A forward-looking user interface for CBIR and CFIR systems

Liam M. Mayron
and Oge Marques
Florida Atlantic University
Boca Raton, FL 33431
{lmayron, omarques}@fau.edu

Gustavo B. Borba
and Humberto R. Gamba
Universidade Tecnológica Federal do Paraná
Curitiba, PR - Brazil
{gustavo,humberto}@cpgei.cefetpr.br

Vladimir Nedovic
University of Amsterdam
Amsterdam, Netherlands
vnedovic@science.uva.nl

***Abstract*— This demonstration highlights the benefits that image retrieval systems can enjoy by use of a thoughtful interface. We present a live demonstration of PRISM, a new Web-based application that, through a simple set of natural actions, allows the user the ability to elicit sophisticated responses from easily-formed queries. Additionally, the system was designed with not only content-based, but also content-free image retrieval and semantic annotation in mind. The demonstration system uses an attention-based method for extracting and determining the relevance of salient regions of interest in an unsupervised way, but can be adapted to a variety of implementations.**

I. Introduction

Content-Based Image Retrieval (CBIR) systems traditionally consist of a query interface and a mechanism to process the query. The field of CBIR has seen diverse and notable contributions to both components. However, many CBIR implementation couple a sophisticated retrieval system with a basic interface. This demonstration presents a new CBIR prototype, the Perceptually-Relevant Image Search Machine (PRISM) that instead focuses on the benefits a thoughtful interface may provide.

We will demonstrate the Web-based interface of PRISM. PRISM was designed with contemporary technologies (DHTML, Ajax) and introduces new possibilities for image retrieval. It is flexible enough to be adapted to diverse CBIR and CFIR systems. Our demonstration uses a region-of-interest-based analysis to determine image relevance. Possibilities for semantic annotation within the same interface will also be introduced.

II. Background and Motivation

Our previous work implemented an attention-driven model for grouping similar images [1]. We postulated that users frequently search for salient object within images. The Itti-Koch model of visual attention [2] is combined with the Stentiford model [3] to extract regions of interest. One extension to this work generated clusters of related images in an unsupervised manner [4].

PRISM extends the previous implementation with a new, human-centric interface for image search. PRISM provides a query interface for image retrieval systems that allows the user to place and scale images freely. The work was motivated by our belief that an image search system, regardless of the sophistication of its algorithmic capabilities, is limited by the ability of the query interface.

Further motivation for creating an interface such as the one employed by PRISM was provided by recent advances in Content-Free Image Retrieval (CFIR). Liu and Chen demonstrate a method for retrieving images absent of any content-specific analysis [5]. The PRISM interface was designed with the ability to record user interactions in mind, eventually allowing CFIR ideas to be tested in the same environment as well. By collecting user data throughout the session and allowing query-by-relevant-content to be exercised during this time the interface provides means to combine conventional CBIR and relevance feedback functionality in parallel with CFIR functionality.

PRISM was also inspired by the interfaces developed by Jefferson Y. Han and demonstrated by him and Davis [6]. One demonstration of this technology shows a user moving, scaling, and rotating images with nothing more than their fingers directly affecting the images. While the standard mouse pointer is certainly far less elegant, we appreciate their minimalist and natural approach to interface design.

Finally, the ability to semantically annotate groups of images was an additional motivational factor in the design of PRISM. The system currently allows the user to annotate groups of images with a short amount of text identifying each workspace tab (a container for a group).

III. Functional Description

In this particular application of PRISM the user has the ability to position and resize related images, with the only instruction being that larger images are judged as more relevant by the system. This new method of relevance feedback is intuitive and does not require expert knowledge of the system's inner workings.

The main interface of PRISM is divided into two primary partitions: the "filmstrip", and the workspace, as shown in Figure 1. PRISM allows the user to maintain multiple independent workspaces at the same time. Tabs are used to switch between workspaces and can be renamed as the user desired. In the bottom-left corner additional controls provide the ability to refresh the filmstrip with a new set of random or relevant

Fig. 1. The PRISM interface

images. In the bottom-right corner a familiar trash can icon deletes images that are dragged onto it.

The design of the filmstrip is purposefully metaphorical. It is an always-replenished collection of images that can either be selected at random or relevant to the user's query. Images are simply dragged and dropped from the filmstrip to the workspace (to keep) or the trash can (to delete). Once an image is removed from the filmstrip it is immediately replaced.

The workspace occupies most of the interface. Here the user is free to arrange images in any way that is meaningful. When the mouse pointer is positioned over an image, controls appear that allow the user to enlarge or reduce the image's size. This ability, while simply explained, is key to the system's functionality. Larger images are considered more relevant and given more weight when computing relevant images. This form of relevance feedback is easily understood yet makes its function immediately clear.

The tabbed interface allows the user to organize independent groups of images. Tabs have become a popular interface construct, and for good reason. They dramatically expand and segment the functional area while occupying a minimal amount of space. Additionally, since the tabs can be named, the user can choose to add semantic information (up to several words) to annotate groups of images. In the future this information can be used to enhance search results with previously annotated images.

It is not insignificant that PRISM is a Web-based application. Ultimately, this provides the broadest platform for the product. Accessibility is of paramount important to any interface, and designing for widely-used Web browsers helps satisfy this need. The client-server architecture allows query processing to occur on the remote server, freeing the client of processing and storage requirements. The image database communicates independently with both the client and server, providing images as requested. A live demonstration of PRISM will be available online, accessible to anyone with a compatible Web browser.

The PRISM interface was implemented using contemporary technologies and design methodologies. On the client-side JavaScript, CSS, and the DOM (Document Object Model) realize the interface's functionality (collectively known as DHTML). Ajax (Asynchronous JavaScript and XML) is used throughout to communicate with the PHP-based server and image database.

The image database consists of 1471 images. The images constitute twelve categories of salient objects in a variety of indoor and outdoor environments.

IV. Conclusion

The demonstration introduces a live, Web-based application, known as PRISM. PRISM approaches the image retrieval task from a user-centric way, providing an interface that provides considerable freedom for the user to organize images as they see fit. We will demonstrate the effectiveness of this approach using an attention-based method of extracting and comparing salient regions-of-interest within images.

The future potential of this system is a key point of our presentation. We intend to show how the same interface, while demonstrated in a particular CBIR implementation, can be adapted to CFIR and semantic annotation tasks as well.

Acknowledgment

This research was partially sponsored by UOL (`www.uol.com.br`), through its *UOL Bolsa Pesquisa* program, process number 200503312101a.

References

[1] O. Marques, L. M. Mayron, G. B. Borba, and H. R. Gamba, "An attention-driven model for grouping similar images with image retrieval applications," *EURASIP Journal on Applied Signal Processing (accepted)*.

[2] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. on PAMI*, vol. 20, no. 11, pp. 1254–1259, Nov 1998. [Online]. Available: http://klab.caltech.edu/~itti/attention/publications/98_PAMI/

[3] F. Stentiford, "An attention based similarity measure with application to content based information retrieval," in *Proceedings of the Storage and Retrieval for Media Databases conference, SPIE Electronic Imaging*, Santa Clara, CA, 2003. [Online]. Available: citeseer.ist.psu.edu/stentiford03attention.html

[4] G. B. Borba, H. R. Gamba, O. Marques, and L. M. Mayron, "An unsupervised method for clustering images based on their salient regions of interest," in *The 2006 ACM Multimedia Conference*, Santa Barbara, CA, 2006.

[5] D. Liu and T. Chen, "Content-free image retrieval using bayesian product rule," in *IEEE International Conference on Multimedia & Expo*, 2006.

[6] P. Davidson and J. Han, "Synthesis and control on large scale multi-touch sensing displays." in *NIME*, 2006, pp. 216–219.