

Interactive Image Search by 2D Semantic Map^{*}

Hao Xu
University of Science and
Technology of China
Hefei, 230026, P.R.China
xuhao657@ustc.edu

Jingdong Wang
Microsoft Research Asia
49 Zhichun Road
Beijing, 100190, P.R.China
jingdw@microsoft.com

Xian-Sheng Hua
Microsoft Research Asia
49 Zhichun Road
Beijing, 100190, P.R.China
xshua@microsoft.com

Shipeng Li
Microsoft Research Asia
49 Zhichun Road
Beijing, 100190, P.R.China
spli@microsoft.com

ABSTRACT

In this demo, we present a novel interactive image search system, *image search by 2D semantic map*. This system enables users to indicate what semantic concepts are expected to appear and even how these concepts are spatially distributed in the desired images. To this end, we design an intuitive interface for users to formulate a query in the form of 2D semantic map, called concept map, by typing textual queries in a blank canvas. In the ranking process, by interpreting each textual concept as a set of representative visual instances, the concept map query is translated into a visual instance map, which is then used for comparison with the images in the database. Besides, in this demo, we also show an image search system with a simplest semantic map, a 2D color map, where the concepts are limited from the colors.

Categories and Subject Descriptors

H.3.1 [Information Storage and Retrieval]: Content Analysis and Indexing

General Terms

Algorithms, Experimentation, Performance

Keywords

Interactive image search, concept map, color map

1. INTRODUCTION

Image search engines play an important role in enabling people to easily access the desired images. Image search engines often present a query interface to allow users to submit a query in some forms, e.g., textual input, or visual input, to indicate the search goal. To help image search, query formulation is required not only to be convenient and effective to indicate the search goal clearly, but also to be easily interpreted and exploited for the image search engine. The research community on image search has paid a lot of effort on helping users formulating the search requirement [2, 4, 5], for example, some techniques were developed on image search

interface design [1, 7, 8, 9, 10]. In this demo, we present a novel image search system to enable users to conveniently find desired images with the requirements, what semantic concepts are expected to appear and how these concepts are spatially distributed.

1.1 Related work

Existing image search schemes can be roughly divided into three types: text-based search, similar image search, and sketch-based search, which are illustrated in Fig. 1. Most commercial image search engines, e.g., Google image search, Microsoft Bing image search, and Yahoo! image search, belong to the first type, and provide a text box to allow users to type several textual keywords to indicate the search goal. Then the text-based search technique is adopted to match the textual query with the text meta data associated with images. This type of search interface is easy to use for users. But it is limited in that the associated texts may not necessarily be relevant to the image content.

Traditional content based image retrieval techniques [5] usually require users to input a visual query, e.g., reverse image search¹. As an application of the second type, an example image is selected among the search results from textual keywords, and then the results are reranked, and such search functions are released in “show similar images” from Microsoft Bing image search, and “similar image search” from Google image search. For the third type, a painted sketch is drawn to represent the shapes of objects in the desired images, for example, an online similar image search engine, similar image search², presents such a technique. The two schemes then extract visual features for the comparison with images in the database. Such systems are often inconvenient because users have to submit a visual image that may not be available at hand, or paint sketches that may not be easy to exactly indicate the search goal, and the visual images cannot clearly reflect the semantic requirement of the search.

1.2 Our approach

The aforementioned three types of image search schemes all suffer from a limitation that it is incapable of search images with spatial requirements of desired objects. Let’s look at the task of “finding images containing a butterfly on the top-left of a flower”. The results, from the text-based techniques with “butterfly flower” (more complex textual queries even lead to worse results), similar image search, and sketch-based search, are shown in Fig. 1, re-

^{*}This work was performed at Microsoft Research Asia.

Copyright is held by the International World Wide Web Conference Committee (IW3C2). Distribution of these papers is limited to classroom use, and personal use by others.
WWW 2010, April 26–30, 2010, Raleigh, North Carolina, USA.
ACM 978-1-60558-799-8/10/04.

¹<http://www.tineye.com/>

²<http://www.gazopa.com/>

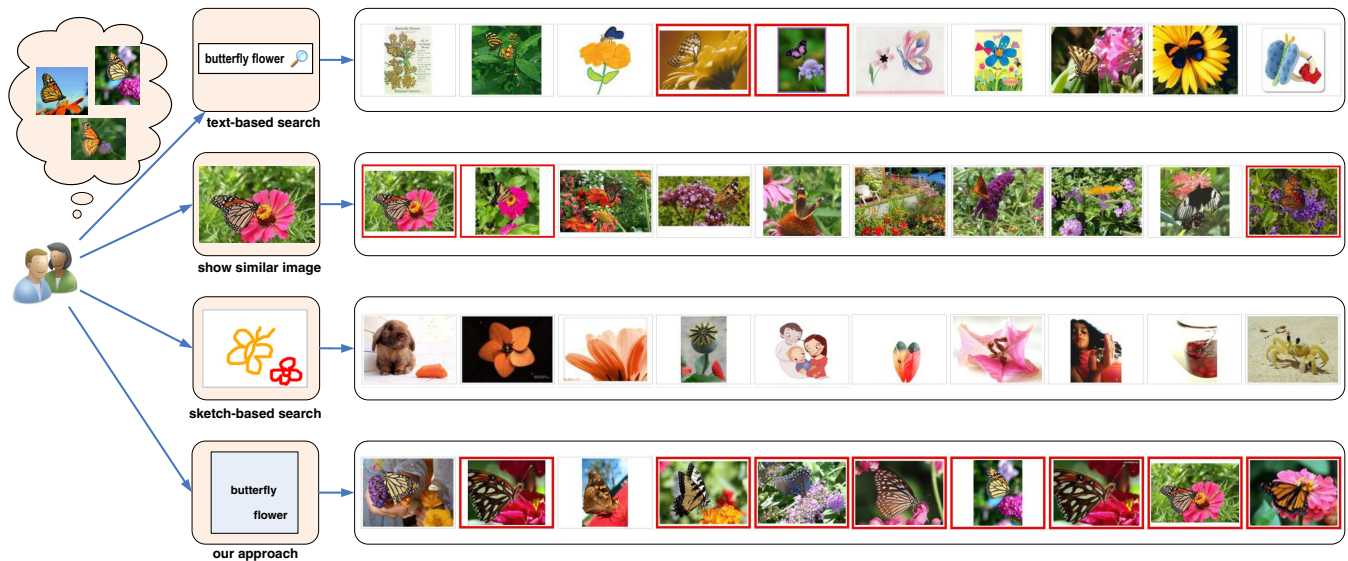


Figure 1: Illustration of four types of image search schemes: text-based image search, “show similar images” of Microsoft Bing image search, sketch-based image search, and our approach. The images satisfying the search requirement are highlighted in red boxes.

spectively. It can be observed that the results are far away from the expectation.

In this demo, we propose a novel image search system, which presents a novel intuitive query inputting interface to enable users to conveniently indicate the search goal by typing the textual queries in a 2D canvas to show where the concepts are expected to appear. Such a query, in this demo, is called concept map. Fig. 1 shows the concept map formulated by a user for accomplishing the aforementioned image search task and the search results, in which most images are exactly what we expected. We also show the search system by a simple 2D semantic map, color map, with limiting the concepts in the colors.

The key contributions include the following aspects.

1. We present a novel image search system to enable users to search images that satisfy the requirements, what semantic concepts are expected to appear and how these concepts are spatially distributed.
2. We propose a novel image search interface to enable users to intuitively input a concept map by typing textual concepts in a blank canvas to formulate the search goal.
3. Technically, we introduce an effective technique to translate the concept map into visual instance map, and a robust scheme to evaluate the relevances of images. This technique is very adoptable for large scale of images and concepts, while conventionally concept detection techniques usually suffer from the large scale problems.

2. SYSTEM

The user interface of the interactive image search system by 2D semantic map is shown in Fig. 2. It consists three modules: query formulation canvas, where users can specify the query, search result panel, where the search results are presented, and advanced function, which enables users to perform advanced operations. In the following, we first describe the basic steps to play our system

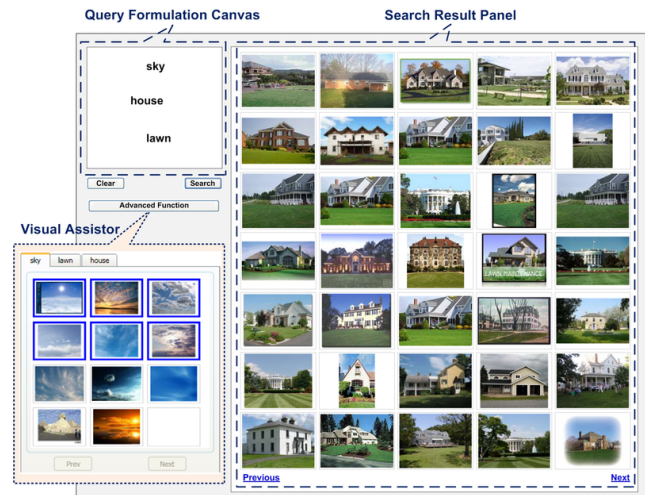


Figure 2: User interface of the proposed system.

to search images. Then we present two optional operations, spatial influence scope specification, and visual assistant to specify what concepts visually look like. Finally, we introduce a simple 2D semantic map, color map, in which the concepts are limited from the colors.

Basic operations

To perform image search, the basic steps are as follows. First, a user formulates a query in the form of 2D concept map, by typing a few keywords (concepts) in *query formulation canvas* to show where the concepts are expected to appear. An example is shown in Fig. 2, in which the query consists of three concepts, “sky”, “house” and “lawn”, which are expected to appear from top to bottom in desired images. Then, after submitting the query, the system returns a list of images according to their relevances with the submitted concept map query, and presents them in *search result panel*.

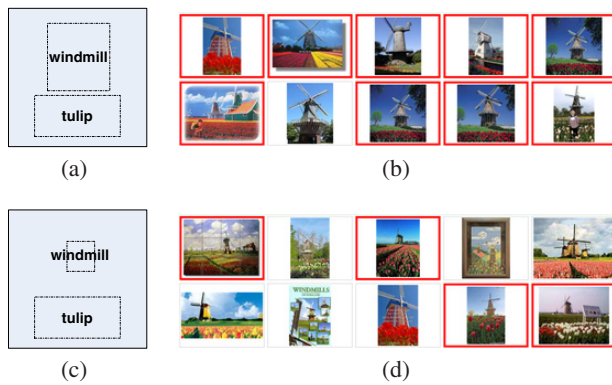


Figure 3: Illustration of influence scope specification. (a) and (c) show two concept maps with the same concepts but different influence scopes for the concept “windmill”, indicating to find a big windmill and a small windmill respectively. (b) and (d) are search results obtained by (a) and (c).

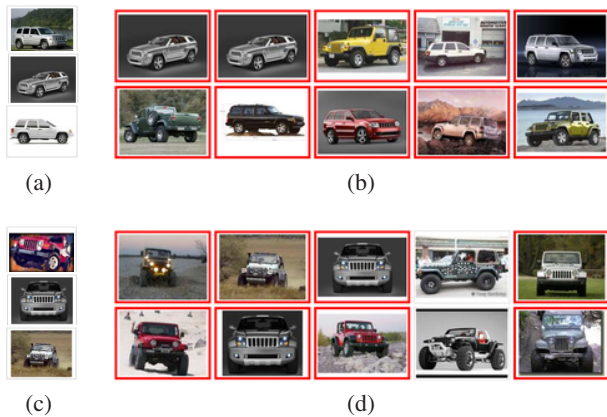


Figure 4: Illustration of visual assistant. (a) and (c) show two different sets of visual examples selected for the concept “jeep”, indicating to find a jeep in the side view and a jeep in the front view, respectively. (b) and (d) are search results corresponding to (a) and (c), using the same concept map query.

Influence scope specification

In the query formulation process, the user may perform an optional operation, influence scope specification (ISS), to indicate the spatial influence scope of each concept. ISS associates each keyword with a rectangle and enables the user to explicitly control the influence scope of a keyword by stretching the rectangle. An example is shown in Fig. 3. Here, the larger influence scope of “windmill” in Fig. 3(a) means that the object, “windmill”, covers a larger region, while the smaller influence scope of “windmill” in Fig. 3(c) means that the object, “windmill”, covers a smaller region. From the results, it can be seen that our system can help reach this goal.

Visual assistance

We find that in some cases it may be insufficient to merely use a textual keyword to represent the search goal, and hence provide an advanced scheme, called *visual assistant*, to enable users to select a few visual examples to assist to visually specify or narrow down what the concept visually looks like. Users may click *advanced function* button, to popup the *visual assistant* window. For each

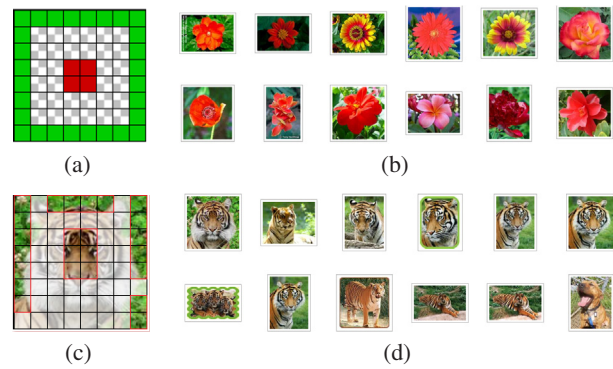


Figure 5: Illustration of image search by color map. (a) and (c) show two types of color maps, and (b) and (d) show the search results, respectively.

concept, a set of visual examples are presented for user selection, as shown in Fig. 2. Fig. 4 shows an illustrative example. The user can select different views of jeeps, shown in Fig. 4(a) and Fig. 4(c), and gets the desired images shown in Fig. 4(b) and Fig. 4(d).

Color map

We introduce a simple version of 2D semantic map, in which each semantic concept is reduced to colors. To be intuitive, we provide a color table to enable users to select a color of interest rather than specifying a color name. In addition, the user can move the mouse and draw color strokes to specify the influence scope of each color. An example of color map is shown in Fig. 5(a). To make the search results semantically relevant, the user is also needed to submit a textual query as pre-search. The results of color map based image search with the textual query “flower” is shown in Fig. 5(b). In this example, the color map may mean to find images with the semantic requirement that the red flower is in the middle and surrounded by green leaves. As a special color map, the user may select an image and highlight a few color regions of interest, and then submit it as the color map. Such an example is shown in Fig. 5(c), which may mean to find images with the semantic requirement, the tiger in the green grass. The search results with pre-searched by textual query “tiger” are shown in Fig. 5(d).

3. IMPLEMENTATION

In this section, we briefly describe the technique to realize the proposed system. The basic idea of the technique is to transform the textual concept to a few representative visual instances, then to estimate the occurrence map of each concept over an image, and finally to rank the images according to the consistency between the occurrence maps of desired concepts and the spatial map that is specified in the concept map by the users. This technique is inspired by the successful instance based classifier, and based on the reliable image search results for single textual concept of existing commercial image search engines. This technique also avoids training parametric classifiers for large scale of concepts and large scale of images. The flowchart of the technique is shown in Fig. 6. It should be noted that for the color map based image search, the visual feature is the color itself.

3.1 Concept map translation

Visual instance transformation

We collect a set of images by exploiting text-based image search

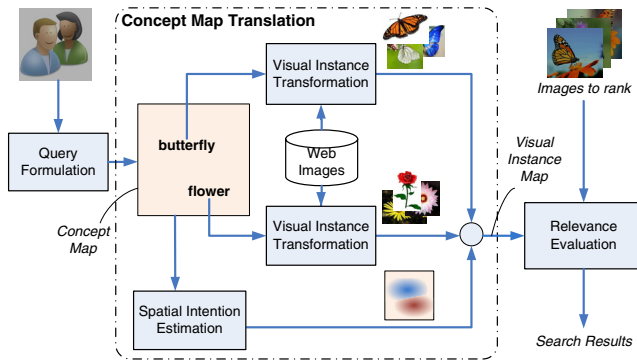


Figure 6: The flowchart of concept map translation.

using the textual concept, and then adopt the affinity propagation (AP) algorithm [3] to find the representative visual instances. To reduce the computation cost, we run AP only over the top images returned by text-based image search which is based on the observation that the main visual appearances of a concept are almost covered in the top images. In our implementation, the visual similarities are evaluated on the regions containing salient objects, which have large probability to correspond to the concept. This helps reducing the affects from the regions irrelevant to the concepts. The salient object is detected offline using the learning based technique [6]. We sort the obtained exemplars in a descending order of the sizes of their associated image groups, and take the detected salient objects of the top exemplars as the visual instances of the concept. The optional operation of visual assistance essentially performs the visual instance specification.

Spatial intention estimation

A concept map includes the rough position and the influence scope. With these information, we estimate the spatial intention that is represented by a spatial distribution. The estimation is based on the observation that the concept has larger probability to occur near the specified position, and less probability to appear far away.

The estimation procedure is as follows. Given the influence scope of concept k , which is represented as a rectangle r_k , the desired spatial distribution of concept k is estimated by a 2D Gaussian distribution parameterized by the rectangle r_k . The center of the 2D Gaussian distribution is the center of the rectangle, and the variance is determined by the size of the rectangle. Later, the continuous distribution is discretized into a 9×9 matrix for the relevance evaluation.

3.2 Relevance evaluation

The evaluation procedure consists of two basic steps: occurrence map estimation for each concept and consistency evaluation between occurrence maps and spatial intention.

We follow the state-of-the-art image representation technique, and extract a bag-of-words (BoW) feature. Rather than building a global BoW model, we build a block-based BoW model, which will be helpful to obtain the spatial occurrence layout for each concept. In our implementation, we divide each image into 9×9 blocks. We adopt two kinds of visual features: the color feature and the SIFT feature, which are used for different concepts. For color feature, each pixel is converted into the HSV (hue, saturation and value) color space and quantized into 192 levels. For SIFT (scale invariant feature transform) feature, the extracted SIFT features are quantized into visual words with a visual vocabulary of size 6K.

We compute an occurrence map O_v^k of dimension 9×9 for a

visual instance I_v of the concept k by computing the similarity between I_v and each cell of the image. We compute the similarity score by comparing their BoW vectors, using a similar technique to the histogram intersection measure. By combining all the occurrence maps of the visual instances of concept k , we can get an overall occurrence map O^k of the concept k .

The consistency evaluation between occurrence maps and spatial intention is based on the distribution consistency criterion between them and the balance criterion. For the first criterion, we compute the consistency between the occurrence map and the spatial intention for each concept, by viewing them as probabilistic distributions and comparing them using the divergence evaluation technique for probabilistic distribution, and in our implementation, we adopt the histogram intersection. The second balance criterion guides how to fuse the consistency values for several concepts, and avoids over-weighting any concept.

3.3 Speedup

It would be very computationally expensive if the image database is very large since we have to scan each image in the database for the evaluation. To reduce the cost, we introduce a trick and first obtain a set of images by performing the text-based image search using the combinations of the desired concepts and selecting top 500 images for each search result. The trick is reasonable since the text-based search has ability to obtain images containing the desired concepts although limited.

4. CONCLUSION

In this demo, we present an interactive image search system, to enable users to find images with the requirements, what semantic concepts are expected to appear and how these concepts are spatially distributed in the desired images. The present system is successful to search some concrete concepts, e.g., jeep, windmill, and colors. In the future, we will investigate other general concepts.

5. REFERENCES

- [1] J. Cui, F. Wen, and X. Tang. Intentsearch: interactive on-line image search re-ranking. In *ACM Multimedia*, pages 997–998, 2008.
- [2] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Image retrieval: Ideas, influences, and trends of the new age. *ACM Comput. Surv.*, 40(2), 2008.
- [3] B. J. Frey and D. Dueck. Clustering by passing messages between data points. *Science*, 315(5814):972, 2007.
- [4] R. Jain. Multimedia information retrieval: watershed events. In *Multimedia Information Retrieval*, pages 229–236, 2008.
- [5] M. S. Lew, N. Sebe, C. Djeraba, and R. Jain. Content-based multimedia information retrieval: State of the art and challenges. *TOMCCAP*, 2(1):1–19, 2006.
- [6] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H.-Y. Shum. Learning to detect a salient object. *IEEE Trans. Pattern Anal. Mach. Intell.*, to appear.
- [7] Y. Luo, W. Liu, J. Liu, and X. Tang. Mqsearch: image search by multi-class query. In *CHI*, pages 49–52, 2008.
- [8] G. P. Nguyen and M. Worring. Optimization of interactive visual-similarity-based search. *TOMCCAP*, 4(1), 2008.
- [9] R. Yan, A. Natsev, and M. Campbell. Multi-query interactive image and video retrieval: theory and practice. In *CIVR*, pages 475–484, 2008.
- [10] Z.-J. Zha, L. Yang, T. Mei, M. Wang, and Z. Wang. Visual query suggestion. In *ACM Multimedia*, pages 15–24, 2009.