MindFinder: Interactive Sketch-based Image Search on Millions of Images*

Yang Cao
Brain-like Computing Lab
Shanghai Jiao Tong University
Shanghai, China
cybersun.sjtu@gmail.com

Hai Wang
University of Science and Technology of China
Hefei, China
wanghai369@gmail.com

Changhu Wang
Microsoft Research Asia
No.49, Zhichun Road
Beijing, China
chw@microsoft.com

Zhiwei Li
Microsoft Research Asia
No.49, Zhichun Road
Beijing, China
zli@microsoft.com

Liqing Zhang
Brain-like Computing Lab
Shanghai Jiao Tong University
Shanghai, China
zhang-lq@cs.sjtu.edu.cn

Lei Zhang
Microsoft Research Asia
No.49, Zhichun Road
Beijing, China
leizhang@microsoft.com

ABSTRACT

In this paper, we showcase the MindFinder system, which is an interactive sketch-based image search engine. Different from existing work, most of which is limited to a small scale database or only enables single modality input, MindFinder is a sketch-based multimodal search engine for million-level database. It enables users to sketch major curves of the target image in their mind, and also supports tagging and coloring operations to better express their search intentions. Owning to a friendly interface, our system supports multiple actions, which help users to flexibly design their queries. After each operation, top returned images are updated in real time, based on which users could interactively refine their initial thoughts until ideal images are returned. The novelty of the MindFinder system includes the following two aspects: 1) A multimodal searching scheme is proposed to retrieve images which meet users' requirements not only in structure, but also in semantic meaning and color tone. 2) An indexing framework is designed to make MindFinder scalable in terms of database size, memory cost, and response time. By scaling up the database to more than two million images, MindFinder not only helps users to easily present whatever they are imagining, but also has the potential to retrieve the most desired images in their mind.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Query formulation, Search process; H.5.2 [**User Interfaces**]: User-centered design

*This work was performed at Microsoft Research Asia.

MM'10, October 25–29, 2010, Firenze, Italy.

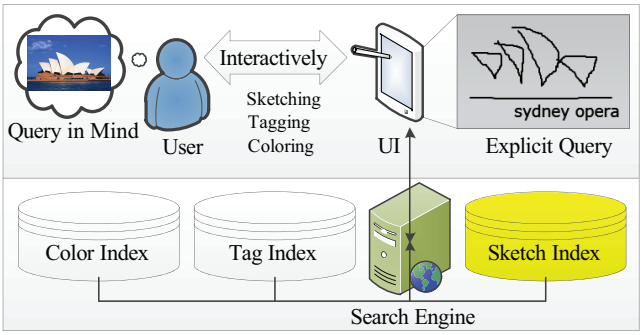

Figure 1: The framework of MindFinder. Using our system, an implicit query in a user's mind is turned into an explicit query with simple features, and after each interaction, resulting images from a database of 2.1 million web images are returned in real time.

General Terms

Design, Experimentation

Keywords

MindFinder, sketching, tagging, interactive search

1. INTRODUCTION

Owing to the prevalence of the Internet and imaging devices, billions of digital pictures are now freely accessible online. How to effectively and efficiently find the best images matching users' search intentions has become an active topic in both academic and commercial circles.

Currently, text-based image search engines, e.g. Google, Bing, and content-based search engines, e.g. tineye[1] are two major types of image search. However, when a user's search intention is complex, neither of these techniques could meet his/her needs. For text-based search, many requirements such as the object shape cannot be easily formulated as a text query. As shown in Fig. 2, although all keyword queries are elaborately chosen by a typical user, top results are still far from what he/she wants, which demonstrates that sometimes text query is insufficient to fully deliver a user's search

[1]http://www.tineye.com/

Figure 2: Comparison between MindFinder and traditional text-based image search engine (Flickr search). Gray images are explicit queries generated by a user, which also indicate the best matching image in his/her mind. Next to it are top results from our system. For some complex tasks, the user also adds tags (indicated by words in each sketch image) and colors (indicated by small colored circles in the top-right corner). We also provide top results from the Flickr search engine. Each text box indicates the keywords tried out by the user to describe his/her search intention to a great extent, and next to it are the most relevant images.

intention. For content-based search, a typical query is a similar image, however, whose absence is usually the reason we cannot start a search.

To overcome existing problems in current search engines, a natural solution is to enable users to flexibly express what they want by drawing strokes (contours of objects), providing tags, and specifying main colors. As illustrated in Fig. 1 & 2, a query to our system, MindFinder, may contain a sketch, a few tags and several dominant colors, and the top returns are natural images with similar contours, semantics, and colors. We can see that, the sketch-based search is more accurate and convenient than the traditional search when a user's search intent is specific and complex. With the prevalence of devices with touchable screens, almost all necessary prerequisites to support this UI are ready.

Since tremendous works have been proposed on searching images via tags and colors, we mainly focus on the sketch-based search in this paper. Actually, among various query modalities, sketch is probably the most challenging one. Although extensive studies about sketch-based search started from 1990 (see [1] for a survey), little progress has been made in the past decade due to the following three barriers. 1) There is an unavoidable gap between the binary map sketched by the user and full color natural images in the database. It is not easy to extract representative curves for

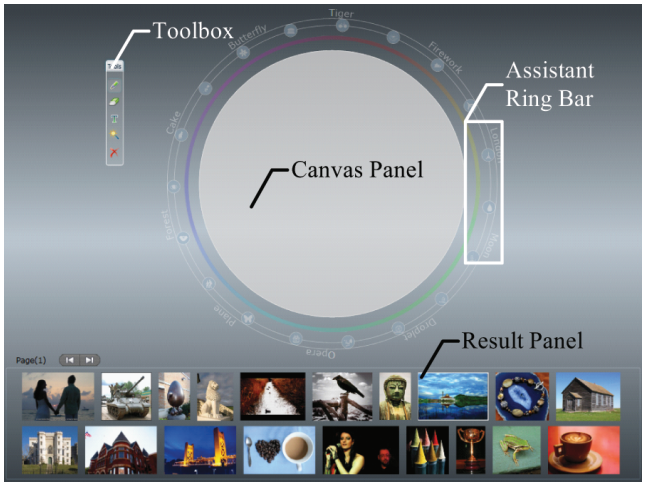

Figure 3: The user interface of MindFinder.

a natural image to match with a query sketch. 2) The second barrier is the difficulty in meeting the requirement for both efficiency and accuracy when matching a query sketch with curves in a natural image. 3) Most existing sketch-based search engines [3] have no indexing mechanism, which means they will face scalability problems when scaling up to millions of images. Actually, these three issues are tightly coupled with each other. On the one hand, the representation and matching schemes determine whether the system could be adapted to an efficient indexing structure. The indexing mechanism is a key point to make the system practical in huge databases, which is crucial to ensure that the system can always find adequate images to match users' queries. On the other hand, a large scale of database also raises the possibility of returning false positives in top returns, which poses a more rigid requirement on precision of the representation and matching scheme.

In this work, we systematically solve the following three problems, i.e. image representation, matching and indexing. To bridge the gap between the sketch query and the full color image, we represent a natural image by its salient curves, which has potential to be closer to the sketch queries from users. Besides, a raw curve-based algorithm is used to efficiently and precisely calculate the similarity between the salient curve representation of natural images and a user's sketch query. We also design an indexing strategy to speed up the matching process and make the system scalable to millions of images. Under this framework, we have built a real-time sketch-based multimodal search system using more than 2 million web images with less than 8GB memory cost for the sketch index on a common machine. Few state-of-the-art sketch-based search engines could handle databases at million level. The web site Retrievr [2] applied Jacobs *et al* 's method [4] on a database with 1M Flickr images, and Eitz *et al.* [2] announced a system to retrieve more than 1.5M images. Both of them try to extract effective and isometric vector representation for each image to facilitate the linear scan of the whole database, which discard local detail of structure information and may face scalability problem. To the best of our knowledge, MindFinder is the first large-scale indexing framework for sketch-based image search.

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

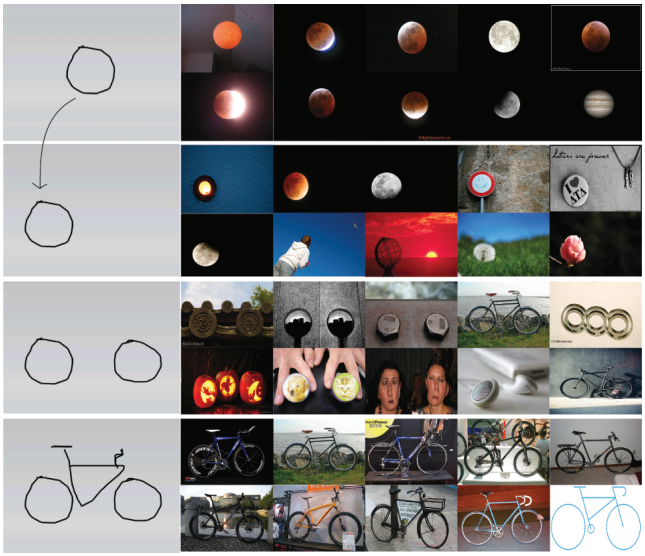

Figure 4: Illustration of interactive sketch-based image search. For each search, top 10 resulting images from the database of 2.1M images were shown. To find bicycle images, the user first drew a circle and then adjusted its location. Then, more parts were progressively sketched. After each modification of the query, the system accordingly updated the top returns in real time. We can see that, every newly added stroke leads to more impressive results, and the final results are all relevant images.

2. SYSTEM OVERVIEW

In this section, we introduce the MindFinder system. A user-friendly interface (as shown in Fig. 3) is designed to help users to flexibly express their search intentions. Our system has two main panels: *canvas panel* and *result panel*. The canvas panel, i.e. a circular interface in the center, is the major component where users present their ideas by directly sketching or adding other supporting information. Surrounding the canvas panel are assistant ring bars, which are designed for users to quickly specify color, or add commonly used tags and sketch patterns. The result panel is a two-line display bar under the canvas panel, which shows top returned images and will be immediately refreshed when users alter the query. Our system also provides a floating toolbox to facilitate users to switch querying mode. Using this UI, users could interactively refine their initial thoughts with real-time response. Millions of images are indexed by separate indices of sketch, tag and color at back-end to support the interactions between users and the system. Technique details will be introduced in Section 3.

The major querying operation of MindFinder is sketching, which describes the main curve and structure of the image in user's mind. In addition, for some abstract concepts or complex objects that cannot be clearly drawn by normal users, our system allows users to add tags to constraint the semantic subject of return images. Since color is also an indispensable feature within an image, our system supports to specify at most 3 dominant color cues during a search. In the rest part of this section, we introduce the main functions of MindFinder step-by-step.

Figure 5: Illustration of the multimodal search to find "the Sydney Opera House in side-view" by MindFinder. Top row: the top three search results queried by only a keyword "sydney opera". Middle row: a mixed query combined with tag and sketch (the grey image) and its top results. Bottom row: the top one results after specifying different dominant colors, which are indicated by circles in the top-right corner of the four images.

1. Sketch Querying

   Draw strokes within the canvas panel. When finished, clicking the background will start the search. In addition, users could easily move or erase some strokes. In this case, search results will be automatically updated according to users' modification. An illustration of searching bicycles by interactively sketching is provided in Fig. 4. We can see that, the sketch-based search in our system is precise and structure-sensitive, which guarantees that the curves in natural images are highly matched with the strokes drawn by users.

2. Sketch + Tag Querying

   To add some text queries, switch to the text mode, then type keywords in the canvas panel. When finished, a new search will be automatically triggered. Users are also allowed to drag and drop some commonly used tags onto the panel for saving time. Fig. 5 shows the search results of tag+sketch querying. A user wanted to collect some photos about the famous side-view of the Sydney Opera House. By only using the keyword-based search, top images were diverse (top row) and far from the user's intention. After adding a sketch query to confine the main structures, all top results (middle row) met the user's requirement.

3. Sketch + Tag + Color Querying

   To add some color queries, click the color bar among assistant ring bars, and then specify dominant colors and corresponding mutual weights. Once the state of the color bar is changed, a new search will be automatically triggered. The bottom row in Fig. 5 shows the results after adding a color query. We can see that, by specifying a certain dominant color, the user could easily find pictures of the Sydney Opera House in any lighting color he/she wants. Besides, the composition of the returned images still meet the user's search intention (side-view of the Sydney Opera House).

It should be noted that, we provide above examples in order to emphasize sketch-related performance of our system. Actually, our system supports any combination of

Figure 6: Example queries and corresponding top search results. Grey images in the left column are explicit queries generated by a user, part of which also contain tag or color search conditions.

these three query types, and they could be added in any order. More illustrations could be seen in Fig. 6.

3. TECHNICAL DETAILS

To generate the back-end database, we totally downloaded 2,114,085 Flickr photos with tag information using the top 1000 hot queries. We first downsampled each image to a suitable size, then adopted a saliency discovering method to extract major curves, by which each full color images is transformed into a binary map. Unlike former query-by-sketch methods, we adopt an raw curve-based algorithm to achieve precise matching between sketch queries and database images. Besides, this process is further speeded up by an index structure. The tag and color features are indexed by inverted file structures, which totally take less than 1GB memory. The sketch index occupies less than 7GB memory, and thus MindFinder could be easily applied on a normal Intel machine with 4 cores and 16GB memory. Since all index structures were pre-built offline, by our well designed architecture, a typical response time of a complex query is between 1 and 2 seconds.

An important issue should be considered is that how to collaborate multimodal search conditions. A simple way is to separately retrieve top M images by sketch, tag and color, and then merge them into one list. However, it is impractical since in such a huge database, overlap images among all returned sets are rare, which significantly reduces the quality of the final combination. To deal with this problem, we use only one query type in the mixed query to retrieve an image set, and other query types are worked as reranking conditions on this set. In the implementation, we choose tag as the dominant search condition (to cross the semantic gap), followed by dominant color (to specify the global hue). By elaborately choosing a suitably large size of this set, our solution could make sure that top returned images meet all search conditions to a great extent.

4. CONCLUSIONS

The main contributions of the proposed system are summarized as follows:

1. MindFinder is the first sketch-based multimodal search engine for more than two million web images.
2. Our system provides a convenient interface for users to freely express their search intentions, and enables real-time interactions for users to more efficiently locate their desired images.
3. It is the first index-based query-by-sketch solution for million level database.

5. ACKNOWLEDGEMENTS

The work of Yang Cao and Liqing Zhang was partially supported by the Science and Technology Commission of Shanghai Municipality (Grant No. 08511501701) and the National Natural Science Foundation of China (Grant No. 60775007, 90920014).

6. REFERENCES

[1] R. Datta, D. Joshi, J. Li, and J. Wang. Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys*, 2008.

[2] M. Eitz, K. Hildebrand, T. Boubekeur, and M. Alexa. A descriptor for large scale image retrieval based on sketched feature lines. In *Proceedings of the 6th Eurographics Symposium on Sketch-Based Interfaces and Modeling*, 2009.

[3] M. Eitz, K. Hildebrand, T. Boubekeur, and M. Alexa. PhotoSketch: a sketch based image query and compositing system. In *SIGGRAPH*, 2009.

[4] C. Jacobs, A. Finkelstein, and D. Salesin. Fast multiresolution image querying. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, 1995.