if target color map is given as shown in Fig. 8(a). However, common sense will tell us that color image 2 in Fig. 9(c) is more similar than color image 1 in Fig. 9(b) because the additional blue cells in image 2 are nearer the blue cells in the target color map.

4.3.1 *Intention Map.* To involve the above intentions into the similarity evaluation, we propose a so-called intention map to represent the intention for each color in the target color map. Let $\mathbf{w}_i$ be the intention map of the color $\mathbf{c}_i$, a real-valued vector of dimension 8×8, in which each entry is the intention weight on the corresponding cell. For presentation convenience, we use influence scopes to represent a set of cells for each color. The influence scope $\mathcal{S}_i$ for color $\mathbf{c}_i$ consists of two parts: The cells $\mathcal{S}_i^1$ in which color $\mathbf{c}_i$ appears, and the cells $\mathcal{S}_i^2$ which color $\mathbf{c}_i$ is propagated to. Specifically, the intention map is constructed as follows:

—For each color $\mathbf{c}_i$, we first set the intention weight of its influence scope. For the cells $\mathcal{S}_i^1$, we set the intention weight to be 1, and for the remaining cells $\mathcal{S}_i^2$ in the influence scope, we set the intention weight according to its distance from $\mathcal{S}_i^1$. The rule is that the larger the distance, the smaller the weight. Particularly the weights in the cells that are immediately neighboring to $\mathcal{S}_i^1$ are set to be 0.5. This will result in that the cells of the desired image in the influence scope are intended to have the similar color with color $\mathbf{c}_i$, which corresponds to the consistency intention and intention propagation. We denote the current intention map of color $\mathbf{c}_i$ by $\mathbf{w}'_i$.

—We now add the relation intention to the intention map. For color $\mathbf{c}_i$, it is usually expected that the scopes, dominantly-influenced by other colors, are not intended to have color $\mathbf{c}_i$. Hence, we set the intention map for color $\mathbf{c}_i$ as

$$\mathbf{w}_i = \mathbf{w}'_i - \alpha \sum\nolimits_{j \neq i} \delta_{[\mathrm{sim}(\mathbf{c}_i, \mathbf{c}_j) > \beta]} \mathbf{w}'_j, \tag{2}$$

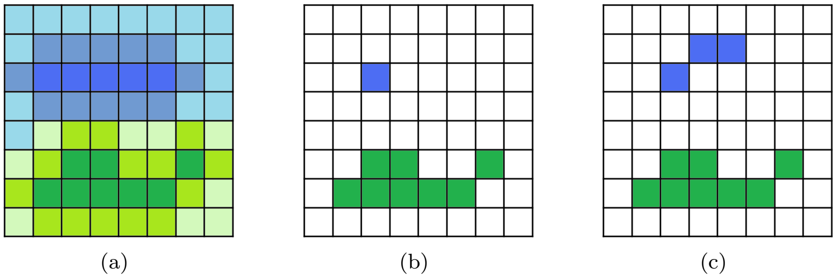


Fig. 9. Intention propagation. (a) the target color map (white represents unpainted), (b) image 1, and (c) image 2.

where α is a trade-off parameter to control the effect of the other colors. The indicator function $\delta_{[\mathrm{sim}(\mathbf{c}_i,\mathbf{c}_j)>\beta]}$ is valued as 1 if $\mathrm{sim}(\mathbf{c}_i,\mathbf{c}_j) > \beta$ and as 0 otherwise. This term is used to discriminate whether the similarity between colors $\mathbf{c}_i$ and $\mathbf{c}_j$ is larger than β. Specifically, when the two colors are too similar (larger than β), the penalization is not imposed, which is helpful in avoiding the over-penalty.

4.3.2 *Intention Map for Single Color.* In the above, we presented the approach to compute the intention map, which is accordant with the user's intention when the number of colors in the target color map is at least two. But we made investigations and found that the intention propagation is not proper for the single color case. Instead, users desire that the color they have painted in one area does not appear in any other area, which forms a negative intention map. On the other hand, the painted cells may not be exact, and the nearby cells around them may have some intention to have the same color. Therefore, in the single color case, the propagated intention map consists of two parts: Positive and negative. The positive part means that the colored area will be associated with positive weights, and intuitively means that the specified color is expected to appear in the area. In our implementation, the positive part is composed of the cells immediately propagated from source cells. By contrast, the negative part means that the specified color is not expected to appear in the corresponding area, and the negative part includes the propagated cells remaining from the positive part.

4.4 Similarity Evaluation

We have represented each image by $\mathcal{I} = \{(\mathbf{c}_t, \mathbf{m}_t)\}_{\mathbf{c}_t \in \mathcal{C}_i}$ with $\mathbf{c}_t$ being the color appearing in the image and $\mathbf{m}_t$ being the associated (binary-valued) color map. The target color map is a little different from the color map, and is represented by colors and intention maps (that can be viewed as weighted versions of the binary-valued color map), $\mathcal{Q} = \{(\mathbf{c}_q, \mathbf{w}_q)\}_{\mathbf{c}_q} \in \mathcal{C}_q$. To evaluate the similarity, we consider each color in the target color map, and compute and sum up the similarities between it and all the colors in the image color map,

$$\mathrm{sim}((\mathbf{c}_q, \mathbf{w}_q), \mathcal{I}) = \sum\nolimits_{\mathbf{c}_k \in \mathcal{C}_i} \mathrm{sim}(\mathbf{c}_q, \mathbf{c}_k)\mathrm{sim}(\mathbf{w}_q, \mathbf{m}_k),$$

where $\mathrm{sim}(\mathbf{w}_q, \mathbf{m}_k) = \mathbf{w}_q^T \mathbf{m}_k$ is the similarity between the intention maps and can be viewed as a weighted count operator, and $\mathrm{sim}(\mathbf{c}_q, \mathbf{c}_k)$ is the color similarity between $\mathbf{c}_q$ and $\mathbf{c}_k$, computed in the HSV color space. Then we aggregate the similarities over all the colors in the target color map,

$$\mathrm{sim}(\mathcal{Q},\mathcal{I}) = \sum\nolimits_{\mathbf{c}_q \in \mathcal{C}_q} \mathrm{sim}((\mathbf{c}_q, \mathbf{w}_q), \mathcal{I}), \tag{3}$$

to get the overall similarity between the target color map $\mathcal{Q}$ and the image color map $\mathcal{I}$. The similarity evaluation scheme is very fast, and only takes about 50 ms per 1000 images.

5. EXPERIMENTS

We conducted an experiment to justify that the proposed system can enhance text-based image search to help users find images more conveniently. There are some available image data sets, e.g., NUS-WIDE [Chua et al. 2009] and ImageNet [Deng et al. 2009]. However, most of them are collected for concept modeling and detection, and are not very suitable for our experiment because they do not indeed reflect the real search results. Therefore, we collected a data set from commercial search engines (e.g., Google, Bing), which is available from [Li et al. 2009]. For each engine, we crawled the top 1000 images from the returned images for a specific text query. There are in total 51 queries, which are selected from the top popular queries such that they cover different types of queries, such as object, scene, and portrait. We collected target color maps from the users, and asked them to help label the ground truth. To differentiate different relevance degrees, we adopted a graded relevance scale, and used four levels from level 0 (the worst match) to level 3 (the best match). We asked 5 users to manually label the relevance level for each target color map, and then selected the most frequent level as the final level for each image. To avoid any bias on the labeling, users were selected so that they have no special knowledge on image search and do not know the proposed technique.

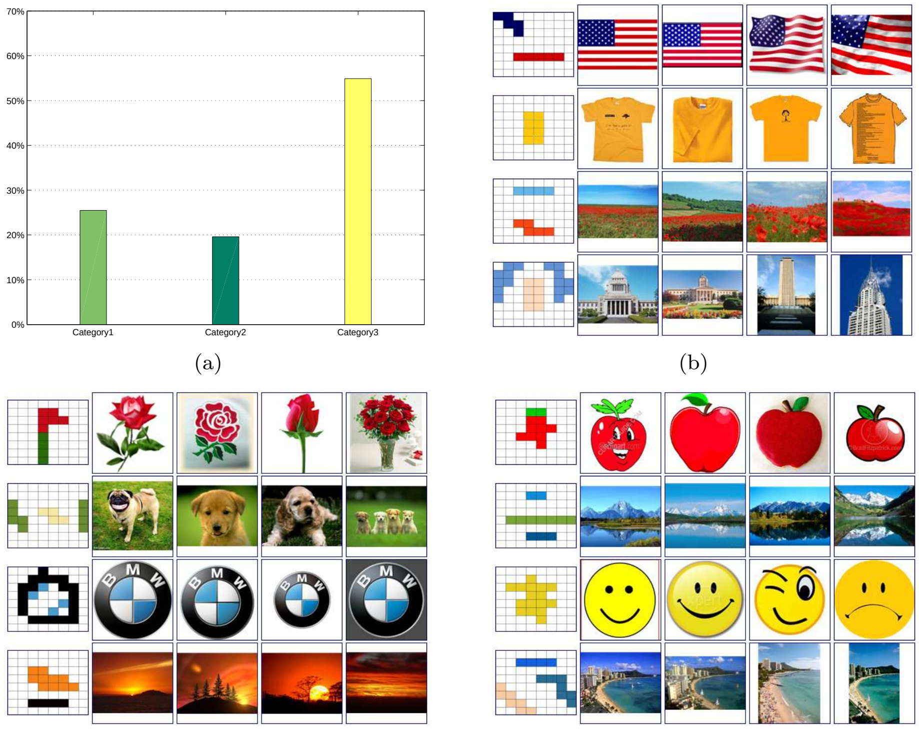
5.1 Evaluation Criterion

We use the normalized discounted cumulative gain (nDCG) measure to evaluate the performance. 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 the 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 the documents of a result list by the ground truth, producing an ideal DCG at position p. Mathematically, nDCG at position p is calculated as

$$\mathrm{nDCG}(p) = \mathrm{DCG}(p) / \mathrm{IDCG}(p), \tag{4}$$

$$\mathrm{DCG}(p) = \sum\nolimits_{i=1}^{p} r_i / \log_2(i+1), \tag{5}$$

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 being the ground truth level of the image at position i, $\mathrm{IDCG}(p)$ is an



(a)

(b)

(c)

(d)

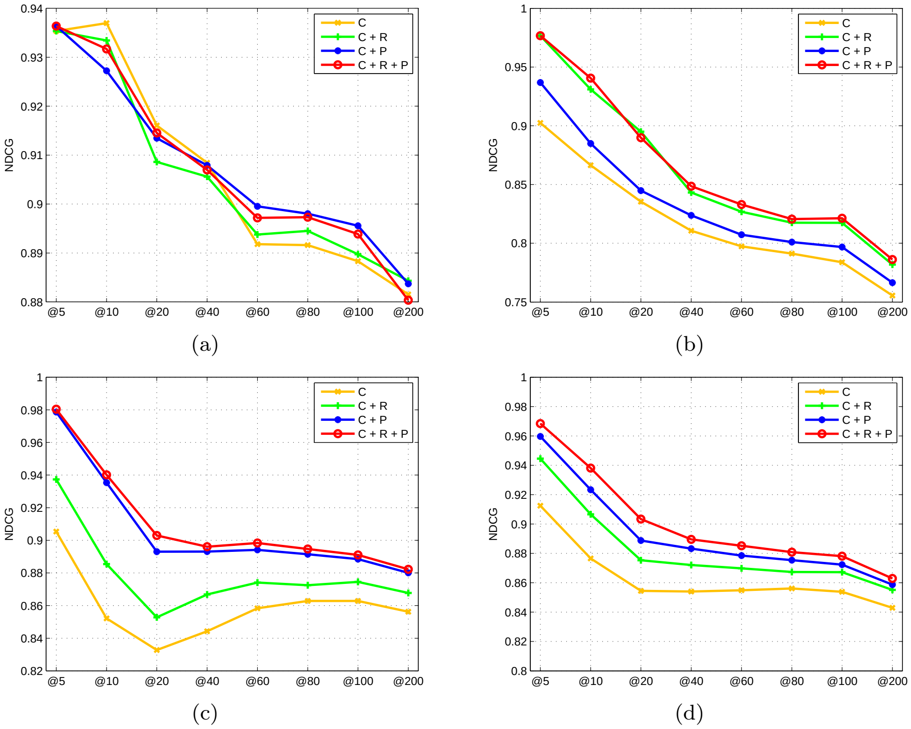
Fig. 10. Target color map categorization. (a) shows the statistics of different categories. (b), (c) and (d) show target maps of three categories and their corresponding images of interest.

ideal DCG at position p. The nDCG values for all queries (all target color maps in our experiments) can be averaged to obtain a measure of the average performance for many queries.

5.2 Performance Illustration

To illustrate the performance clearly, we divide target color maps into three categories, corresponding to consistency intention, relation intention, and propagation intention, respectively, which are shown in Fig. 10, to analyze the effects of the three intentions. It should be noted that the target color maps are collected from the users that do not know anything about our approach, i.e., they scribble the target color maps for those image search results without any special hints that favor the three intentions. Hence, the distribution of the different categories reflect the practice, without being specially designed for our approach. From Fig. 10(a), it can be observed that realistic target color maps indeed cover those three categories. Figs. 10(b), 10(c) and 10(d) visually illustrate the different categories.

The performances of three kinds of intention maps over the collected data set are shown in Fig. 11. "C" stands for "consistency intention", "R" stands for "relation



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

Fig. 11. Illustration of the effects of different intention components. (a), (b) and (c) show the performance over three categories of data sets. (d) shows the performance over the whole data set.

intention", and "P" stands for "intention propagation". "C", "C + R", "C + P", and "C + R + P" represent the four combinations of the three methods. The "consistency intention" map is the basic one, and hence is involved in all four combinations. We report the average nDCG scores over all 51 queries on different rank positions, from position 5 to position 200.

The experimental results are shown in Fig. 11, and we can have the following observations:

(1) Fig. 11(a) shows the performance over the first category of target color maps. It can be observed that in such a case consistency intention (C) is sufficient since the performances of all the methods are almost the same.

(2) Fig. 11(b) shows the performance over the second category. This figure discloses that in some cases relation intention (R) takes effects.

(3) Fig. 11(c) shows the performance over the third category. The fact that "C + P" has a similar performance to "C + R + P" shows that P actually fulfills a role in performance improvement.

(4) Fig. 11(d) shows the performance over the whole data set. "C + R + P", i.e., the proposed approach, gets the best performance.

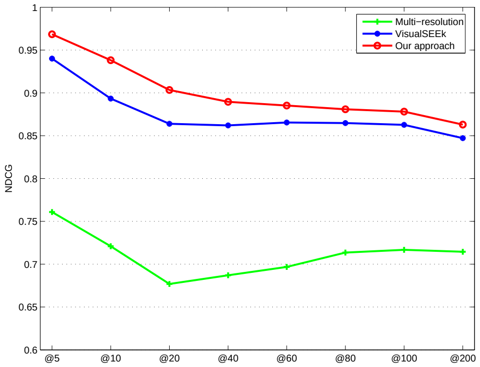


Fig. 12. Quantitative comparison with two representative related techniques.

Table I. Comparison of our approach 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

5.3 Quantitative Comparison

We present quantitative comparison results to justify the superiority of the proposed approach, by comparing our approach with two representative techniques, VisualSEEk [Smith and Chang 1996] and the multi-resolution based method [Jacobs et al. 1995] (demonstrated in Retrievr[8]). The VisualSEEk approach, prior to the query, automatically extracts salient color regions from the images and jointly indexes the color and spatial information for each image. In the query process, the 2D string based similarity computation algorithm is proposed to compare the query and the image in the database. The multi-resolution based approach makes use of multi-resolution (wavelet) decompositions of the query and database images to make the similarity computation more effective and efficient.

The comparison over the whole data set is shown in Fig. 12. It can be observed that our approach is the best and that the multi-resolution based method achieves the worst performance. Some analysis is given as follows. Our approach is superior over VisualSEEk at least in two aspects: 1) The color map representation has a larger ability to represent propagation intention; and 2) our similarity measure is more capable of discriminating relation intention and hence is more compatible with users' intent. Compared with the multi-resolution based method, our approach exploits both consistency intention and relation intention, while the multi-resolution based method only considered consistency intention Tab. I summarizes their differences.

[8] `http://labs.systemone.at/retrievr/`

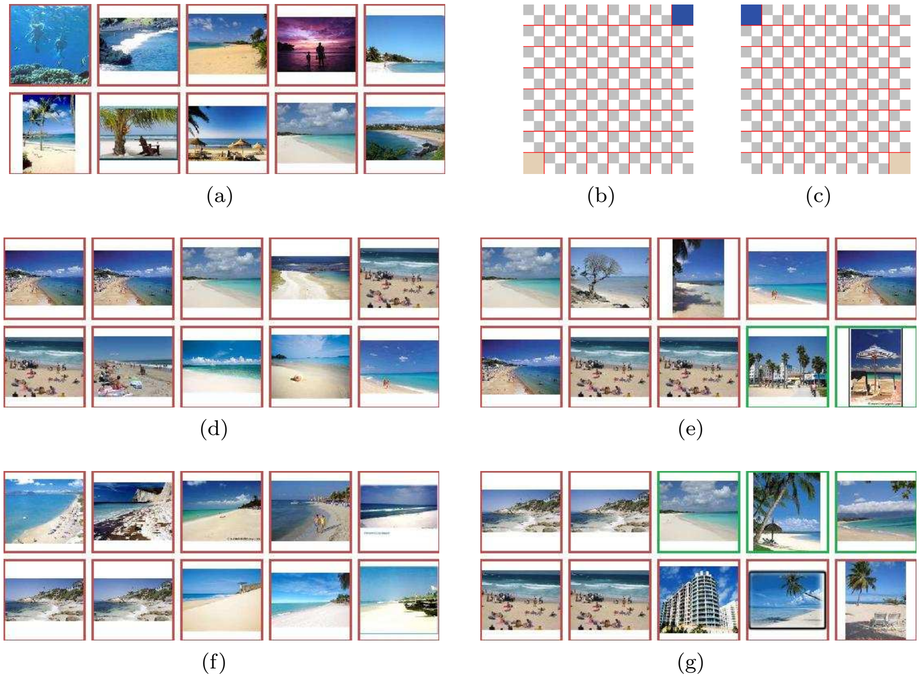


Fig. 13. Visual comparison between our approach and the consistency intention based approach. (a) shows the original top 10 images. (b) and (c) show two target color maps with the same colors and different spatial maps. (d) and (f) show the top 10 images of our approach, corresponding to the two target color maps. (e) and (g) show the top 10 images of the consistency intention based approach. It can easily be observed that the results of our approach meet with the target color map better.

5.4 Visual Comparison

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

In Fig. 13, we present a visual comparison to show that our approach has the ability to distinguish different relation intentions. Fig. 13(a) shows the original top ten images corresponding to the query "beach". Figs. 13(b) and 13(c) show two target color maps from the users. It should be noted that these two target color maps have the same colors but a different spatial relation. For the target color map in Fig. 13(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 consistency intention are shown in Fig. 13(d) and Fig. 13(e), respectively. We can see that the images in Fig. 13(d) meet with the user's intent, i.e., the yellow sand appears in the bottom left and the blue sky appears in the top right, while the ninth and tenth images in Fig. 13(e) are not consistent with the requirements. For the target color map in 13(c), our results shown in Fig. 13(f) are very satisfactory, and obviously are better than the results in Fig. 13(g). In the results based on consistency intention shown in Fig. 13(g), the third, fourth, and fifth images are not consistent with the user's requirements.

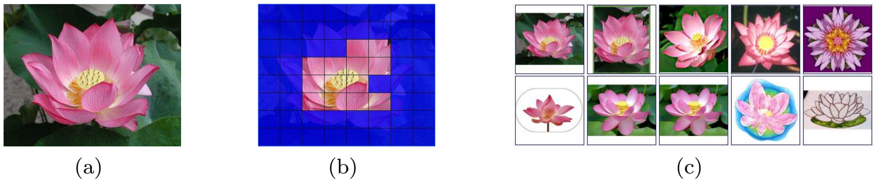


Fig. 14. Visual results with the target color map from an example image. (a) shows the selected image, (b) shows the specified color map, in which the center region without superposing the color blue corresponds to target color map, (c) shows the refined results.

This comparison shows that our approach is superior over the consistency intention based method and can distinguish different spatial relations even with the same colors.

In addition, we present an experiment with the target color map from an example image, as shown in Fig. 14. Figs. 14(a) and 14(b) show the selected example image and the specified color map. Fig. 14(c) shows the original and refined results, respectively.

6. USER STUDY

We conducted user studies to compare our system with two Web image search engines, Google image search and Bing image search. In the following, we use engine A to represent one of the two engines, and engine B to represent the other engine. We justify our system by comparing the performance with Engines A and B, investigating the requirement of search-by-color-map, and inspecting the usability of the user interface.

6.1 Design

Participants We recruit 30 volunteers, students from a university campus and our research lab, to take part in the user study. Their grades vary from freshman to graduate grade 3. Their ages range from 19 to 24. All participants are Web image search engine users.

Data set We collect a large scale of images that come from Microsoft Bing image search for our system. For each image in the database, besides the text-based indexing, we compute two kinds of features offline, dominant colors and color map features, and we organize the images based on the dominant color using the text based technique for efficient dominant color filtering.

Search tasks We compare the performance of our approach with the other two image search engines over a set of tasks. We collect the tasks that are conducted by the volunteers when they tried our system. We select a part of tasks (40 tasks) from those collected tasks. These tasks are given in descriptions that express the search intent, e.g., “the beach on the bottom left and the sky on the top right”. Sample tasks are listed as follows:

(1) White dog on grass;

(2) Island in the sea;

(3) Eiffel tower on the lawn;

(4) Brooke Hogan, dressed in a purple gown, walking on a red carpet;

(5) Ronaldo, on a soccer pitch, wearing the Brazil team strip.

Search procedure The users are asked to find desired images in our system and the other two image search engines. For each task, the users may formulate different queries and use the existing search options supported by the image search engines. For each search engine, we define one unit of operation as performing any interaction so that the search results are refreshed. Ideally, our image search system can combine the search options, existing in Google and Bing image search engines, but in the experiment, users only conduct the color map option and change the text query. When users use Google and Bing image search engines to search images, they can use different text queries, and use the color related search options and similar image search options. Submitting a new text query, or clicking a search option to get a new search result, is viewed as a unit of operation. In our system, either submitting a novel target color map, or submitting a new text query, is one unit of operation.

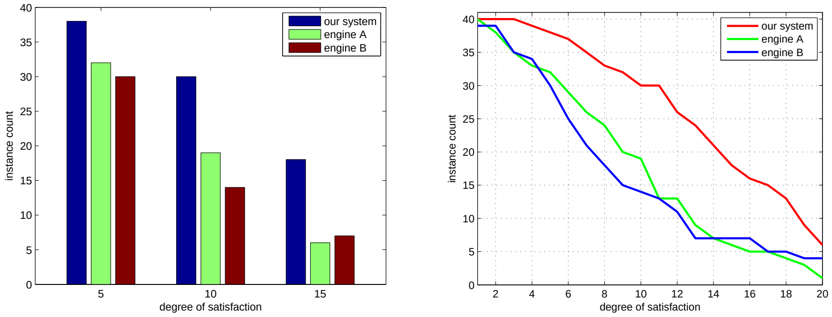
During the interaction process, for each task we ensure that the number of unit operations does not exceed a preset number (5) as we find that users may not try different queries too often. For each unit of operation, we allow the user to evaluate the satisfaction degree in the first 20 images by labeling each image as satisfactory or dissatisfactory. Then, for each task and each participant, we use a vector to record the satisfaction degrees for a series of units of operations.

6.2 Performance Comparison

To compare the performance, we consider two measures. The first is the degree of satisfaction (DoS) in the top 20 images. We do not check all the search results since it is observed that users often check the images in the first several pages. The measure reflects whether users can successfully find their desired images using the search system. The second one is the number of units of operations required to get the satisfied result, which indicates the effort involved.

First, we consider the number of tasks that reach some specific degrees of satisfaction against varying degrees of satisfaction. The comparison is shown in Fig. 15. The horizontal axis represents the maximum degree of satisfaction through five operations. The vertical axis represents the number of tasks, whose degrees of satisfaction exceed some specific degree of satisfaction. Fig. 15(a) considers three typical DoS thresholds, and Fig. 15(b) shows more DoS thresholds. From the performance comparison in Fig. 15(a), our system performs better than those of Engines A and B. We can also observe that 75% of the tasks achieve the 10 satisfactory images in the top 20 images, which shows that our system is very powerful and that it indeed helps users to find desired images.

We also show the performance of the number of the tasks reaching different degrees of satisfaction against varying operations. Fig. 16(a), Fig. 16(b), and Fig. 16(c) show the curves of our system, Engine A, and Engine B with 1-5 operations. The horizontal axis represents the degree of satisfaction, and the vertical axis represents the number of tasks. Note that the first operation is the result of only using the text queries. We can see that our results from operation 2 to operation 5 are better than the results of A and B. Fig. 16(d) shows the performance comparison of the



(a) (b)

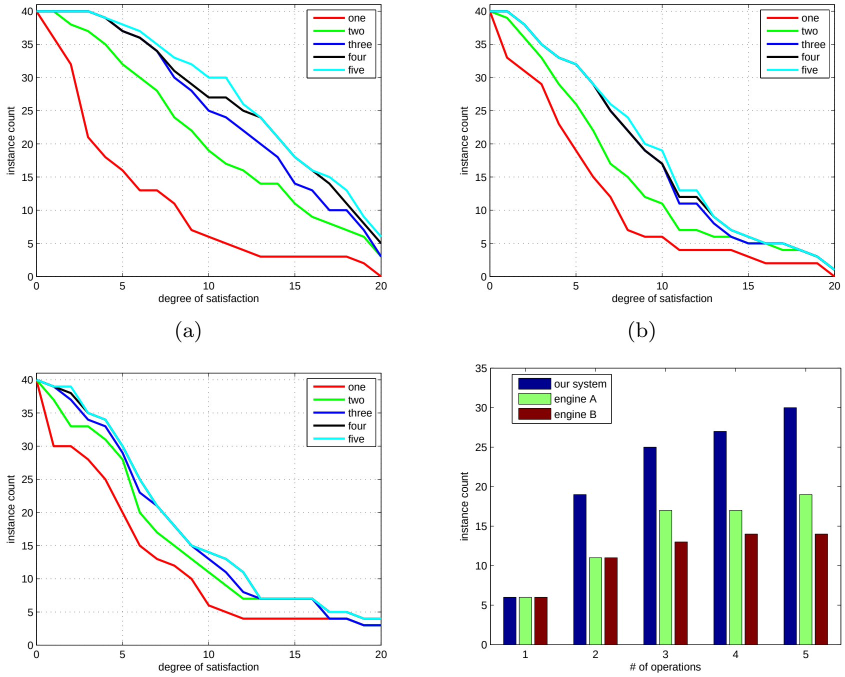
Fig. 15. The performance comparison with the number of tasks against the degree of satisfaction.

three systems considering the specific degree of satisfaction being 50%. The horizontal axis is the number of operations, and the the vertical axis is the number of tasks having 10 satisfactory images among the fixed operations. From this comparison, we can see that our system gets the best performance along all five operations and that our system helps users find desired images in an acceptable amount of operations.

Besides this, we also present visual results to illustrate the performance difference between our system and Google image search. Fig. 17 shows such results on two example search goals, finding images containing "a red apple below green leaves" and finding images containing "a BMW on green grass". Our results come from the textual queries, "apple" and "BMW", with the color maps shown in the left of Figs. 17(c) and 17(d). The results from Google image search are the best from several trials on similar image features and search by color with textual queries, "apple" and "BMW", and finally the results shown in Figs. 17(a) and 17(b) are obtained with the first image as the seed for the similar image feature from Google image search. Furthermore, we also present the results, from the unique feature of our system that is the user can edit the selected image to satisfy his desire. The results are shown in Fig. 18. Figs. 18(a), 18(b) and 18(c) show the results expecting a similar visual appearance with the selected image, the results expecting a blue sky, and the results expecting the color pink at the top.

7. CONCLUSION

In this paper, we propose a new interactive image search system to enhance text-based image search, e.g., Google/Bing image search, by allowing users to indicate their visual requirements. The proposed system enables users to conveniently specify a color map to indicate how the colors are spatially distributed. We propose an effective approach to mine the search intention from the input color map and a two-step scheme to scale our system up to Web-scale image search. We conducted experiments to demonstrate our approach over image search results. We also conducted user studies to compare our system with Google image search and Microsoft Bing image search to demonstrate that our system can indeed help users find desired images. It cannot be denied that the color map has limits in showing other intentions, such as the shape, although it can partially use example regions to indi-



(a)

(b)

(c)

(d)

Fig. 16. The performance comparison with the number of tasks reaching different degrees of satisfaction against the number of operations. (a), (b), and (c) show the results of our system, Engine A, and Engine B, respectively. (d) shows the performance comparison of the number of tasks against the number of operations with the degree of satisfaction being 50%.

cate such intentions. Therefore, in the future, we will investigate more interactive ways to enhance text-based image search.

REFERENCES

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

Chen, T., Cheng, M.-M., Tan, P., Shamir, A., and Hu, S.-M. 2009. Sketch2photo: internet image montage. *ACM Trans. Graph. 28,* 5.

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

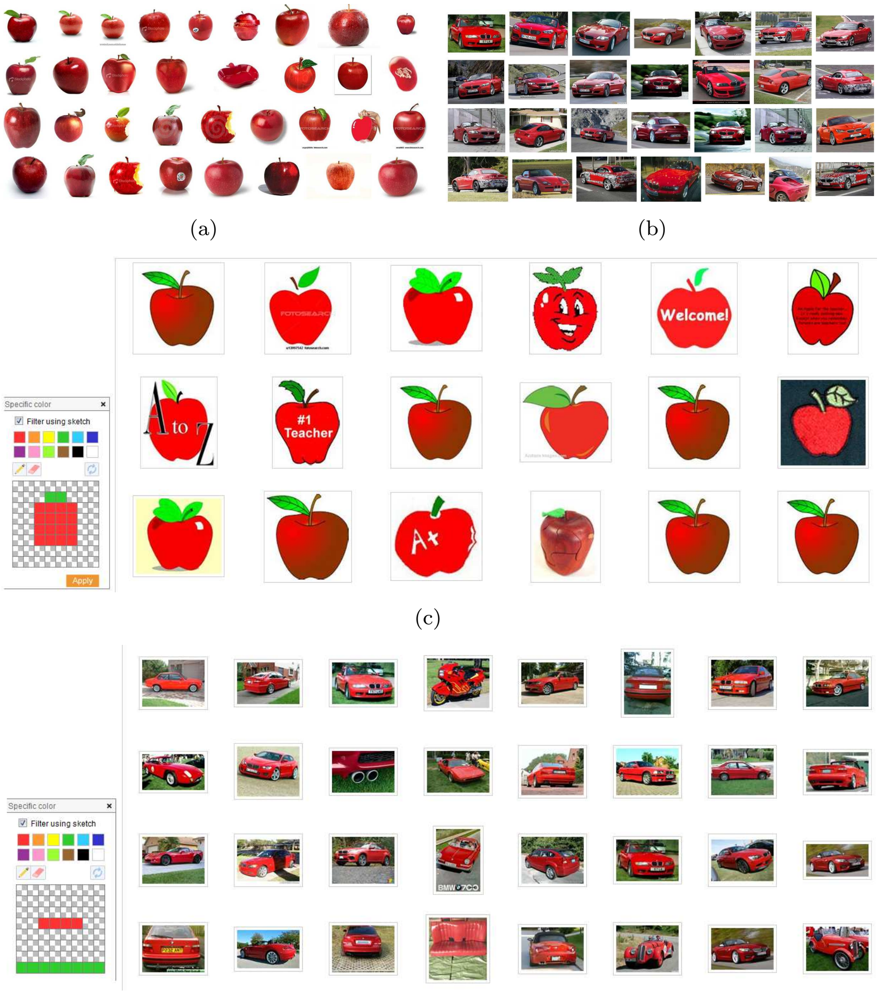
Chua, T.-S., Tang, J., Hong, R., Li, H., Luo, Z., and Zheng, Y. 2009. NUS-WIDE: a real-world web image database from National University of Singapore. In *CIVR*.

Cieplinski, L. 2001. MPEG-7 Color Descriptors and Their Applications. In *CAIP*. 11–20.

Colombo, C. and Bimbo, A. D. 1999. Color-induced image representation and retrieval. *Pattern Recognition 32,* 10, 1685–1695.

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

Datta, R., Joshi, D., Li, J., and Wang, J. Z. 2008. Image retrieval: Ideas, influences, and trends of the new age. *ACM Comput. Surv. 40,* 2.



(d)

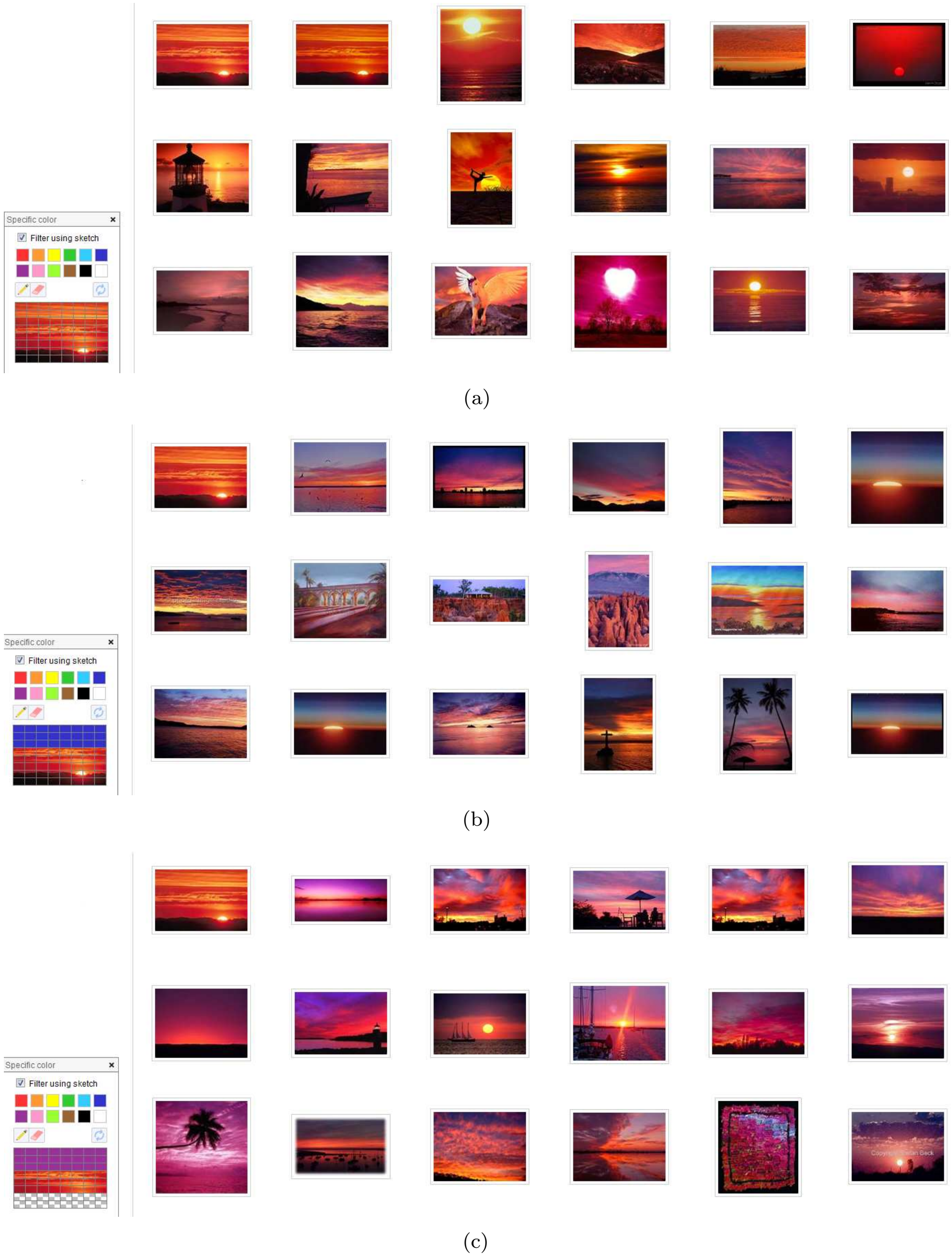
Fig. 17. Visual illustration of the differences between our system and Google image search. (a) and (c) show the results from Google image search and our system, with the search goal of finding images containing a red apple below green leaves, (b) and (d) show the results from Google image search and our system, with the search goal of finding images containing a BMW on green grass.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and 0002, F.-F. L. 2009. ImageNet: A large-scale hierarchical image database. In *CVPR*. 248–255.

Deselaers, T., Keysers, D., and Ney, H. 2008. Features for image retrieval: an experimental comparison. *Inf. Retr. 11,* 2, 77–107.

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

Flickner, M., Sawhney, H. S., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D., and Yanker, P. 1995. Query by Image and Video Content:



(a)

(b)

(c)

Fig. 18. An example to show that our system can enable users to edit an image example to satisfy his/her desire. (b) shows the results expecting a blue sky and (c) shows the results expecting the color pink at the top.

The QBIC System. *IEEE Computer 28,* 9, 23–32.

Fogarty, J., Tan, D. S., Kapoor, A., and Winder, S. A. J. 2008. CueFlik: interactive concept learning in image search. In *CHI*. 29–38.

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

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

Lew, M. S., Sebe, N., Djeraba, C., and Jain, R. 2006. Content-based multimedia information retrieval: State of the art and challenges. *TOMCCAP 2,* 1, 1–19.

Li, H., Wang, M., and Hua, X.-S. 2009. MSRA-MM 2.0: A Large-Scale Web Multimedia Dataset. In *ICDM Workshops*. 164–169.

Lin, S., Özsu, M. T., Oria, V., and Ng, R. T. 2001. An Extendible Hash for Multi-Precision Similarity Querying of Image Databases. In *VLDB*. 221–230.

Luo, Y., Liu, W., Liu, J., and Tang, X. 2008. MQSearch: image search by multi-class query. In *CHI*. 49–52.

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

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

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

Rui, Y., Huang, T. S., and Chang, S.-F. 1999. Image Retrieval: Current Techniques, Promising Directions, and Open Issues. *JVCIR 10,* 1 (March), 39–62.

Sebe, N. and Lew, M. S. 2001. Color-Based Retrieval. *Pattern Recognition Letters 22,* 2, 223–230.

Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A., and Jain, R. 2000. Content-Based Image Retrieval at the End of the Early Years. *IEEE Trans. Pattern Anal. Mach. Intell. 22,* 12, 1349–1380.

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

Stehling, R. O., Nascimento, M. A., and Falcão, A. X. 2000. On “shapes” of colors for content-based image retrieval. In *ACM Multimedia Workshops*. 171–174.

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

Xu, H., Wang, J., Hua, X.-S., and Li, S. 2010a. Image search by concept map. In *SIGIR*. 275–282.

Xu, H., Wang, J., Hua, X.-S., and Li, S. 2010b. Interactive image search by 2D semantic map. In *WWW*. 1321–1324.

Yee, C. Y., Tan, K.-L., Chua, T.-S., and Ooi, B. C. 1998. An Empirical Study of Color-Spatial Retrieval Techniques for Large Image Databases. In *ICMCS*. 218–221.