

USING A GAME TO EVALUATE IMAGE RETRIEVAL, ORGANIZATION, AND ANNOTATION

Liam M. Mayron and Oge Marques

Department of Computer Science and Engineering
Florida Atlantic University
Boca Raton, FL USA

ABSTRACT

Methods of retrieving images that incorporate human-generated metadata, such as keyword annotation and collaborative filtering, are less vulnerable to the semantic gap than content-based image retrieval. However, generating such metadata is time-consuming, expensive, and difficult to evaluate. This paper discusses an interface for image retrieval that is able to simultaneously compose queries based on content-based features, keywords, and collaborative relations. We propose using a game metaphor to evaluate the system.

Index Terms— content-based image retrieval, annotation, gameplay, multimedia information retrieval

1. INTRODUCTION

Content-based image retrieval (CBIR) systems are vulnerable to the *semantic gap* – the disparity between a user’s high-level interpretation of an image and the information that can be extracted from an image’s low-level physical properties – particularly due to their reliance on low-level features for describing image content.

The semantic gap may be narrowed by including high-level, user-generated information. Such high-level descriptions may be more capable of capturing the semantic meaning of image content, but it is not always practical to collect this information. Keyword annotation and collaborative filtering are two methods of collecting information from human users.

This paper presents the evaluation of an image retrieval system using a game metaphor. Because the user is an essential part of the performance of the system, the user must be included in the evaluation process. Thus, we present the use of a game to evaluate the system. We focus on the connection between the user and the system – the user interface. A summary of a new interface for image retrieval, PRISM, the Perceptually-Relevant Image Search Machine, is presented. PRISM provides the ability to query by image content, keywords, or previously-developed collaborative filters simultaneously through a single interface.

2. BACKGROUND

In a given situation, gameplay is the set of actions actions, bounded by certain rules, undertaken to reach an objective [3]. This pattern of interactions (perceptual, cognitive, and motor operations) known as the *gameplay gestalt* [3].

The context of the game is the key factor distinguishing gameplay from performing an otherwise tedious task. For example, it is more enjoyable to “complete the quest” using a series of keystrokes than to simply press those keys out of context. For similar reasons, games have been applied to education, as a way to motivate students [1].

The degree of interactivity is an important aspect of gameplay. “It is not enough to just sit and watch and possibly activate some cognitive schemas. Instead, the player must become an active participant. When successful, this type of participation leads to strong gameplay experiences that can have [a] particularly powerful hold on the player’s actions and attention” [2]. The PRISM Game enables interactivity in its design and persistently displays a timer, score meter, and progress meter throughout gameplay. Additionally, runs are ranked upon completion.

There must be a clear benefit to the user providing the keyword annotation and collaborative information. This benefit may belong to one of two classes:

- **Improved retrieval:** a system requires a good amount of human effort to annotate items, but does not provide feedback as to the significance of the person’s annotation or encouragement to continue. It is important to convey the contribution the human is making to the system without overwhelming a nonexpert user.
- **Entertainment:** in some cases the retrieval task overlaps with one that provides the user with a diversion. It may even be possible to completely separate the retrieval task from the collection of annotation. In this case, annotation can be presented to the user as a game. The only incentive to participate may be the entertainment value derived from playing the game. This has been successfully demonstrated in the ESP Game [7],

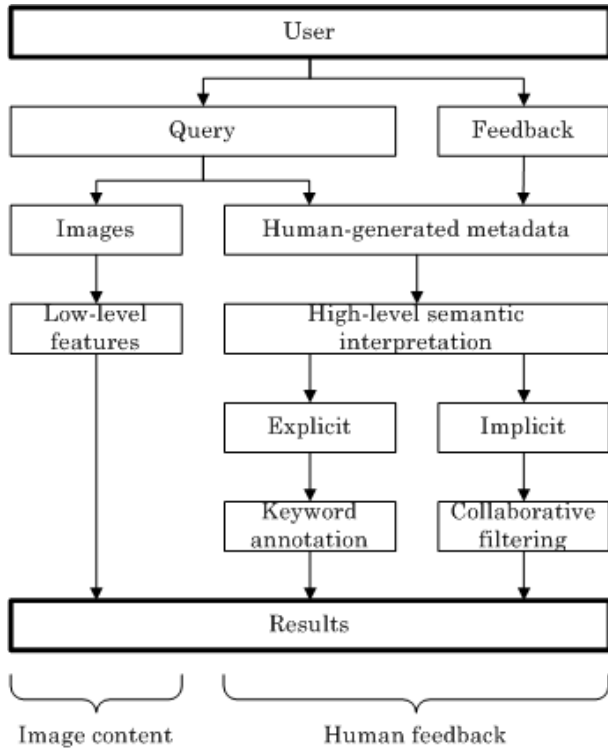


Fig. 1. High-level representation of the integration of retrieval by image content, keyword annotation, and collaborative filtering

which has proven to be an effective method of annotating large amounts of images.

3. ARCHITECTURE

The integration of content-based, keyword annotation, and collaborative filtering together in the image retrieval system described in this paper is illustrated in Figure 1. The user is the most critical component of a visual information retrieval system – it is the user who specifies the query and the user who is the ultimate judge of the results. The user composes their query and provides feedback through a user interface (consisting of the *Query* and *Feedback* blocks in Figure 1). The content-based track (the left side of Figure 1) relies solely on content-based data (i.e. metrics derived from low-level visual features). The other track is based on human-generated metadata. This may either be explicitly stated by the user (in the case of keyword annotation), or implicit (as with collaborative filtering). Searching based on keyword annotation and retrieving images using collaborative filters occurs entirely without knowledge of the image content itself, only relying on previous human actions. Collaborative filtering has previously been applied to image retrieval and is sometimes referred to as content-free image retrieval [4].

Table 1. Required interface features for each subsystem. *Q* indicates a requirement for querying whereas *L* indicates a requirement for learning

Method	Multiple Examples	Scale	Position	Annotate
Content-based	Q	Q		
Collaborative	Q, L		Q, L	
Keyword	Q	Q		Q, L

4. DESIGN

Table 1 summarizes the requirements for implementing a single interface that allows the three selected query methods to be simultaneously implemented. In Table 1 *Q* indicates a feature which is required for composing a query while *L* indicates a feature which is needed for learning (annotation or inferring relationships). In the context of this discussion *learning* is synonymous to modifying metadata in the database through user actions (i.e. annotating images and weighing collaborative filters).

This system’s design includes an interface that simultaneously enables content-based and content-free queries. The design of this interface allows the system to learn from user actions, improving the results without direct intervention. Through a single interface access is provided to three query subsystems. While each subsystem executes its query independently, results from all subsystem are displayed in the same interface once the queries are complete. The information used to compose queries – image annotation and spatial arrangement is also used for learning.

5. USER INTERFACE

The PRISM client’s user interface enables the user to construct expressive queries that meet the requirements of Section 4 through a set of intuitive, purpose-driven actions [5, 6].

The PRISM interface is shown in Figure 2. This view shows the initial screen the user sees before any actions are performed. It is separated into four functional areas:

- **Banner:** allows access to the least frequently used control and utility functions (e.g. help, sign out, change user settings).
- **Filmstrip:** a region near the top of the interface. It is the only source for new images. These images can either be random or related to the user’s query.
- **Console:** displays tabs, allows switching between tabs, and the annotation of tabs.



Fig. 2. The PRISM interface

- **Canvas:** an area, divided into multiple tabs (a tabbed interface design metaphor), where the user composes a query by arranging images, annotating images, and scaling images. Images are dragged to the canvas from the filmstrip and can be deleted by dragging them to a trash can area in the bottom-right corner.

In Figure 2 the user has created three tabs representing three broad categories (“transportation”, “landscapes”, and “animals”). The “transportation” tab is displayed. It has been populated by eight images featuring buses. Several images have been enlarged, with one being labeled “bus”. This will enable PRISM to search by the text “bus” next time the user requests a “Related images” query. Furthermore, the user has indicated a stronger relationship between three pictures of single-level buses (towards the right side in Figure 2) by overlapping the images. The information gathered by the interface will improve future queries by providing information that will improve the collaborative filters.

6. USING A GAME FOR EVALUATION

A game must establish clear objectives from two perspectives: that of the system designer, and of the user. There are the considerations from the system designer’s perspective:

- **Which features must be evaluated?** A wide range of variables can be evaluated. Interface features have a significant impact on the user experience. Overt changes to the interface can be compared between user groups. Alternatively, the behind-the-scenes retrieval methods can be evaluated.
- **How are outcomes evaluated?** The practitioner must decide the method used to evaluate results. This may be in the form of free responses in a user survey, a survey consisting of statements graded on a Likert scale, or a more subtle method, such as a timer or score. Covert

evaluation methods, such as counting the number of clicks or computing the quality of the user’s queries can be employed, or even a combination of evaluation methods can be used.

- **What is the size and composition of the user group?** The number of control and test users must be established. Users may be divided into *control* and *test* user groups. The practitioner must decide if a few users will be enough to evaluate the system, or if many are needed. The level of technical expertise required of the users must also be established.

The player has a different set of objectives:

- **What is the task I must perform?** The user’s task must be clearly and unambiguously defined. Furthermore, the complexity, challenge, and time demands must be communicated to the user.
- **How can I evaluate my performance?** The user must know if they are doing poorly or well. This can be communicated during gameplay or after gameplay. During gameplay, a running score, progress meter, or a timer can indicate success or failure. After gameplay a user’s performance can be compared to that of other players to produce a relative ranking.
- **How can I improve my performance?** Incentive to replay the game should be established by communicating ways in which they can improve their performance.
- **What incentive do I have to complete this task?** The game must be demonstrated to be either fun, scientifically meaningful, or personally rewarding.

7. THE PRISM GAME

The PRISM system combines CBIR, collaborative filtering, and keyword queries in one system. This hybrid approach created unique challenges in evaluating the quality of the results. The open-ended nature of the interface allows a variety of expressive of queries to be composed. Furthermore, the learning nature of the system means that the results may change over time.

A variation of PRISM was created in order to effectively evaluate the content-free components of the system. This variation presents PRISM not as an image retrieval system, but as a *game* where the objective is to obtain a high score. The game also serves to instruct users regarding the use of PRISM by rewarding more effective actions with more points.

The PRISM Game tracks and stores user actions for analysis. The frequency and nature of these actions indicate the effectiveness and intuitiveness of the interface was observed and evaluated. The PRISM Game allows user feedback, the evaluation of the usefulness of features of the user interface, and the comparison between baseline and enhanced features.

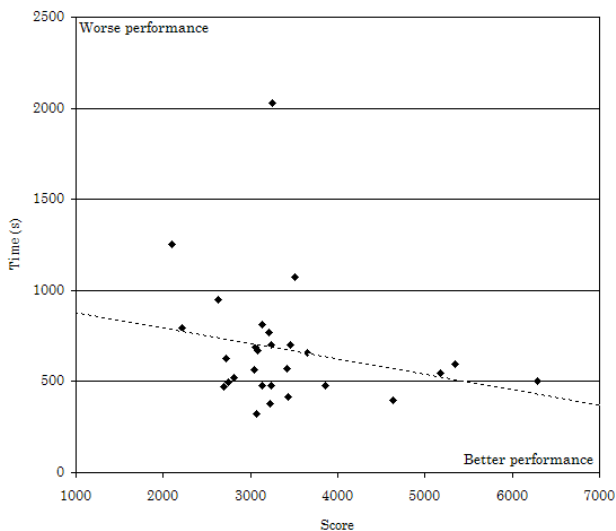


Fig. 3. Time vs. score for the PRISM Game

8. PRELIMINARY RESULTS

We present initial, preliminary results of a first trial of the PRISM Game. We share the highlights of this trial as there are currently no comparable results in the literature. We used this limited test to prepare for upcoming, more focused trials.

The PRISM Game (available online at <http://mlab.fau.edu/prism/site/>) was advertised for play between February 7 and February 24, 2008. During this period, 28 sessions of the PRISM Game were completed. Seventeen individuals also completed a user survey after their game concluded. 93% of the users (26 out of 28) users searched using the “related images” query. 39% of the users annotated at least one image. Images were annotated 234 times (8.75% of all images). Users were most satisfied with the ability to retrieve images using keywords. Searching for keywords is an overt method (the user explicitly specifies the keywords), whereas content-based retrieval and collaborative filtering are covert, behind-the-scenes methods. The abilities to create new tabs and to annotate tabs were highly rated – organization capabilities are important to users. Furthermore, being able to view a large version of the images in the system was also rated highly. This is a simple feature, but it eases viewing images in detail.

Figure 3 plots scores achieved in the PRISM Game versus the time required to complete the game for all users. The *score* metric indicates the efficiency with which the user organized and annotated images, whereas *time* is the amount of seconds taken to complete the task. Higher scores and lower times indicate better performance. The graph demonstrates the inverse relationship between score and time indicating that users who achieved higher scores were able

to more efficiently organize images using the features of the PRISM interface, although this is a weak correlation (indicated by the dashed line in Figure 3). However, higher scores do not necessarily correlate to higher satisfaction. Furthermore, the correlation coefficient between score and users responses to the question “what is your overall opinion of PRISM?” was only 0.0874 – highly ranked user performance does not necessarily indicate increased satisfaction.

Several improvements can be made to this initial trial. In the PRISM Game all users were rated against each other. Instead, the users should be divided into control and test groups. We learned that the task we assigned users – to organize all images as they see fit – was too broad, ambiguous, and complex. Instead, a smaller, more focus tasks needs to be given to users (e.g. asking users to find several images of a specific type of scene within a certain amount of time). We are currently implementing these improvements and evaluating these new game-based experiments.

9. REFERENCES

- [1] G. Denis and P. Jouvelot. Motivation-driven educational game design: applying best practices to music education. In *ACE '05: Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology*, pages 462–465, New York, NY, USA, 2005. ACM.
- [2] L. Ermi and F. Myr. Fundamental components of the gameplay experience: Analysing immersion. In *DIGRA Conf.*, 2005.
- [3] C. A. Lindley. The gameplay gestalt, narrative, and interactive storytelling. In *Computer Games and Digital Cultures Conference*, Tampere, Finland, June 2002.
- [4] D. Liu and T. Chen. Content-free image retrieval using bayesian product rule. In *IEEE International Conference on Multimedia & Expo*, 2006.
- [5] L. M. Mayron, G. B. Borba, V. Nedovic, O. Marques, and H. R. Gamba. A forward-looking user interface for CBIR and CFIR systems. In *IEEE International Symposium on Multimedia (ISM2006)*, San Diego, CA, USA, December 2006.
- [6] L. M. Mayron and O. Marques. Design of a web-based interface for image retrieval. In *International Conference on Web Information Systems and Technologies (WEBIST)*, Barcelona, Spain, March 2007.
- [7] L. von Ahn and L. Dabbish. Labeling images with a computer game. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 319–326, New York, NY, USA, 2004. ACM Press.