PBIR: Perception-Based Image Retrieval— A System That Learns Subjective Image Query Concepts

Edward Chang, Kwang-Ting Cheng,
Wei-Cheng Lai, Ching-Tung Wu
Electrical & Computer Engineering
University of California
Santa Barbara, CA 93106
{echang,timcheng}@ece.ucsb.edu

Chengwei Chang, Yi-Leh Wu
Morpho Software Inc.
3944 State Street, Suite #340
Santa Barbara, CA 93105
{bchang,ywu}@morphosoft.com

ABSTRACT

We describe the Perception-Based Image Retrieval (PBIR) system that we have built on our recently developed query-concept learning algorithms, MEGA and SVM_{Active}. We show that MEGA and SVM_{Active} can learn a complex image-query concept in a small number of user iterations (usually three to four) on a large, multi-category, high-dimensional image database.

Keywords
active learning, image retrieval, query concept, relevance feedback, support vector machines.

1. OVERVIEW

For multimedia search tasks, a query concept is hard to articulate, and articulation can be subjective. For instance, in an image search, it is difficult to describe a desired image using low-level features such as color, shape, and texture. Different users may perceive the same image differently; and even if an image is perceived similarly, users may use different vocabulary (i.e., different combinations of low-level features) to depict it. Furthermore, most users are not trained to specify simple query criteria using for example, SQL. It is thus both necessary (for capturing subjective concepts) and desirable (for alleviating users from specifying complex queries) to build intelligent search engines that can quickly learn users' query concepts via relevance feedback.

Unfortunately, traditional relevance feedback and learning techniques may not be suitable for online query-concept learning for at least two reasons. First, an online user is typically impatient and cannot be expected to provide a great deal of feedback, but traditional learning methods (e.g., decision trees and neural networks) often require a large number of training instances to learn a concept [6]. Second, almost all traditional relevance feedback methods [3, 7, 9] require users to provide good examples to seed a query. However, finding good seeds is the job of the search engine itself, and this circular requirement leaves the core problem— learning and understanding users' query concepts—unsolved.

In our demo, we present an image search engine that employs our query-concept learning algorithms—MEGA [1, 4] and SVM_{Active} [8]—to learn users' query concepts. We show that both MEGA and SVM_{Active} can learn complex image-query concepts quickly and in a small number of user iterations. To demonstrate that MEGA and SVM_{Active} are scalable in feature dimension, in dataset size, and in concept complexity, we construct our demo system with more than 50K *diversified* images. Each image is characterized by more than 200 features.

The focus of this demo is to show that our learning algorithms can quickly grasp a complex query concept with a small number of training examples. (Although our system also relies on high-dimensional indexers [5], and perception-based image characterization [2] to be fast, robust, and scalable, we only discuss our learning algorithms in this description.)

In the rest of this demo description, we briefly depict in Section 2 how MEGA and SVM_{Active} work, and we present in Section 3 query scenarios and flows conducted on our system.

2. LEARNING ALGORITHMS

While we describe MEGA and SVM_{Active} in detail in [1] and [8], respectively, we highlight these two learning algorithms in this section.

2.1 MEGA

The *Maximizing Expected Generalization Algorithm* (MEGA) models query concepts in k-CNF, which can model virtually all practical query concepts. To ensure that target concepts can be learned quickly and with a small number of samples, MEGA employs two steps iteratively: sample selection (S-step) and feature reduction (F-step).

1: In its S-step, MEGA uses a k-DNF to bound the sampling space and judiciously selects samples that can collect maximum information from users to remove irrelevant k-CNF and k-DNF terms in its subsequent F-step.

2: In its F-step, MEGA removes irrelevant terms from the query-concept (i.e., a k-CNF), and at the same time, refines the sampling boundary (i.e., a k-DNF) so that the most informative samples can be selected in its subsequent S-step.

Unlike traditional query refinement methods, which use only the S-step or only the F-step, MEGA uses these two steps in a complementary way to achieve fast convergence to target concepts.

Note that MEGA does not require seeding a query. MEGA presents randomly selected images as the first round of examples. Even if all of the images generated in the first round are irrelevant, MEGA uses the irrelevant images to reduce the set of potentially relevant images substantially (the k-DNF that bounds the sampling space can be shrunk substantially by the irrelevant images). Thus, the probability is higher that a relevant image is sampled in the next round.

2.2 SVM$_{Active}$

SVM$_{Active}$ is an *active learning* scheme that uses support vector machines. Intuitively, SVM$_{Active}$ works by combining the following three ideas:

1. SVM$_{Active}$ regards the task of learning a target concept as one of learning an SVM binary classifier. An SVM captures the query concept by separating the relevant images from the irrelevant images with a hyperplane in a projected space, usually a very high-dimensional one. The projected points on one side of the hyperplane are considered relevant to the query concept and the rest irrelevant.
2. SVM$_{Active}$ learns the classifier quickly via active learning. The active part of SVM$_{Active}$ selects the most informative instances with which to train the SVM classifier. This step ensures fast convergence to the query concept in a small number of feedback rounds.
3. Once the classifier is trained, SVM$_{Active}$ returns the top-k most relevant images. These are the k images that are farthest from the hyperplane on the query concept side.

3. ILLUSTRATIVE EXAMPLES

We present examples in this section to show the learning steps of MEGA and SVM$_{Active}$ in two image query scenarios: *image browsing* and *similarity search*.

- *Image browsing*. A user knows what he/she wants but has difficulty articulating it. Through an interactive browsing session, MEGA or SVM$_{Active}$ learns what the user wants.
- *Similarity search*. After MEGA or SVM$_{Active}$ knows what the user wants, the search engine can perform a traditional similarity search to find data objects that appear similar to a given query object.

3.1 MEGA Query Steps

In the following, we present an interactive query session using MEGA. This interactive query session involves seven screens that are illustrated in seven figures. The user's query concept in this example is "wild animals." Due to the space limitation, we skip presenting screens #3 and #5.

Screen 1. Initial Screen. Our Perception-Based Image Retrieval (PBIR) system presents the initial screen to the user as depicted in Figure 1. The screen is split into two frames vertically. 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, PBIR learns what the user wants via an intelligent sampling process. The similarity search frame displays what the system thinks the user wants. (The user can set the number of images to be displayed in these frames.)

Screen 2. Sampling and relevance feedback starts. Once the user clicks the "submit" button in the initial frame, the sampling and relevance feedback step commences to learn what the user wants. The PBIR system presents a number of samples in the learner frame, and the user highlights images that are relevant to his/her query concept by clicking on the relevant images. As shown in Figure 2, three images (the third image in rows one, two and four in the learner frame) are selected as relevant, and the rest of the unmarked images are considered irrelevant. The user indicates the end of his/her selection by clicking on the submit button in the learner screen. This action brings up the next screen.

Screen 4. Sampling and relevance feedback continues. Figure 3 shows the fourth screen. First, the similarity search frame displays what the PBIR system thinks will match the user's query concept at this time. As the figure indicates, the top nine returned images fit the concept of "wild animals." 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, the fourth image in the second row and the third image in the fourth row are selected as relevant to the concept. After the user clicks on the submit button in the learner frame, the fifth screen is displayed.

Screen 6. Sampling and relevance feedback ends. Figure 4 shows that all returned images in the similarity search frames fit the query concept (wild animals).

Screen 7. Similarity search. At any time, the user can click on an image in the similarity search frame to request images that *appear similar* to the selected image. This step allows the user to zoom in onto a specific set of images that match some appearance criteria, such as color distribution, textures and shapes. As shown in Figure 5, after clicking on one of the tiger images, the user will find similar tiger images returned in the similarity search frame. Notice that other wild animals are ranked lower than the matching tiger images, since the user has concentrated more on specific appearances than on general concepts.

In summary, in this example we show that our PBIR system effectively uses MEGA to learn a query concept. The images that match a concept do not have to appear similar in their low-level feature space. The learner is able to match high-level concepts to low-level features directly through an intelligent learning process. Our PBIR system can capture images that match a concept through MEGA or SVM$_{Active}$, whereas the traditional image systems can do only appearance similarity searches. Again, as illustrated by this example, MEGA can capture the query concept of wild animal (wild animals can be elephants, tigers, bears, and etc), but a traditional similarity search engine can at best select only animals that appear similar. Another strength of the system is its runtime efficiency. Each round of relevance feedback takes less than half a second to complete. Finally, our PBIR system does not require users to seed a query with positive images as systems like MARS [7] and FALCON [9] require.

4. CONCLUSION

Both MEGA and SVM$_{Active}$ are pioneer work for learning transient online user behavior. In an on-line learning setting, a learner is severely constrained by time and by scarcity of data, and the past access pattern of a user may not be applicable to his/her current behavior. Through our demo, we show that our learning algorithms can quickly grasp a complex query concept with a small number

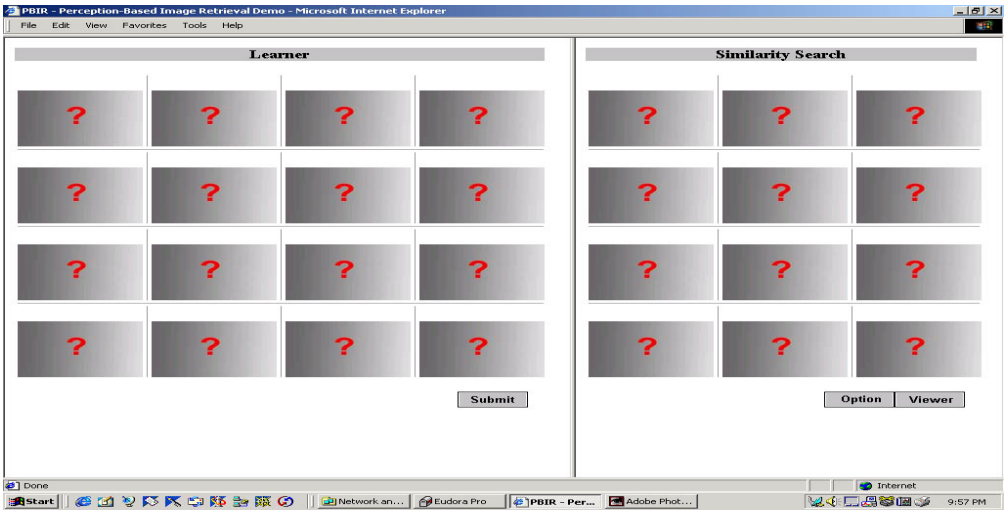

Figure 1: Wild Animal Query Screen #1.

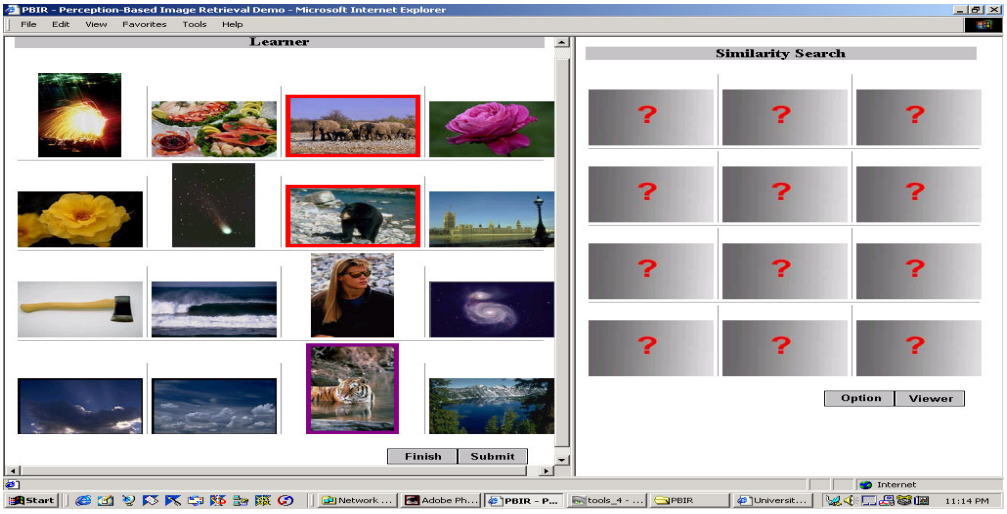

Figure 2: Wild Animal Query Screen #2.

of labeled instances. The learning is done directly in a very high-dimensional feature space via relevance feedback. Although we demonstrate MEGA and SVM_{Active} through an image system, we believe that these learning algorithms can be applied to search for other types of media data and can be used in applications such as information filtering and e-commerce.

5. REFERENCES

[1] 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.ps*, November 2000.

[2] E. Chang, B. Li, and C. Li. Towards perception-based image retrieval. *IEEE Content-Based Access of Image and Video Libraries*, pages 101–105, June 2000.

[3] K. S. Jones and P. W. (Editors). *Readings in Information Retrieval*. Morgan Kaufman, July 1997.

[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] C. Li, E. Chang, H. Garcia-Molina, and G. Wiederhold. Clustering for approximate similarity queries in high-dimensional spaces. *IEEE Transaction on Knowledge and Data Engineering (to appear)*, 2001.

[6] T. M. Mitchell. *Machine Learning*. McGraw-Hill, 1997.

[7] K. Porkaew, K. Chakrabarti, and S. Mehrotra. Query refinement for multimedia similarity retrieval in mars. *Proceedings of ACM Multimedia*, November 1999.

[8] S. Tong and E. Chang. Support vector machine active learning for image retrieval. *ACM International Conference on Multimedia*, October 2001.

[9] L. Wu, C. Faloutsos, K. Sycara, and T. R. Payne. Falcon: Feedback adaptive loop for content-based retrieval. *The 26th VLDB Conference*, September 2000.

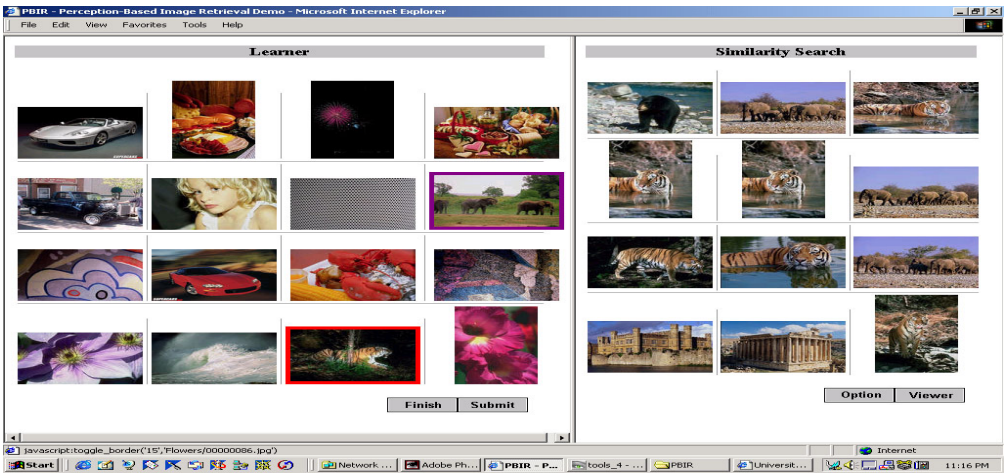

Figure 3: Wild Animal Query Screen #4.

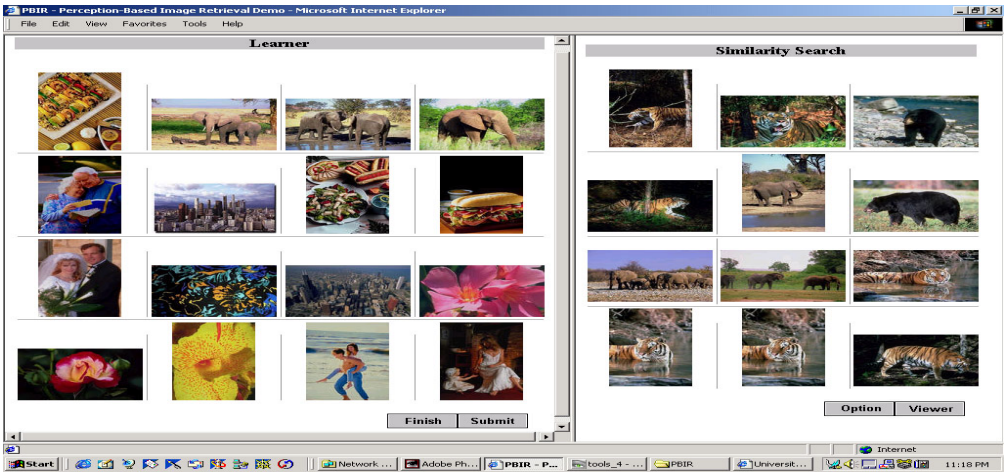

Figure 4: Wild Animal Query Screen #6.

Figure 5: Wild Animal Similarity Query (Screen #7).