An Anatomy of a Large-scale Image Search Engine

Wei-Cheng Lai, Edward Chang, and Kwang-Ting (Tim) Cheng
Morpho Software Inc.
3944 State Street, Suite #340
Santa Barbara, CA 93105
{wlai,echang,timcheng}@morphosoft.com

1 Introduction

As the World-Wide Web moves rapidly from text-based towards multimedia content, and requires more personalized access, we deem existing infrastructures inadequate. In this paper, we present critical components for enabling effective searches in Web-based or large-scale image libraries. In particular, we propose a *perception-based search component*, which can *learn* users' subjective query concepts quickly through an intelligent sampling process. Through an example, we demonstrate how, and explain why our perception-based search paradigm can effectively personalize a query and achieve high recall.

1.1 An Illustrative Example

In this example, we compare a keyword-based image retrieval system with our proposed perception-based image retrieval system. We use the Yahoo! Picture Gallery (i.e., http://gallery.yahoo.com) as a test site for keyword-based image retrieval. Suppose a user wants to retrieve images related to "bird of paradise." Given the keywords "bird of paradise" at the test site, the gallery engine retrieves five images of this flower.

However, there are more than five images relevant to "bird of paradise" in the Yahoo image database. Our system can retrieve more of these relevant images. First, we query Yahoo's keyword-based search engine using "bird" and "flowers" and store the returned images (both birds and flowers) in a local database. Second, we apply our perception-based search engine to the local database. The learning steps for grasping the concept "bird of paradise" involves three screens that are illustrated in the following three figures.

- Screen 1. Sampling and relevance feedback starts. The screen is split into two frames horizontally. On the left-hand side of the screen is the learner frame; on the right-hand side is the similarity search frame. Through the learner frame, the system learns what the user wants via an active learning process. The similarity search frame displays images that match the user's query concept. The system presents a number of samples in the learner frame, and the user marks images that are relevant to his or her query concept by clicking on the relevant images. As shown in Figure 1, one image (the last image in the first row) is selected as relevant, and the rest of the unmarked images are considered irrelevant. The user indicates the end of his or her selection by clicking on the submit button in the learner screen. This action brings up the next screen.
- Screen 2. Sampling and relevance feedback continues. Figure 2 shows the second screen. First, the similarity search frame displays what the system thinks will match the user's query concept at this time. As the figure

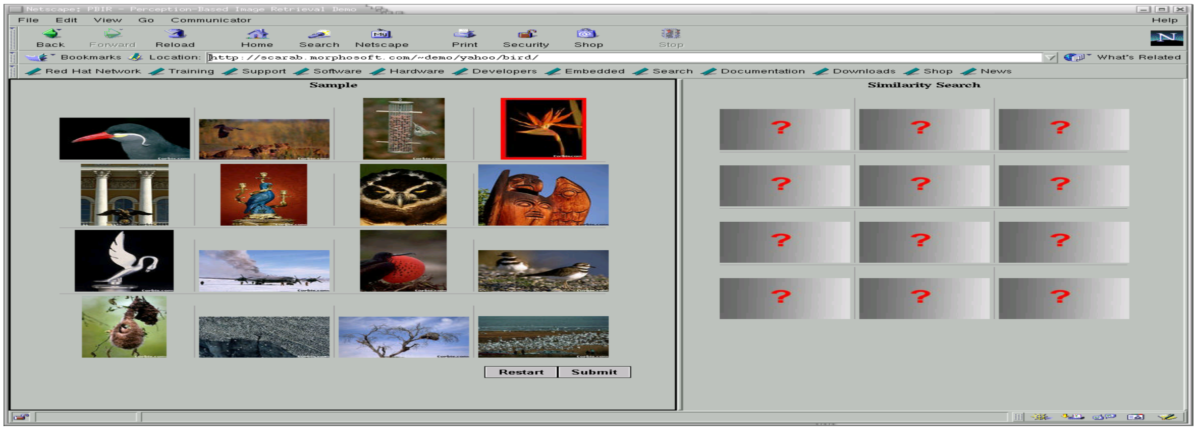

Figure 1: “Bird of Paradise” Query Screen #1.

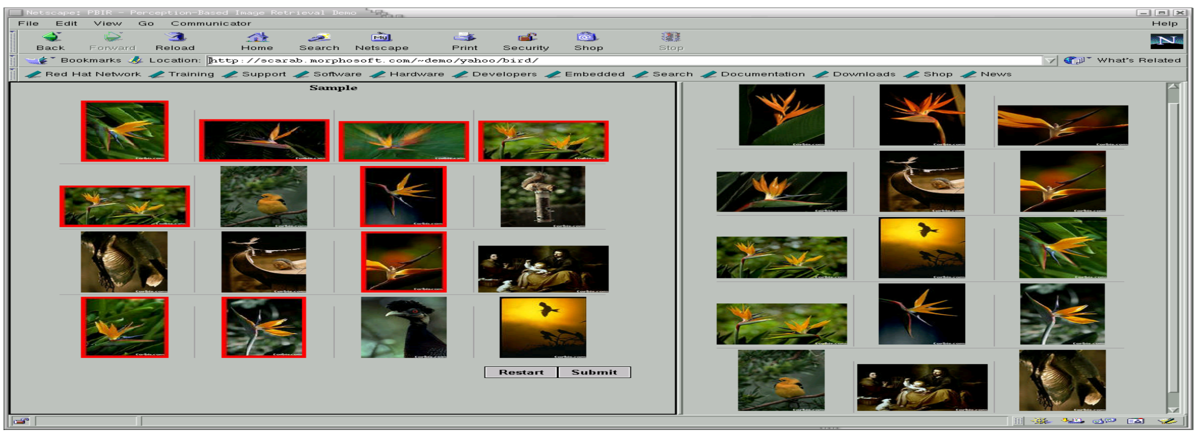

Figure 2: “Bird of Paradise” Query Screen #2.

indicates, eleven returned images fit the concept of “bird of paradise.” The user’s query concept has been captured, though somewhat fuzzily. The user can ask the system to further refine the target concept by selecting relevant images in the learner frame. In this example, nine images (four images from the first row, the first and the third images from the second row, the third image from the third row, and the first two images from the last row) are relevant to the concept. After the user clicks on the submit button in the learner frame, the third screen is displayed.

- Screen 3. Sampling and relevance feedback ends. Figure 3 shows that all returned images in the similarity search frame fit the query concept (bird of paradise).

As observed, in two iterations, our system is able to retrieve fifteen relevant images from the image database. In this example, we use the keyword-based search engine to seed our perception-based search engine. The keyword-based search engine can be used to quickly identify the set of images relevant to the specified keywords. Based on these relevant images, the perception-based search engine can explore the feature space and discover more images relevant to the users’ query concept. Note that our perception-based search system will also work without seedings from a keyword-based search engine. An on-line system prototype is available at [1].

The above example illustrates that the perception-based search paradigm achieves much higher recall because it avoids the following limitations that the traditional keyword-only search paradigm encounters:

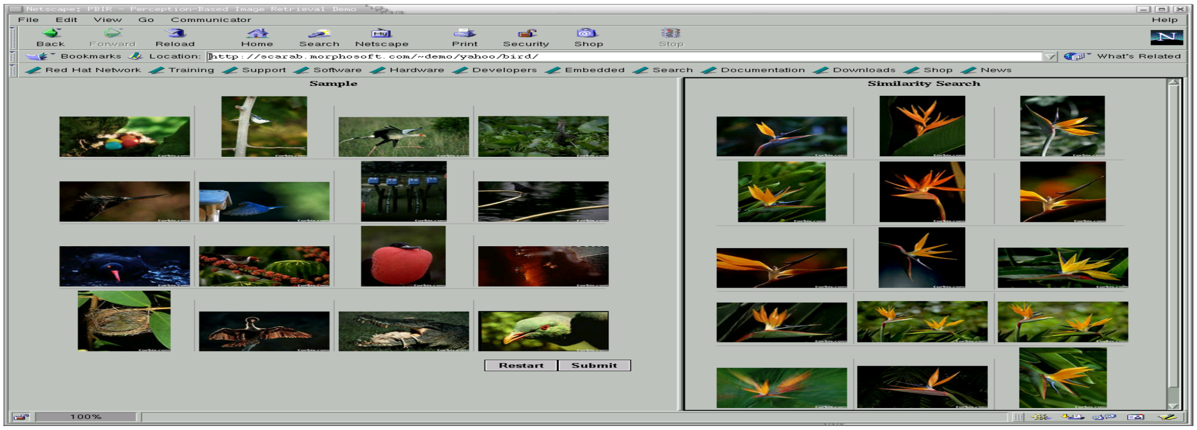

Figure 3: "Bird of Paradise" Query Screen #3.

1. *Subjective annotation.* As we can see from the example, a "bird of paradise" image may be annotated as "bird," "flower," "hawaii," and many other possible words. Using one of the words to conduct a keyword search cannot get the images labeled by the other words.

2. *Terse annotation*. The annotation of an image typically does not have as many words as that in a text document. With limited number of keywords, keyword annotation often cannot faithfully and completely capture images' semantics.

3. *Incomplete query-concept formulation*. A picture is worth more than a thousand words. Thus, a few query keywords can hardly characterize a complete query concept.

In summary, it is evident that with incomplete query-concept formulation and incomplete image characterization, the keyword-only search approach faces severe limitations to achieve high recall.

2 System Architecture

We present the system components that make perception-based image retrieval work: *multi-resolution feature extractor*, *perception-based search engine*, and *high-dimensional indexer*.

2.1 Multi-resolution Feature Extractor

Feature extractor extracts perceptual features from images. Common perceptual features are color, shape, texture, and spatial layout of these features. Feature extraction can be performed off-line; however, since the number of images can be large, feature extraction should be both efficient and effective. For representative features and how they are organized in a multi-resolution fashion to speed up learning performance, please consult [2, 4, 5].

2.2 Perception-based Search Engine

The perception-based search engine is the heart of enabling personalized image retrieval. The engine learns users' query concepts as learning a binary classifier that separates the images relevant to the query concept from the irrelevant ones. The learning takes place in an iterative process: The system presents examples to the users to refine the class boundary. The final class boundary is learned based on users' relevance feedback.

Relevance feedback is not new. Unfortunately, traditional relevance feedback methods require a large number of training instances to converge to a target concept, and therefore not practical. In our perception-based engine, we explore several active learning algorithms, which can "grasp" a query profile with a small number of training instances. We recently proposed two active learning algorithms, MEGA (The Maximizing Expected Generalization Algorithm) [2, 4] and SVM_{Active} (Support Vector Machine Active Learning) [6], to tackle the problem effectively. Please consult these publications for details.

2.3 High-dimensional Indexer

An image is often characterized by a large number of features (more than one hundred). A query concept may be best characterized by a selected subset of the features. To deal with the "dimensionality-curse" problem and to support dynamic subspace searching, we propose an indexing scheme using clustering and classification methods for supporting *approximate similarity searches*. Our indexing method is a statistical approach that works in two steps. It first performs non-supervised clustering using Tree-Structured Vector Quantization (TSVQ) [3] to group similar objects together. To maximize IO efficiency, each cluster is stored in a sequential file. A similarity search is then treated as a classification problem. Our hypothesis is that if a query object's class prediction yields C probable classes, then the probability is high that its nearest neighbors can be found in these C classes. This hypothesis is analogous to looking for books in a library. If we want to look for a calculus book and we know calculus belongs in the math category, by visiting the math section we can find many calculus books. Similarly, by searching for the most probable clusters into which the query object might be classified, we can harvest most of the similar objects.

3 Conclusion

This paper proposes an image retrieval system that uses active learning to capture complex and subjective query concepts. We presented key supporting components—a multi-resolution image-feature extractor and a high-dimensional indexer— for making both query-concept learning and image retrieval efficient. An on-line system prototype is available at [1].

References

[1] E. Chang, K.-T. Cheng, W.-C. Lai, C.-T. Wu, C.-W. Chang, and Y.-L. Wu. PBIR — a system that learns subjective image query concepts. *Proceedings of ACM Multimedia, http://www.mmdb.ece.ucsb.edu/˜demo/corelacm/*, pages 611–614, October 2001.

[2] E. Chang and B. Li. Mega — the maximizing expected generalization algorithm for learning complex query concepts (extended version). *Technical Report http://www-db.stanford.edu/∼echang/mega-extended.pdf*, November 2000.

[3] A. Gersho and R. Gray. *Vector Quantization and Signal Compression*. Kluwer Academic, 1991.

[4] B. Li, E. Chang, and C.-S. Li. Learning image query concepts via intelligent sampling. *Proceedings of IEEE Multimedia and Expo*, August 2001.

[5] Y. Rui, T. S. Huang, and S.-F. Chang. Image retrieval: Current techniques, promising directions and open issues. *Journal of Visual Communication and Image Representation*, March 1999.

[6] S. Tong and E. Chang. Support vector machine active learning for image retrieval. *Proceedings of ACM International Conference on Multimedia*, pages 107–118, October 2001.