

FutureView: Enhancing Exploratory Image Search

Sayantana Hore, Dorota Głowacka, Ilkka Kosunen, Kumaripaba Athukorala and Giulio Jacucci
Helsinki Institute for Information Technology HIIT, Department of Computer Science, University of Helsinki
first.last@cs.helsinki.fi

ABSTRACT

Search algorithms in image retrieval tend to focus on giving the user more and more similar images based on queries that the user has to explicitly formulate. Implicitly, such systems limit the users exploration of the image space and thus remove the potential for serendipity. As a response, in recent years there has been an increased interest in developing content based image retrieval systems that allow the user to explore the image space without the need to type specific search queries. However, most of the research focuses on designing new algorithms and techniques, while little research has been done in designing interfaces allowing the user to actively engage in directing their image search. We present an interactive FutureView interface that can be easily combined with most existing exploratory image search engines. The interface gives the user a view of possible future search iterations. A task-based user study demonstrates that our interface enhances exploratory image search by providing access to more images without increasing the time required to find a specific image.

Categories and Subject Descriptors

H.4 [Information Systems Applications]: Miscellaneous

Keywords

Interactive user interfaces, Content Based Image Retrieval (CBIR), Exploratory search

1. INTRODUCTION

In recent years, image retrieval techniques operating on meta-data, such as textual annotations or tags, have become the industry standard for retrieval from large image collections, e.g. Google Image Search. This approach works well with sufficiently high-quality meta-data, however, with the explosive growth of image collections, it has become apparent that tagging new images quickly and efficiently is not always possible. Secondly, even if instantaneous high-quality image tagging was possible, there are still many instances where image search by query is problematic. It might be easy for a user to define their query if they are looking for an image

of a cat but how do they specify that the cat should be of a very particular shade of ginger with sad looking eyes.

A solution to this problem has been content based image retrieval (CBIR) [5, 12] combined with relevance feedback [24]. However, evidence from user studies indicates that relevance feedback can lead to a context trap, where the user has specified their context so strictly that the system is unable to propose anything new, while the user is trapped within the present set of results and can only exploit a limited area of information space [11]. Faceted search [22] was an attempt to solve the problem of context trap by using global features. However, the number of global features can be very large thus forcing the user to select from a large amount of options, which can make the whole process inconvenient and cognitively demanding. Employing various *exploration/exploitation* strategies into relevance feedback has been another attempt at avoiding the context trap. The exploitation step aims at returning to the user the maximum number of relevant images in a local region of the feature space, while the exploration step aims at driving the search towards different areas of the feature space in order to discover not only relevant images but also informative ones. This type of systems control dynamically, at each iteration, the selection of displayed images [18, 7].

However, in spite of the development of new techniques to support queryless exploratory image search, not much attention has been devoted to the development of interfaces to support this type of search [19]. Most research in CBIR interface design concentrates either on faceted search [20, 22] or enabling CBIR through a query image or a group of images [15]. In fact, most of the existing techniques and interfaces rely for exploration on iterative trial-and-error. All of the above techniques provide only limited support for the recent emerging trend of combining interactive search and recommendation [2]. One key question in this respect is how to utilise relevance feedback in optimising not only the narrowing but also the broadening of the scope of the search. We contribute to this problem with FutureView – an interface that supports queryless CBIR image search through more fluid steering of the exploration. This system uses a novel technique that allows users to preemptively explore the impact of the relevance feedback before operating a query iteration. We investigate in an evaluation whether this approach is useful in allowing users to explore more pictures. The evaluation of FutureView is carried out in a comparative user study and we conclude with implications for future development of image search systems that blur interactive search and recommendation.

2. RELATED WORK

Most image search systems still rely on search queries in order to return to the user a set of images associated with a tag related to the search query [1, 19]. There are also a number of alternative in-

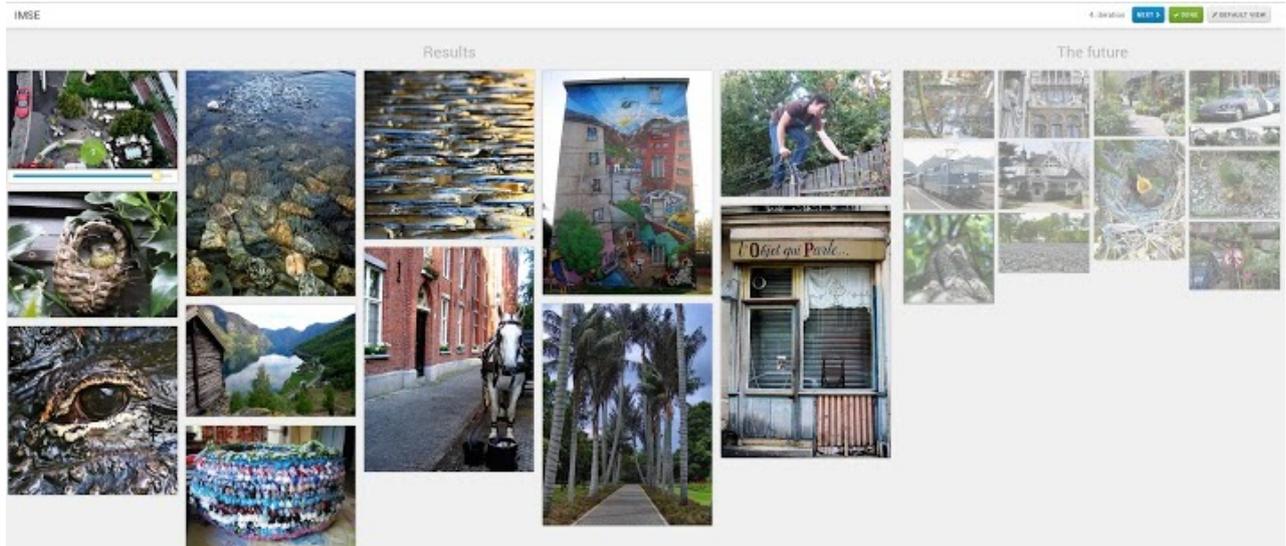


Figure 1: The FutureView interface: users can rate images on the panel on the left-hand side of the screen and the future view of the next iteration is presented on the right-hand side of the screen.

terfaces that group similar images based on various clustering techniques [21], or display similar images close to one another [14, 17, 16, 23]. However, all of these techniques rely on the availability of a dataset of tagged images or an automatic expansion of an initial textual query. Another approach is to rank images based on features extracted from a set of query images provided by the user [4, 6]. Faceted search [22] is another technique applied in CBIR to allow the user to browse through a collection of images using high-level image features, such as colour or texture. However, this approach often leads to a very large number of features, which can make the search process cognitively demanding.

3. OUR APPROACH

The main idea behind interactive interfaces used in most query-less exploratory CBIR systems [3, 13, 18] is that instead of typing queries related to the desired image, the user is presented with a set of images and navigates through the contents by indicating how “close” or “similar” the displayed images are to their ideal image. Typically, the user feedback is given by clicking relevant images or through a sliding bar at the bottom of the image. At the next iteration, the user is presented with a new set of images more relevant to his interest. The search continues until the user is satisfied with the results. Previous studies of CBIR systems show that this type of interface is intuitive and easy to use [3], however, users often feel that the new set of images does not reflect the relevance feedback they provided earlier: users do not feel fully in control of the system.

Our solution to this problem is an interface that provides the user with a “peek into the future”. The FutureView interface, illustrated in Figure 1, is divided into two sections. The left-hand part of the screen is similar to a traditional interface, where the user can rate images by using a sliding bar at the bottom of each image. However, after rating one or more images, the user is not taken to the next search iteration but instead presented with the future view of the next iteration on the right-hand side of the screen. This allows the user to “try out” what impact providing feedback to different images will have on future iterations. When the user is satisfied with one of the future views, he clicks the “next” button in the right

upper corner of the screen to confirm his choice and then is taken to the next search iteration.

4. EXPERIMENTAL STUDY

We conducted a comparative user study to evaluate the impact of FutureView on three types of image search tasks: target, category and open. The study included two conditions: 1) our FutureView interface; 2) a version of our interface without the future view, which from now on we will refer to as “single view”. The same backend system was used with both user interfaces. We used as our backend an existing exploratory image search system, the details of which can be found in [9]. We also recorded the gaze behavior of the participants to determine how much time they spent observing the future during the FutureView condition. Gaze data was recorded during both conditions, and the participants were not informed that only the data in the FutureView condition would be used. We used the Tobii X2-60 eye tracker with sampling rate of 60Hz.

4.1 Participants

We recruited 12 post-graduate students from our university to participate in the study (3 female). The average age of the participants was 24 years (from 20 to 30). Google image search is the most frequently used images search tool by all the participants.

4.2 Design

We used the MIRFLICKR-25000 dataset with three types of features: colour, texture and edge, as described in [10]. We followed the most commonly used categorization of image search to design our tasks[3]:

- *Target search* - the user is looking for a particular image, e.g. a white cat with long hair sitting on a red chair.
- *Category search* - the user does not have a specific image in mind and will be satisfied with any image from a given category, e.g. an image of a cat.
- *Open search* - the user is browsing a collection of images

without knowing what the final target may look like, e.g. looking for an illustration to an essay about “youth”.

We used a within subject design so that every participant performed three tasks covering all task types in both systems (six tasks in total = 3 (task types) × 2 (systems)). We designed two tasks for each category to assign unique task for each system. The subject of the two tasks for target search are: red rose, and tall building. In category search, we asked the participants to find images from the following categories: city by night, seashore. In open search, we asked the participants to imagine they were writing a newspaper article on a given topic and they had to find an image to accompany their article. The topics of the articles were: (1) happiness; (2) gardening. We selected these topics because they are well covered in the MIRFLICKR-2500 dataset. We showed 12 images per iteration in the single view interface and in Futureview. After receiving feedback, FutureView shows the next 12 images on the right-hand side.

4.3 Procedure

At the beginning of the experiment, we briefed the participants as to the procedure and purpose of the experiment before they signed the informed consent form. We then provided them with practice tasks to get them familiar with both systems. The participant would then proceed to perform six search tasks, divided into two groups of three tasks so that each participant would complete each different type of search task once with both systems. Before they started the target search tasks, we presented three example images and a short description of the image that they should look for. We did not provide any example images for category search and open search tasks. We randomized the order of tasks as well as the order of systems. After training, the eye tracker was calibrated.

We instructed the participants to finish each task when they find the target image (in case of target search) or when they feel they found the ideal image for the tasks from category search and open search. In all the tasks, we limited the search to 25 iterations to ensure that the participants did not spend an excessive amount of time on any task. After finishing each task, the participants completed the NASA TLX questionnaire [8]. After the completion all 6 tasks, we conducted a semi-structured interview with every participant to understand their overall satisfaction with the FutureView. A study lasted approximately 45 minutes. We compensated the participants with a movie ticket.

5. FINDINGS

Overall 12 users completed 72 tasks and all the participants completed all the tasks in fewer than 25 iterations. Figure 2 shows the average duration of a search session and the average number of images shown over a search session. On average, category searches were the shortest (104 seconds with single view and 109 seconds with FutureView), while open searches took the longest (145 seconds with single view and 140 seconds with FutureView). The Wilcoxon signed rank test indicates no significant difference in search session duration for any search type with the two interfaces ($p > 0.6$). In spite of the fact that no additional time is required to complete each type of search with FutureView, users are exposed to a much higher number of images – on average three times more than with single view. The Wilcoxon signed rank test shows that this number is significantly higher in open and target searches ($p < 0.05$) and marginally higher ($p = 0.05$) in category search with FutureView. These results indicate that FutureView supports more exploration.

Figure 3 shows the average scores of the NASA TLX question-

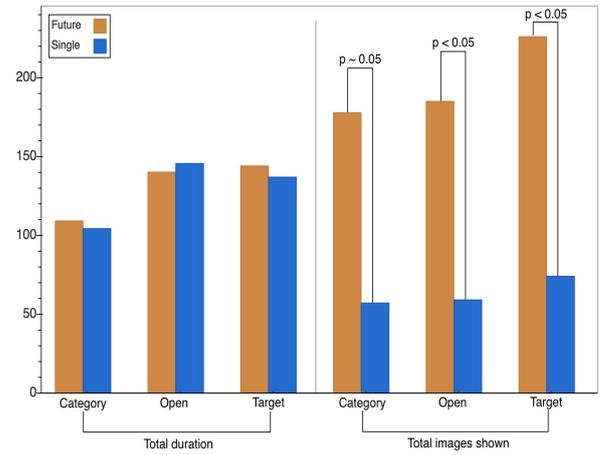


Figure 2: Average duration of a search session (in seconds) and average number of images shown over a search session for the three type of searches with a single view interface and FutureView

naire. In spite of the fact that with FutureView users were exposed to three times as many images as with the single view interface within the same period of time, users did not report feeling hurried, stressed or irritated. Similarly, users did not feel that FutureView made the task more mentally or physically demanding and they did not feel that they had to work any harder to achieve their goal. The Wilcoxon signed rank test indicates that there was not significant difference between the two interfaces in terms of scores for questions 1,2, 4, 5 and 6 ($p > 0.2$). The users, however, felt significantly more successful completing the task with FutureView ($p < 0.04$ according to Wilcoxon signed rank test).

The eye tracking results show that the participants spent similar amount of time looking at both the current search results and the future view. Out of the 12 participants, three had excessive amount of errors in the eye tracking data, so only nine participants were considered. On average, the users spent 41.8% of the time looking at the future section of the screen, with standard deviation of 11.8%.

The post-experiment interviews with the participants also indicate that they found the FutureView interface helpful and easy to use. Some of the comments include: “The FutureView is pleasant to use and play with”; “The FutureView helps in reaching target quicker than the single view”; “The FutureView is helpful for people whose job is to search for images”. These comments are in striking contrast to the remarks the participants made in the pre-study questionnaire, where they stated that most existing image search engines are tiring and cumbersome to use. The participants also remarked that “Single View can be discouraging as the user has no idea what is coming next”, “once deviated from the actual path, there is no way to come back [in single view]”.

6. CONCLUSIONS

In this paper, we introduced the FutureView interface for query-less exploratory content based image search. It allows the user to see the effect of the relevance feedback on currently presented images on future iterations, which, in turn, allows the user to direct their search more effectively. Initial experiments show that users take advantage of the FutureView interface and engage in more exploration than in a system with a single view interface.

Our future plans include more extensive user studies with various types of image datasets and various image feature representations.



Figure 3: Average score for the NASA TLX questionnaire for tasks conducted with a single view interface and the FutureView.

Currently, the FutureView does not save the search history. We are planning to add this feature to our system to allow the user to branch out their searches using any point in the history as a new starting search point.

7. ACKNOWLEDGEMENTS

This work was supported by The Finnish Funding Agency for Innovation (projects Re:Know and D2I) and the Academy of Finland (the Finnish Centre of Excellence in Computational Inference).

8. REFERENCES

- [1] P. André, E. Cutrell, D. S. Tan, and G. Smith. Designing novel image search interfaces by understanding unique characteristics and usage. In *Proc. of INTERACT*, 2009.
- [2] E. H. Chi. Blurring of the boundary between interactive search and recommendation. In *Proc. of IUI*, 2015.
- [3] I. Cox, M. Miller, T. Minka, T. Paphomas, and P. Yianilos. The bayesian image retrieval system, pichunter: theory, implementation, and psychophysical experiments. *Image Processing*, 9(1):20–37, 2000.
- [4] J. Cui, F. Wen, and X. Tang. Real time google and live image search re-ranking. In *Proc. of MM*, 2008.
- [5] R. Datta, J. Li, and J. Wang. Content-based image retrieval: approaches and trends of the new age. In *Multimedia information retrieval*, pages 253–262. ACM, 2005.
- [6] J. Fogarty, D. Tan, A. Kapoor, and S. Winder. Cueflik: Interactive concept learning in image search. In *Proc. of CHI*, 2008.
- [7] D. Głowacka and J. Shawe-Taylor. Content-based image retrieval with multinomial relevance feedback. In *Proc. of ACML*, 2010.
- [8] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Advances in psychology*, 52:139–183, 1988.
- [9] S. Hore, L. Tervainen, J. Pyykko, and D. Glowacka. A reinforcement learning approach to query-less image retrieval. In *Proc. of Symbiotic*, 2014.
- [10] M. J. Huiskes, B. Thomee, and M. S. Lew. New trends and ideas in visual concept detection: The mir flickr retrieval evaluation initiative. In *Proc. of MIR*, 2010.
- [11] D. Kelly and X. Fu. Elicitation of term relevance feedback: an investigation of term source and context. In *Proc. of SIGIR*, 2006.
- [12] H. Kosch and P. Maier. Content-based image retrieval systems-reviewing and benchmarking. *JDIM*, 8(1):54–64, 2010.
- [13] J. Laaksonen, M. Koskela, S. Laakso, and E. Oja. Pictom—content-based image retrieval with self-organizing maps. *Pattern Recognition Letters*, 21(13):1199–1207, 2000.
- [14] H. Liu, X. Xie, X. Tang, Z.-W. Li, and W.-Y. Ma. Effective browsing of web image search results. In *Proc. of MIR*, 2004.
- [15] M. Nakazato, L. Manola, and T. S. Huang. Group-based interface for content-based image retrieval. In *Proc. of the Working Conference on Advanced Visual Interfaces*, 2002.
- [16] N. Quadrianto, K. Kersting, T. Tuytelaars, and W. L. Buntine. Beyond 2d-grids: A dependence maximization view on image browsing. In *Proc. of MIR*, 2010.
- [17] G. Strong, E. Hoque, M. Gong, and O. Hoerber. Organizing and browsing image search results based on conceptual and visual similarities. In *Advances in Visual Computing*, pages 481–490. Springer, 2010.
- [18] N. Suditu and F. Fleuret. Iterative relevance feedback with adaptive exploration/exploitation trade-off. In *Proc. of CIKM*, 2012.
- [19] B. Thomee and M. S. Lew. Interactive search in image retrieval: a survey. *International Journal of Multimedia Information Retrieval*, 1(2):71–86, 2012.
- [20] R. Villa, N. Gildea, and J. M. Jose. A faceted interface for multimedia search. In *Proc. of SIGIR*, 2008.
- [21] S. Wang, F. Jing, J. He, Q. Du, and L. Zhang. Igroup: Presenting web image search results in semantic clusters. In *Proc. of CHI*, 2007.
- [22] K.-P. Yee, K. Swearingen, K. Li, and M. Hearst. Faceted metadata for image search and browsing. In *Proc. of CHI*, 2003.
- [23] E. Zavesky, S.-F. Chang, and C.-C. Yang. Visual islands: Intuitive browsing of visual search results. In *Proc. of CIVR*, 2008.
- [24] X. Zhou and T. Huang. Relevance feedback in image retrieval: A comprehensive review. *Multimedia systems*, 8(6):536–544, 2003.