PARAgrab: A Comprehensive Architecture for Web Image Management and Multimodal Querying

Dhiraj Joshi [1] Ritendra Datta [1] Ziming Zhuang [2]
WP Weiss [2] Marc Friedenberg [2] Jia Li [3,1] James Z. Wang [2,1].

[1]Department of Computer Science and Engineering,
[2]College of Information Sciences and Technology, and [3]Department of Statistics
The Pennsylvania State University, University Park, PA, USA.

ABSTRACT

We demonstrate PARAgrab - a scalable Web image archival, retrieval, and annotation system that supports multiple querying modalities. The underlying architecture of our large-scale Web image database is described. Querying and visualization techniques used in the system are explained.

1. INTRODUCTION

Digital image databases have attracted a large amount of commercial and academic interest in the recent years. Research has shown that image databases require differential treatment. The semantic content of pictures is significantly different from their digitized representation, a problem often referred to as *semantic gap* [10]. Moreover, only a fraction of images have manual labels and these labels often suffer from *synonymy* and *polysemy* [5]. As an alternative, people have attempted to automate image interpretation by their visual content, a technology known as content-based image retrieval (CBIR). Comprehensive surveys of CBIR can be found in [10, 1].

Design and management of very large multimedia databases is a challenging task. An experimental system supporting multi-paradigm querying for multimedia databases is presented in [12]. Adaptation of similarity searches to user needs in large multimedia databases has been addressed in [8]. Clustering in very large spatial image databases has been studied in [9]. With the explosion of the Internet, a number of Web-based image retrieval systems have been proposed. Among those that incorporate visual features into the retrieval process are the iFind system [13] and the Cortina system [7]. The SIMPLIcity system [11] has been applied to art, photography and other domain-specific image databases.

VLDB'06 September 12–15, 2006, Seol, Korea.

2. SYSTEM ARCHITECTURE

PARAgrab has been designed in a modular fashion to ensure (1) efficient processing of large volumes of dynamically changing WWW image data, and (2) smoothness and parallelism of the work-flow. The efficiency in the work-flow is achieved primarily from the fact that the dynamic nature of the Web is reflected by the crawling/indexing modules independent and transparent to the query processing module. Fig. 1 shows the PARAgrab architecture.

2.1 Modular Design

The PARAgrab system has the following main modules:

Crawler Module: A focused crawling strategy has been adopted for our system. If a URL has successfully led to an image of interest, the URLs nearby and those at the same level are considered more *promising*. As a result, they are prioritized so that in the future, they will be explored first. Authoritative image Websites, such as PBase and Flickr, are used as seed URLs.

Indexing Module: Building indexes offline in order to expedite online retrieval is critical to any large scale database. Since the data being collected is highly multi-modal in nature, separate indexes need to be maintained to facilitate simple or combined query modalities. We attack this problem by using a MySQL database to store image meta-data and text attributes. We currently use MySQL to build text indexes. The method of building indexes on image features involves running a batch job periodically to process newly added pictures. The job consists of multiple parallel processes run on a Linux cluster.

Query Processing Module: The query processing module interacts directly with the meta-data and image indexes, a cache of ranked images, and a content filtering module to generate search results for the interface module. The query modalities supported are described in Sec. 3.

Interface Module: In PARAgrab, we aim at achieving a unified visualization of search results in order to maintain high user interest levels. Our interface, as shown in Fig. 2, presents search results in a simple and concise way so that users can search and browse with ease. Moreover, the interface provides usage logs for potential improvements in retrieval performance.

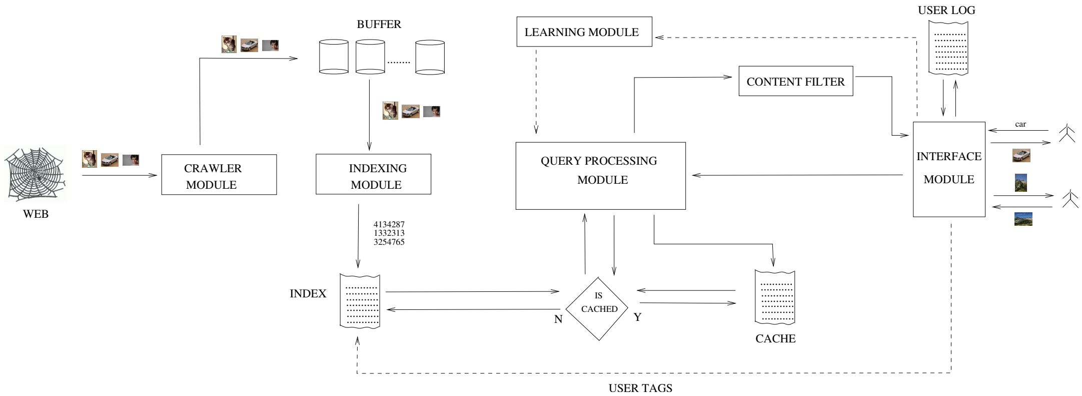

Figure 1: A schematic view of the PARAgrab architecture.

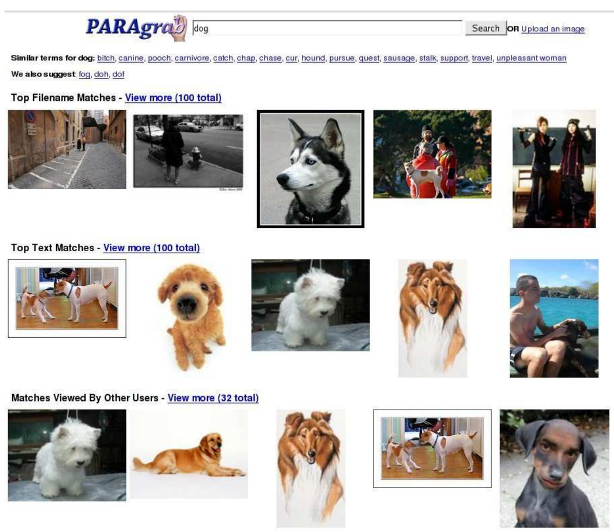

Figure 2: A screenshot of the search results obtained for the query *dog*. Note the unified visualization of search results.

2.2 Image Ranking

In PARAgrab, visual content based image retrieval is performed using the Integrated Region Matching (IRM) distance [11]. A key advantage of using IRM is that image ranking, over large collections, can be performed in real time. This scalability is achieved by modularizing the retrieval process into separate database indexing and distance computation steps.

Among text-based methods, our system currently supports user tag and URL based image search. Tag matching is performed in a way analogous to text based querying. URLs may not always be useful in determining image semantics. Yet, sometimes very high level information is embedded in the URL, e.g., name of the photographer, subject, organization, institute, etc. Our algorithm for URL based matching is based on two intuitions: (1) An image is more relevant to a query if it is closer to the right-most segment (2) An exact query match has greater relevance than a match with its stemmed version.

3. QUERYING AND VISUALIZATION

The ultimate verdict on the success of a Web-based system is typically made by its end-users. In the PARAgrab system, we essentially walk the tight-rope between a feature-rich system and a user-friendly interface. The user is welcomed with a random sample of 10 images from the picture database. The database journey can be started by (1) keyword query, (2) clicking on an image, and (3) personal file upload. For any picture, the user can add/remove tags associated with the picture. Fig. 2 depicts a typical keyword-based search result, consisting of three ranked lists put together in a compact representation. Clicking on a picture launches the visual similarity search. PARAgrab supports the following query refinement features:

Spelling Suggestions: For every query word, the system computes its Levenshtein (edit) distance with each unique tag in order to enlist closely spelled words (within a threshold), if any. We rank spelling suggestions in the increasing order of their edit distances to the query.

Semantics Expansion: Query keywords yield WordNet based synonyms, hypernyms, hyponyms, and sister terms ordered by familiarity of use [5]. These words are recommended to the user, hyper-linked in a way that initiates new keyword-based queries.

The primary advantage of content based image retrieval is to discover images pertaining to a given concept in the absence of reliable meta-data. In Fig. 3, we show how a combination of text-search followed by visual-search achieves this goal. A two-level search beginning with the text query *cat* yields a number of pictures of cats which did not have any cat-related meta-data associated with them. Boxes are drawn around the images used to initiate further visual searches, in Fig. 3 (a). The newly discovered cat images are shown enclosed within boxes in Fig. 3 (b) and (d). This strategy has immense potential for human-guided tag propagation i.e., the newly discovered cat images can now be tagged by the user to assist in future keyword based queries. Interestingly, another second level visual search identifies near duplicates of the query image, as shown in Fig. 3 (c). Fig. 4 shows another instance where visual search retrieves near-duplicates.

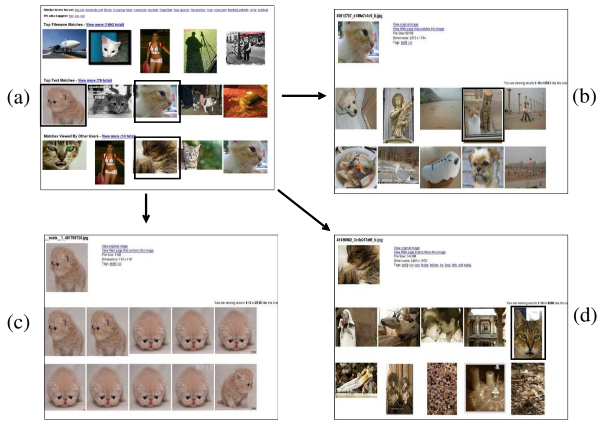

Figure 3: Depiction of a two level search using the PARAgrab system. The top-left figure shows the results obtained for the text query *cat*. The other three figures show the result of visual search performed on images from the first step. Boxes are drawn around the specific pictures clicked.

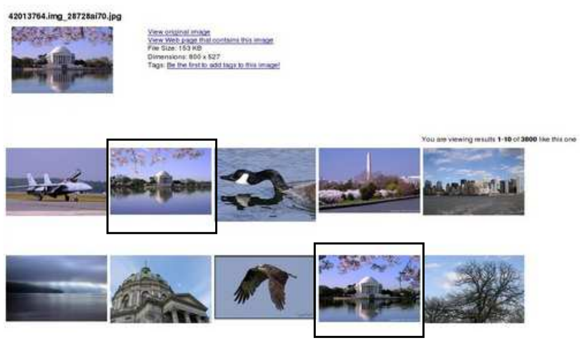

Figure 4: A classic case where visual similarity helps discover near-duplicates.

Typically, in a text search engine the only way a user can browse for relevance is through inspection of result pages in rank order. However, in PARAgrab, given the facility to search by image content, a user can explore pictures in sequential chains of successive similarity queries. The variety in such chains arises from the choice the user makes at each step in order to proceed with browsing. We refer to these as:

Strategy 1: Begin with a single query image and browse result pages strictly in the order of visual similarity to query.

Strategy 2: Begin with a single query image and then hierarchically explore the next level, making each top level result as a subsequent query.

Strategy 3: Begin with a single query, and keep performing visual searches based on personal preference and relevance to query. The sequence thus can potentially deviate further and further from the original query.

Fig. 5 presents a conceptual representation of the discussed search strategies. We next conduct a short study on three text queries namely *mountain*, *building*, and *people*, using the results as seeds for employing different strategies. We

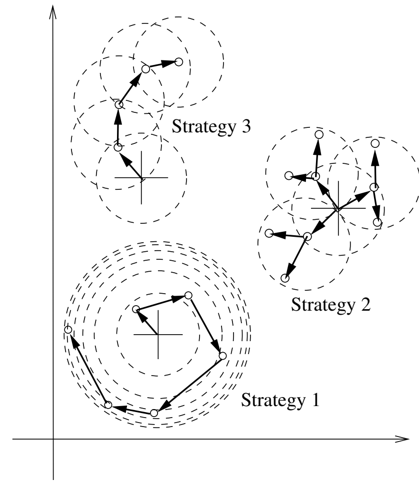

Figure 5: Conceptual representation of three browsing strategies. Circles represent local neighborhoods in the image feature space. Arrows represent jumps during browsing.

restrict this study to the strategies 1 and 2. In Fig. 6, the number of *un-tagged* relevant images obtained are recorded and plotted with the number of images browsed. All numbers are averaged over the first three seed images of *mountain*, *building*, and *people* respectively. It is evident that such compound browsing schemes are capable of revealing images otherwise impossible for simple text based searches. The number of *un-tagged* relevant images discovered in the 110 image browsed (11 user clicks, ignoring browser buttons) for *mountain*, *building*, and *people* is about 200%, 99%, and 62% more than the number obtained by text-query alone. While we observe no significant differences in the two strategies, differences are noticed among different query concepts in relevant image discovery. Given the flexibility of strategy 3, whether its employment makes any difference would be an interesting new direction.

4. SIGNIFICANCE OF CONTRIBUTION

While there has been extensive research in the field of image retrieval, we believe that our key contribution is that of integrating many successful ideas from multiple research domains into a large-scale working system. We feel that our work will promote image retrieval in the database research community and generate interest in database-related aspects of the problem. Some other contributions are summarized below.

(1) In PARAgrab, we combine structured (meta-data) and unstructured (visual features) information into an integrated image management framework.

(2) Our modular architecture is capable of scaling up to handle very large volumes of data. With a database of about half a million Web images and growing, PARAgrab is publicly available for wide usage [6].

(3) Multi-modal searching by keywords, visual examples, and local image file uploads are supported. To the best of our knowledge, PARAgrab is the first large-scale working system that incorporates all these modalities into a single framework.

(4) Besides the visual and text based ranking methods

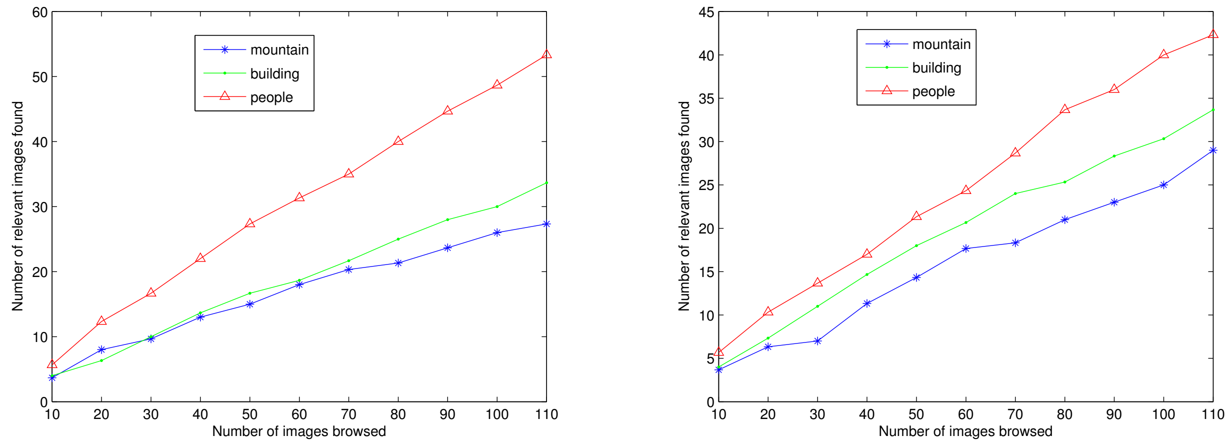

Figure 6: Plots of the number of *un-tagged* relevant images discovered using strategies 1 and 2 respectively. The X axis represents the number of images browsed.

discussed before, PARAgrab also supports popularity-based ranking.
(5) We incorporate natural language processing techniques in image search to enable query refinement.
(6) Offline caching of visual similarity ranking is performed to support real-time search. Caching is performed at regular intervals to reflect the dynamic nature of the database.
(7) In a fashion similar to Flickr, PARAgrab supports adding and removing picture tags. We collect the human provided annotations as meta-data to improve future search by keywords.
We believe that PARAgrab architecture will be a testbed for new ideas and approaches to retrieval. In the near future, we plan to extend the system to support techniques such as clustering, relevance feedback, aesthetics based ranking [2], and story picturing [4]. We will also make the data collected from user interactions available for research purposes.

5. DEMONSTRATION PLAN

Our demonstration will include a comprehensive walk-through of the PARAgrab system.We will present the various querying modalities namely (1) visual image search, (2) text based image search, and (3) search using local image file uploads. Through this, we will demonstrate the advantages of combining multiple modalities for enhanced image search, contrasting it with Google Images. The query refinement features will also be elaborated during the demonstration session. We will compare and contrast the browsing strategies (Fig. 6) and how they affect relevance to query. The advantages of rank combination will be shown through various examples. In this process, feedback through audience participation will help us understand end-user requirements better. We expect that this experience will familiarize the database community with the challenges involved in building useful image retrieval systems.

6. ACKNOWLEDGMENTS

This work is supported in part by the US National Science Foundation under Grant Nos. IIS-0219272, IIS-0347148, and ANI-0202007, The Pennsylvania State University, the PNC Foundation, and SUN Microsystems.

7. REFERENCES

[1] R. Datta, J. Li, and J. Z. Wang, "Content-Based Image Retrieval - A Survey on the Approaches and Trends of the New Age," *Proc. MIR Workshop*, 2005.

[2] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Studying Aesthetics in Photographic Images Using a Computational Approach," *Proc. ECCV*, 2006.

[3] H. Feng, R. Shi, T. S. Chua, "A Bootstrapping Framework for Annotating and Retrieving WWW Images," *Proc. ACM Multimedia*, 2004.

[4] D. Joshi, J. Z. Wang, and J. Li, "The Story Picturing Engine - A System for Automatic Text Illustration," *ACM Trans. on Multimedia Computing and Applications*, 2(1):68-89, 2006.

[5] G. Miller, "WordNet: A Lexical Database for English," *Comm. of the ACM*, 38(11):39-41, 1995.

[6] PARAgrab, `http://paragrab.psu.edu`.

[7] T. Quack, U. Monich, L. Thiele, and B. S. Manjunath, "Cortina: A System for Largescale, Content-based Web Image Retrieval," *Proc. ACM Multimedia*, 2004.

[8] T. Seidl and H.-P. Kriegel, "Efficient User-Adaptable Similarity Search in Large Multimedia Databases," *Proc. Int. Conf. Very Large Databases*, 1997.

[9] G. Sheikholeslami, S. Chatterjee, and A. Zhang, "WaveCluster: A Multi-Resolution Clustering Approach for Very Large Spatial Databases," *Proc. Int. Conf. Very Large Databases*, 1998.

[10] A. W. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-Based Image Retrieval at the End of the Early Years," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, 2000.

[11] J.Z. Wang, J. Li, and G. Wiederhold, "SIMPLIcity: Semantics-Sensitive Integrated Matching for Picture Libraries," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 23(9), 947–963, 2001.

[12] J.-R. Wen, Q. Li, W.-Y. Ma, and H.-J. Zhang, "A Multi-paradigm Querying Approach for a Generic Multimedia Database Management System," *Proc. SIGMOD*, 2003.

[13] H. J. Zhang, L. Wenyin, C. Hu, "iFind - A System for Semantics and Feature Based Image Retrieval over Internet," *Proc. ACM Multimedia*, 2000.