

# A User Study on User Attention for an Interactive Content-based Image Search System

Mahmoud Artemi<sup>a</sup> and Haiming Liu<sup>b</sup>

<sup>a</sup> University of Bedfordshire, University Square, Luton, LU1 3JU, UK

<sup>b</sup> University of Bedfordshire, University Square, Luton, LU1 3JU, UK

## Abstract

User attention is one of the fundamental indications of users' interests in search. For content-based image search systems, it is important to understand what users pay attention to, and thus engage users more in the search process. It remains a big challenge to design a user-centered interactive interface that serves well for both user interaction and search model, and be able to bridge the unsolved problem in content-based image search called the Semantic Gap. In an effort to solve the problem, we designed an interactive content-based image search interface called Search Strategy (SS) based on Vakkari's model. SS enables users to engage in three stages (pre-focus, focus-formulation, and post-focus) during the search process. We carried out a user study to observe which interface attracts more user attention. The user study is conducted in a lab-based setting using a screen-based eye tracker (Tobii Pro Nano) and Galvanic Skin Response (GSR) on the iMotions platform. The preliminary results show that participant attention is noticeably higher on the SS interface. This finding highlights the need for a well-designed interface that enables user interaction at all stages of image search process, and at the same time the interface should allow users to manipulate the search model effectively.

## Keywords

Content-based image retrieval, user interface, active learning, Vakkari model, eye tracking, query formulation.

## 1. Introduction

Most image search systems rely on text-based retrieval. It is often challenging for users to describe their search intents by describing images using keywords; these may lead to unsatisfactory retrieval results containing images irrelevant to the users' search intents [1]. To preserve the users' intent visually and improve the search performance, content-based image retrieval (CBIR) has emerged [2-6]. Since CBIR search uses the representation of visual features (such as color, shape, and texture), it is built for users to express more precisely their intents. [6]. Although CBIR helps cope with the ambiguity in text-based image search systems, it presents a new challenge called the Semantic Gap, a gap between low-level visual features that a computer understands and high-level semantics that users understand.

CBIR mainly works on the representation of visual image features to identify similarity of the images to the users' visual queries. Sustained attention has been made in order to cope with two essential challenges in CBIR system, i.e., intention and semantic gaps. Figure 1 illustrates the intention gap lies between user search intent and desired query [6, 7], whilst the semantic gap refers to the difficulty of mapping high-level concept to low-level image features [4].

*BIRDS 2021: Bridging the Gap between Information Science, Information Retrieval and Data Science, March 19, 2021. Online Event.*

EMAIL: Mahmoud.Artemi@study.beds.ac.uk (M. Artemi); haiming.liu@beds.ac.uk (H. Liu)

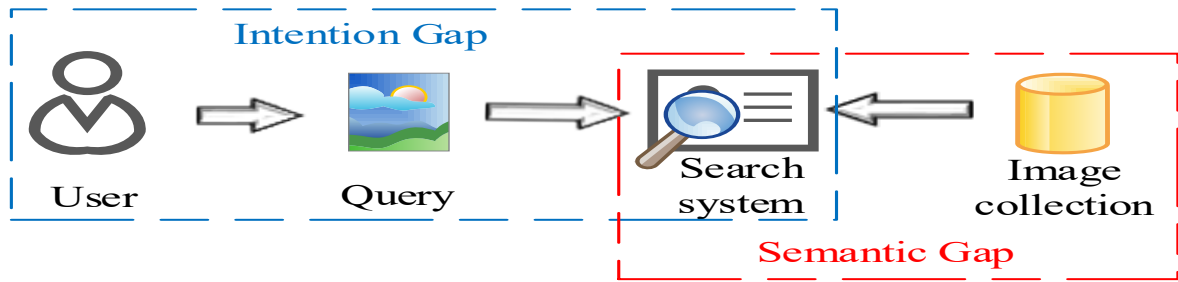
ORCID: 0000-0002-5177-8977 (M. Artemi); 0000-0002-0390-3657 (H. Liu)



© 2021 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

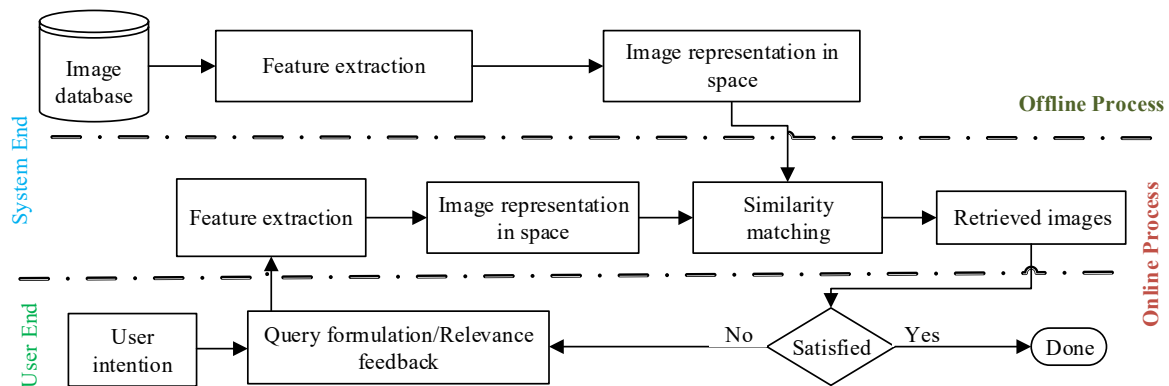
CEUR Workshop Proceedings (CEUR-WS.org)



**Figure 1:** Involvement of intention gap and semantic gap in content-based image retrieval (CBIR). Figure adapted from [7]

The basic framework of a CBIR search system is shown in Figure 2, which comprises of four main components, query formulation / relevance feedback, feature extraction, similarity matching, and results presentation.

- Query Formulation: from the user perspective, the user can use various query formulation schemas to express their intention.
- Feature Extraction: also known as content representation, an image is constructed by an array of pixel distributions containing low visual features such as shape, color, and texture.
- Retrieval Model: also known as similarity matching, the CBIR search model returns a set of ranking images by applying similarity metrics between image query and database images.
- Relevance Feedback (RF): due to lack of sufficient semantics in a given query, the RF provides a mechanism to formulate and modify a given query, aiming to capture user intents more precisely.



**Figure 2:** CBIR system flowchart

Relevance feedback has been an effective way to bring users into the CBIR search loop, which allows them to provide feedback to obtain improved results. Most of the research on relevance feedback focuses on enabling users to provide feedback at the result assessment stage [1, 3]. However, often the underline machine learning mechanisms in many CBIR systems need user feedback at the query formulation stage for better training and search performance [6]. There is a need to design an interactive CBIR search system that not only allows users to interact with the retrieved image results but also allows users to visually explore the image collection and facilitates users to train the underlying search model through a user-centered interactive search interface, therefore improving search performance and users' search experiences and satisfaction [1, 8, 9].

In this paper, we present our CBIR system, developed based on the concept of Vakkari's three-stage model in [6]. We also report a user study that we carried out on our interactive CBIR system design. The user study investigates the advantages of the proposed interfaces detailed in Section 3. The preliminary results enable us to better understand the users' information needs and the influence of user attention.

## 2. Related work

The work presented in this paper is shaped by prior studies in the area of interactive information retrieval, especially relevant feedback approaches through user interface design, and task-based information retrieval for a better search experience.

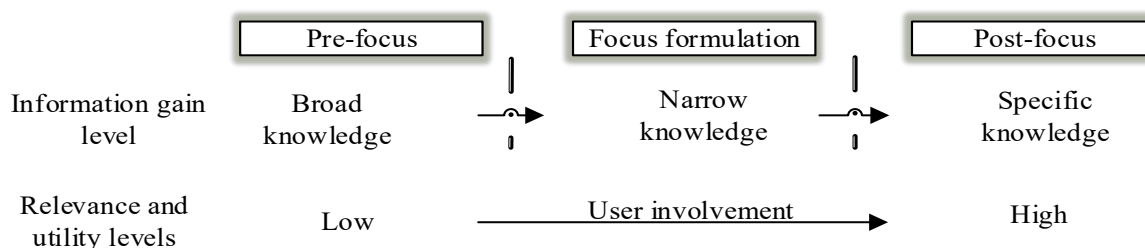
### 2.1. Active learning paradigm

Active learning is a machine learning mechanism that requests labels of data instances to train a model. Various active learning algorithms have been introduced for different applications [10-12]. Active learning is a semi-supervised learning where the learner model has an active role in defining the current data points to be labelled by an oracle (e.g., a user) [11]. The process of active learning starts with proposing the images to be labelled, and then those new-obtained labels are added to the training set to train the learning model. That is, active learning iteratively requests the user to label items (such as document or images) to obtain new data points. This process aims to get the desired results. In the active learning paradigm, the training data are not selected beforehand, unlike in other machine learning problems, which require treating the training data as fixed, selected data. During the training process, the active learner has the role of choosing the data to be acquired for the training purpose. In the training loop, the user typically takes actions that enable them to gain more information [13]. One of the major obstacles in CBIR is the intention gap. In order to bridge the intention gap, Relevance Feedback (RF) is used to capture semantic information from user intention and thus improve the search system. Different RF mechanisms have been introduced in CBIR to enable users to steer the CBIR system during retrieval process [4, 14-16]. This enables users to interact with the image search results. Therefore, RF is used by the users to mark the returned results as being relevant or irrelevant for a given query in an iteration schema. Then, the search model performs another search iteration to improve the system performance and return a more relevant set of images; iterations are continued until the user is satisfied with the search results. Although RF is introduced in CBIR, the search results can still be unsatisfactory [17]; the amount of data provided as feedback may be too small or unreliable to improve system performance as the system already knows about those selected images, which in fact sometimes confuse the system [11].

Due to limitations of conventional RF approaches, we propose applying active learning to the query learning stage (query formulation /focus formulation). Although an active learning procedure still requires users to judge the relevance level of the found items compared to their relevance level to the query, there is a significant difference between conventional RF approaches and the proposed active learning approach in terms of which images need to be labelled [17]. In the RF scenario, the user typically labels the top ranked items of retrieval results to improve the system performance of the next interaction. In this paper, this RF scenario is applied as the baseline system named Information Goal (IG). In the proposed active learning scenario, the learner model actively requests the users to label uncertain items to the search system in order to improve the system accuracy. In this paper, the proposed active learning mechanism is called Search Strategy (SS). It has been reported that the learning rate in an active learning mechanism is faster than in relevance feedback, and thus active learning achieves better accuracy than RF [6, 11]. In this paper, the pool-based active learning is used together with a support active machine in our CBIR system to enable the user's engagement in the query formulation stage (query learning). Our approach employs a Support Vector Machine (SVM) [18], which uses a kernel model for classification [16]. Once again, the users' needs can be defined as the users' intents. In order to capture the users' needs and attention, the users are asked to provide feedback as their preferences to given images. Those images are then used to train the learner model during the query formulation stage. This is different from asking the user to give feedback based on the result images that are already recognized by the search model (learner).

## 2.2. Vakkari’s three-stage model

In this paper, our focal point is the task-based model introduced by Vakkari [19]. It consists of a three-stage information seeking process: *pre-focus* includes three actions performed by users – they may initiate the search by selecting a query image before or after exploring the image collection; *focus-formulation* is where users may refine or change the search activity; and *post-focus* at the end of the search process, where a user can collect and save results of value to their needs. According to Vakkari’s model, the exploratory search process begins with *pre-focus* as the user typically starts with broad knowledge of a topic-based task, and then *focus-formulation* to narrow query formulation [20]. Decision-making may occur during the search process and continue to be presented clearly at the assessment stage (*post-focus*). Users assess a set of returned images to find not only a set of relevant images but also the best images that fit a given task and are useful (utility) for the user’s needs (intents). The effects of applying Vakkari’s three search stages are described in Figure 3.



**Figure 3:** Effects of applying Vakkari’s three search stages

Vakkari’s model is applied in our search system design to support evaluation of user interaction in task performance. Artemi and Liu [6] concluded that a better search system design is needed in order to capture user intention in the early stage of task search process. More clearly, it is insufficient to grasp the user’s search intention by asking the user to provide feedback on returned images that the search model already knows about those images. Furthermore, in most existing image query schemas, the end-user’s visual query formulation handles the form of a single image, which might be insufficient to indicate the user’s search intention, in some cases.

## 2.3. User search interface

Search interfaces play a vital role of intermediary interaction between search systems and end users. In the context of information retrieval, various approaches have been presented to design effective interfaces which fit user needs and more importantly, improve user interactions. Recent studies have concluded that different aspects should be considered for user interface design, such as cognitive aspects and task complexity, which might impede information seeking [21]. As in Section 2.1, it is observed that Vakkari [19] studied the nature process of information seeking, but gave no guidelines to design and implement search systems’ user interface aspects. This issue has also been presented in [22]. There are few studies that have studied the role of low-level user interface functionalities at different stages of the information seeking process.

Hurdeman and Kamps [22] designed a multistage information search system to support the information seeking process; the system was built upon the concept of task-based information seeking theory. Artemi and Liu [6] proposed a three-stage interface based on Vakkari’s model for content-based image retrieval to capture user’ intents during the focus formulation stage. White et al. [23] investigated the usability of implicit and explicit relevance feedback; their findings were that implicit feedback was used more in the early search stage while explicit feedback was used at the end of search process. Niu and Kelly [24] found that query suggestions were used for complex and difficult search tasks in the final stage of the search process. Kules et al. [25] conducted an eye tracking study in which exploratory search tasks were performed on a faceted search interface; the findings showed the user attentions

started at facets, then on query and later moved to results. Huurdeman et al [26] proposed a multistage simulated task approach, where three distinct tasks were performed in a way representing Vakkari's three-stage model. In this paper, a three-stage user interface [6] is used with eye tracking to look further into the impact user engagement evolves on user attention.

### 3. Three-stage-based search interface design

Here, we consider the workflow of the Search Strategy interface (SS) along with baseline Information Goal (IG) search interface (Figure 4-b) as presented in [6]. The SS interface enables the CBIR system, built based on the active learning paradigm, to capture the users' preferences during the query formulation stage, where the users can provide additional image examples within the training stage. The SS interface has three panels (Figure 4-a): the upper left panel is for exploring and selecting  $N$  random images. The upper right panel is the feedback window, where a user marks images in the pool query set as being relevant or irrelevant for selected iterations. In the bottom panel, the CBIR system returns a diversity of resultant sets considered matching the concept learned, where the user assesses the retrieved image set as being relevant and useful. The Explicit Searcher Model (ESM) from [6] represents the sequence of interactions between a searcher and the CBIR system over the course of a search session.



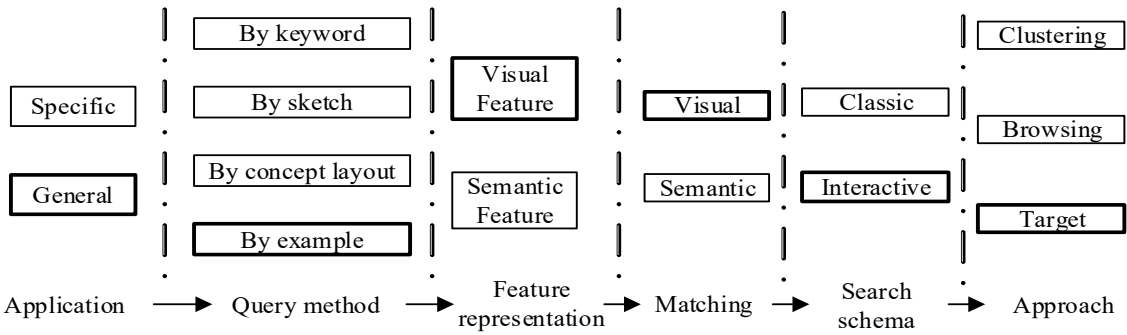
(a) Search Strategy (SS) interface [6]

(b) Information Goal (IG) interface [27]

**Figure 4:** User interfaces used in this study

Using an eye tracker to record user interactions enabled us to investigate the effectiveness of user engagement/attention in the focus stage and the exploratory search process. The SS system used the active learning mechanism where data is abundant [28]. It enabled the users to provide feedback as an intent or preferences. This method is successful in accelerating learning [29].

The feature extractor parameters were applied in these experiments as presented in [3]. The experiments were conducted using two interfaces, SS and IG interfaces (Figure 4). The relevance feedback mechanism was applied to the IG system [9]. Figure 5 shows the study boundary settings of the system proposed in this paper. The query type used was query by example to find a target image through an interactive paradigm; the image-visual features were applied for the image matching process.



**Figure 5:** The boundary settings in our study

## 4. Evaluation

A controlled lab-based user study was conducted on the SS and IG systems using eye tracker and galvanic skin response (GSR) devices with respect to Vakkari’s three-stage model of the information seeking process. This offers additional insight into how the participant’s engagement in query formulation influences user attention at the result assessment stage.

### 4.1. Experimental setup

The experiments were conducted in a UX lab. The aim of the experiments was to find out at which stage the user’s attention was high and why so, through collecting and analyzing eye gaze activity and galvanic skin response (GSR) to capture emotional arousal.

The eye tracker used in this setting was Tobii Pro Nano, selected to capture the activity of eye gaze fixation with sample rate 60Hz, and detecting visual attention. The GSR was used to record the level of emotional response that users experienced with the system. The model of used devices with output metrics is illustrated in Table 1.

**Table 1**  
Devices used in this study with output metrics

| Device      | Model                | Tool                                  | Output metric            | Output metric type |
|-------------|----------------------|---------------------------------------|--------------------------|--------------------|
| Eye tracker | Tobii Pro Nano       | Heatmaps, and areas of interest (AOI) | Fixation, and time spent | Attention          |
| GSR         | Wireless GSR Shimmer | Automated peak detection              | Peak detection           | Emotional arousal  |

Figure 6 shows the experimental setup where eye tracker and GSR devices are connected. Both devices were synchronized with the iMotion platform. iMotion creates users’ behavior data from the biosensor recordings.

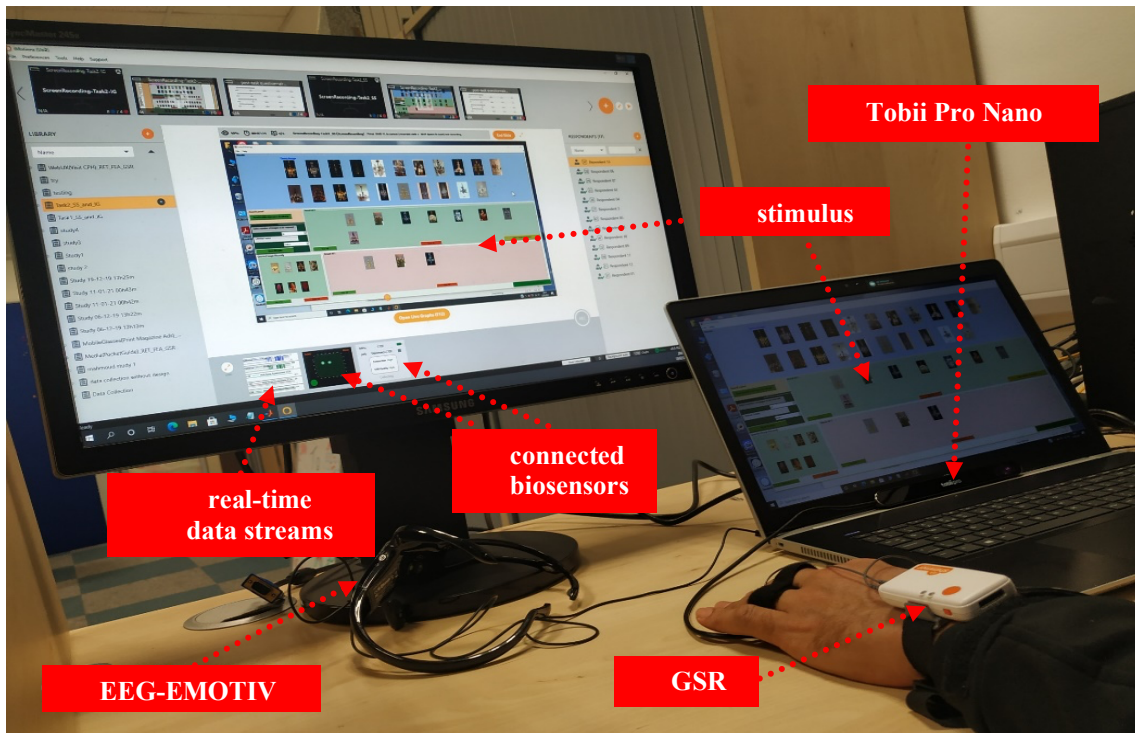


Figure 6: The experimental setup

## 4.2. Experimental design

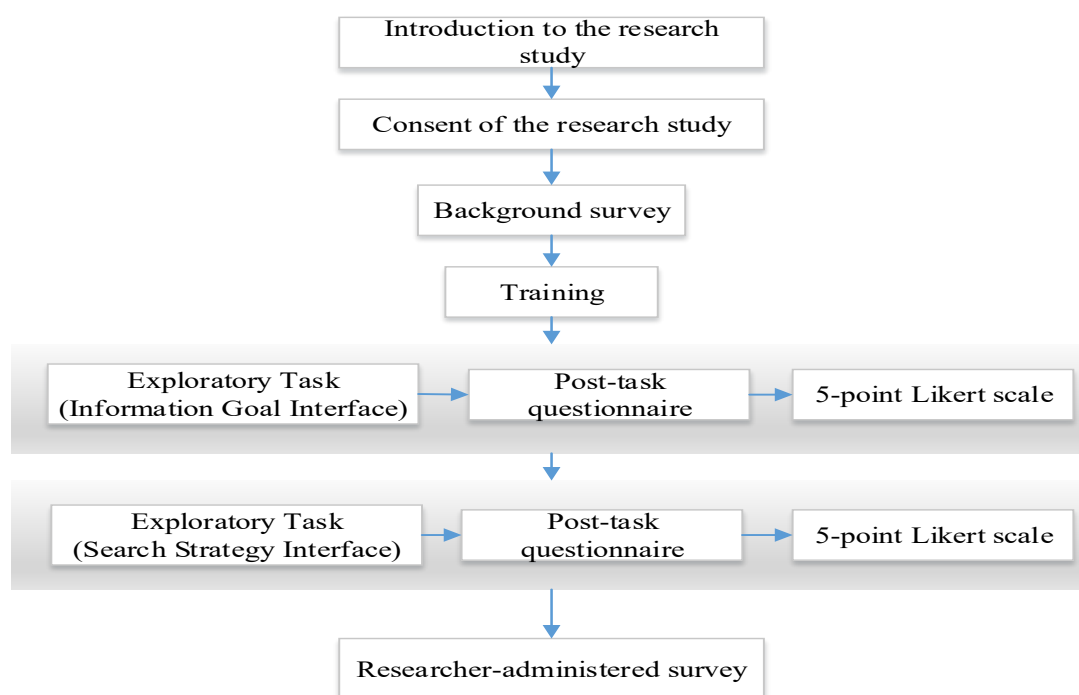
To investigate effectively the effects of the users' interaction with the interface based on three stages as users paid a certain attention when examine the image results. Therefore, this amount of user attention helps us to differentiate in which interface or system the results potentially meet user needs. To obtain evidence from the post-focus stage of what influences user attention behavior when they contribute in all search aspects, visualizing user's gaze path is needed such as heat maps and fixation patterns. We designed a controlled user study to obtain eye tracking data along with explicit feedback on search satisfaction from participants. Each of our participants performed two exploratory-image search tasks on each search interface. The GSR and eye tracker recorded the user activities. Table 2 shows the two image search tasks which participants were asked to perform using the SS and IG interfaces.

**Table 2**

### Exploratory search tasks

- |        |   |
|--------|---|
| Task 1 | Background: Imagine you intend to enter a photo competition on the topic of “Good variety food guide”, where you could win £50. This photo competition is being run by BBC Good Food: they are all about good recipes, and about quality home cooking that everyone can enjoy and like. The images you intend to present in this guide would show a variety of healthy and delicious inspiration, including a decadent dessert. It would also present trustworthy guidance for even some foodie needs. In order to get ideas for the competition, you want to look for already existing photographs conveying a similar subject. Your task is to find as many as diverse images that you think are the best fit to the topic “Good variety food guide”. |
| Task 2 | Background: Imagine you are an interior designer, specialist in lighting with responsibility for the design of a leaflet that illuminates customers about the chandelier options in terms of colors and shapes, which can be designed and intended for practical, or relaxing uses or both combined. Customers do not have knowledge and experience of lighting their homes. Your task is to find diverse chandelier images from a large collection of images that can be included in the leaflet. The leaflet is intended to raise interest among them and to have a variety of chandelier shapes lined up for matching customer requirements, style and budget.   |

Twelve participants were recruited through a mailing list, 9 postgraduate students and 3 undergraduate students, comprising 4 females and 8 males. Only 8 participants had adequate technical knowledge in search system design. All participants were familiar with text-based search, but not with search using query by example (CBIR). The experiment lasted about 55 minutes. In the first step the GSR device was attached, and the eye tracker calibrated before performing each task. The experiment was conducted in a UX lab. Data were collected using the iMotion platform. Post-task questionnaires were presented as stimulated recall after each task was performed. The experimental procedure is depicted in Figure 7. In order to avoid the impact of learning and fatigue, the stimulus order of search tasks was not fixed.



**Figure 7:** Experimental procedure

As shown in the experimental procedure, each participant was informed of the study objectives and their consent obtained; they completed a background survey. Before performing any tasks, we provided each participant with training on each system, since the quality of query formulation has significant impact on search results, and it can be beneficial to involve users in all retrieval processes [5]. The questions addressed here are:

- Q1:** To what extent can user engagement in query formulation improve the user involvement during post-focus stage?
- Q2:** To what extent can user engagement in focus-formulation stage affect user perceptions?

## 5. Results and discussion

Visual analysis of fixation and heat map patterns is presented including the heat maps on objects of the interface; eye gaze fixation path activities are then presented. Heat maps were generated during the two exploratory tasks that participants performed using the IG and SS interfaces. Eye fixation is one of the most widely used indicators in eye tracking studies [30]. They can illustrate precise visual attention of the interaction activities occurring with the search system. To address the first research question (Q1), where the recorded data for both eye fixation and heat maps were generated by eye tracker devices, it is observed among the participant data that high fixation rate denotes high attention paid by participants on a target image. The heat maps are objective attributes, representing the time spent on a certain object (image). Therefore, heat maps are useful to observe potential issues related to user perception, for instance, interface usability, task completion, task performance, and task complexity. Here, we look at



how users assess the result panel on two different interface designs. In this context the eye tracker helps us to spot additional insights from image search elements. The recorded data were aggregated to enable static visualization and then heat maps were generated. In this analysis we look at the aggregated and individual levels, thus eye tracking data of areas of interest (AOI) metrics per participant and AOI fixations per participant were exported to statistical software (SPSS). To spot the importance of user engagement in the focus-formulation stage, Figure (8-a) illustrates the customized heat maps that easily indicate the most heat focused element were where the image results presented on the SS and IG interfaces. Clearly, more attention can be observed on the SS interface, which means more images were selected, unlike the attention rate on the IG interface result panel (Figure 8-b) where the user had intent to change the search system, on the bottom right corner. It is also noticed that high heats were recorded on images that have similar color background with different object texture design.



**Figure 8:** Heat maps generated on the result panels (scenes) of (a) SS and (b) IG interfaces

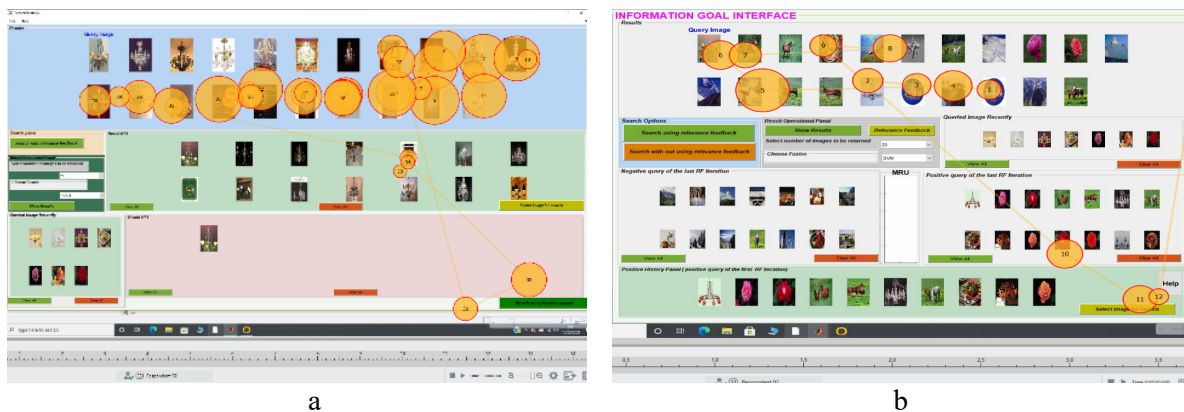
In order to quantify visual attention at both individual and aggregated levels, we first aggregated multiple dynamic events in the recorded video stimuli. Within the recording sensor, we created a scene as a segment. The segment was defined with a fixed size allocated manually along the timeline for each participant. The created scenes were treated as static stimuli at the individual level. Four static AOIs of eye tracking matrices were generated and used for analyzing the created scene. The drawn AOIs are to quantify the visual attention on the result panel. The heat maps indicate highlights where the participant’s attention was focused, in Figures 9-a and 9-b of the SS and IG interfaces, respectively.



**Figure 9:** Created areas of interest (AOIs) on the result panels of the (a) SS and (b) IG interfaces

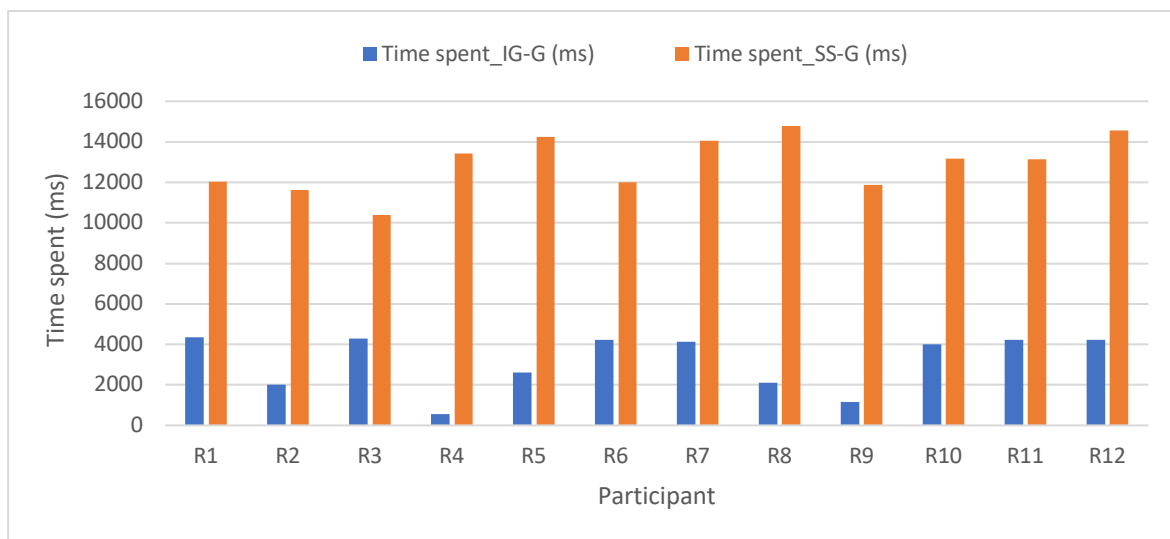
Figures 10-a and 10-b show the gaze path activity at the individual level with option to observe the duration of a fixation, unlike dynamic or static gaze path. The gaze patterns are for Task 2 on both interfaces. The circle size indicates the fixation time, where the radius increases with longer fixation. To obtain more understanding, the fixation values can be observed across the presented AOIs in Figure 9. The number of fixations per image is related to how long a participant engages on interface elements or with useful images they might have seen. Eye tracking data brings evidence that the users’ actions are not randomly taken. When the results do not fit user needs, the gaze fixation path

illustrates how participants try to find alternative search methods to find a desired image (see right corner at the bottom).



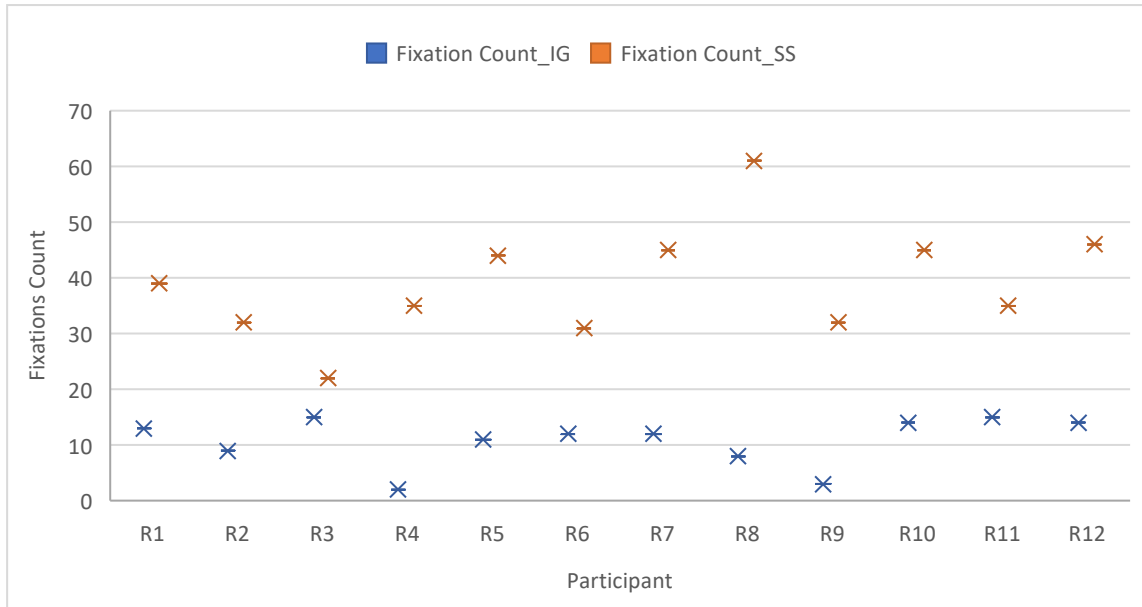
**Figure 10:** Individual gaze maps on the result panels of the (a) SS and (b) IG interfaces

The Mann–Whitney U-test was applied as an independent samples t-test, performed on ordinal recorded data on the result panels of the IG and SS interfaces. This is to find the significance of the difference between the time spent and fixation counts on both results of the IG and SS interfaces. We found that there is a significant difference between the two elements of time spent (ms) in AOI with no fixation for the IG and SS interfaces (based on raw data) at  $p = 0.00032$ . Moreover, the Mann–Whitney U-test shows a high significance at  $p < 0.05$  (at 0.0031) on the total duration spent in AOI of all participants' fixations (excluding data points between fixations). The significant difference between the amounts of time spent in AOI-based raw data for both interfaces can be seen in Figure 11.

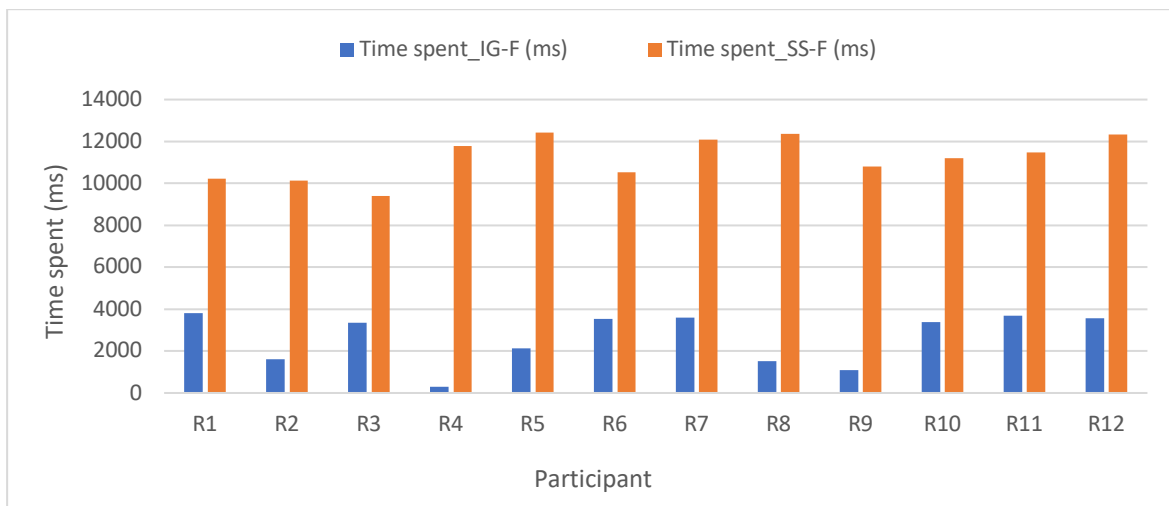


**Figure 11:** Time spent in AOI (not fixation based) of the result panel on IG and SS interfaces

Figure 12 shows the number of fixations recorded inside the AOI of the result panels using IG and SS interfaces. The fixation counts on SS results are distributed with higher scores than the IG results. The higher the amount of total fixation counted inside the AOI, the more time a participant spent on an AOI with high interest. Similarly, as seen in Figure 13, the total times spent in AOI of all participants' fixations (excluding data points between fixations) were significantly higher at the result-assessment stage of the SS than the IG interface.



**Figure 12:** Number of fixations recorded inside the AOI of the result panels on the SS and IG interfaces

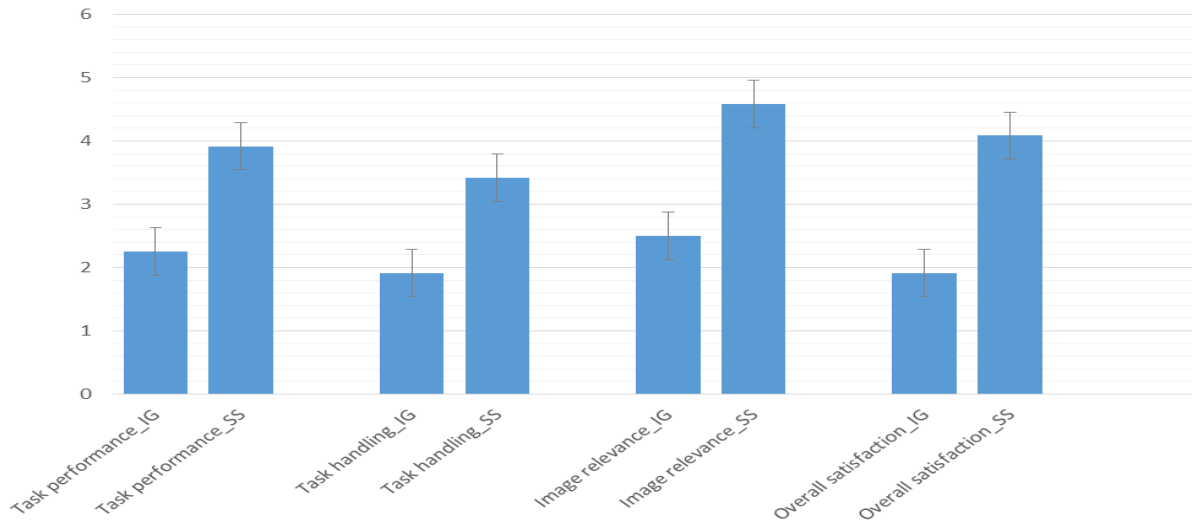


**Figure 13:** Total time spent in AOI of all participants' fixations (excluding data points between fixations) of the result panels on the SS and IG interfaces

Meanwhile, eye tracking data indicates how participants experience the exploratory image search tasks; the participants' perceptions in respect of both search systems (i.e., IG and SS) are important to address the second research question (Q2).

To report the participants' perceptions, we aggregated and exported the survey data per participant from the stimuli of a 5-point Likert scale. The survey data is from post-task questionnaires. Figure 14 shows the average of four elements of users' perceptions during the second task performed on interfaces IG and SS. The results show that among the four factors, the SS interface outperforms the IG interface: this includes task performance rate, the approach to task handling, the number of returned relevant images, and overall user's satisfaction rate.

Figure 15 shows the individual heat maps from the post-task questionnaires (PTQ) for SS and IG. It highlights the benefit of using eye tracker data, which can complement the survey stimuli evaluation, as the participant's attention focuses more on the area that receives a number of mouse clicks.



**Figure 14:** Comparison of user perceptions towards IG and SS interfaces; error bars represent standard deviation



**Figure 15:** Individual heat maps on the PTQ panels of the (a) SS and (b) IG interfaces

## 6. Conclusion and future work

We used eye tracking sensor and GSR sensor in the user study to determine at which stage users pay high attention. The evaluation of user search experience comprised of task performance, task handling, returned relevant images, and overall satisfaction. The aggregated eye tracking data were helpful to identify at which stage more attention was paid. The finding shows that while participants engage in the focus-formulation stage (i.e., SS interface) during both search tasks, the SS aggregated heat maps and gaze fixations were noticeably higher at the end of the search process (result panel) than when the IG interface applied. The analysis of recorded eye-tracking data revealed that the gaze behavior patterns can complement the survey stimuli evaluation by examining the gaze navigation behavior and fixations. The limitation of this paper is that the recorded GSR data were not integrated with the eye tracking data. Further investigation will be devoted to shed light into the key factors of CBIR-approach design by increasing the number of participants and aggregating GSR and eye tracking data. That is in order to pave the way to obtain a better image search paradigm.

## 7. References

- [1] V. Tyagi, *Content-Based Image Retrieval: Ideas, Influences, and Current Trends*. Springer, 2018.
- [2] A. W. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 12, pp. 1349-1380, 2000.
- [3] H. Liu, S. Zagorac, V. Uren, D. Song, and S. Rüger, "Enabling effective user interactions in content-based image retrieval," in *Asia Information Retrieval Symposium*, 2009: Springer, pp. 265-276.
- [4] A. Mohanan and S. Raju, "A Survey on Different Relevance Feedback Techniques in Content Based Image Retrieval," *International Research Journal of Engineering and Technology*, vol. 4, no. 02, pp. 582-585, 2017.
- [5] W. Zhou, H. Li, and Q. Tian, "Recent advance in content-based image retrieval: A literature survey," *arXiv preprint arXiv:1706.06064*, 2017.
- [6] M. Artemi and H. Liu: Content-based Image Search System Design for Capturing User Preferences during Query Formulation. Proc. of BIRDS2020, Xi'an, China, July, 2020, <http://ceur-ws.org/Vol-2741/paper-12.pdf>
- [7] H. Zhang, Z.-J. Zha, Y. Yang, S. Yan, Y. Gao, and T.-S. Chua, "Attribute-augmented semantic hierarchy: towards bridging semantic gap and intention gap in image retrieval," in *Proceedings of the 21st ACM international conference on Multimedia*, 2013, pp. 33-42.
- [8] L. Piras and G. Giacinto, "Information fusion in content based image retrieval: A comprehensive overview," *Information Fusion*, vol. 37, pp. 50-60, 2017.
- [9] H. Liu, P. Mulholland, D. Song, V. Uren, and S. Rüger, "Applying information foraging theory to understand user interaction with content-based image retrieval," in *Proceedings of the third symposium on Information interaction in context*, 2010: ACM, pp. 135-144.
- [10] B. Settles, "Active learning literature survey," University of Wisconsin-Madison Department of Computer Sciences, 2009.
- [11] X.-D. Zhang, "Machine learning," in *A Matrix Algebra Approach to Artificial Intelligence*: Springer, 2020, pp. 223-440.
- [12] B. Settles, M. Craven, and L. Friedland, "Active learning with real annotation costs," in *Proceedings of the NIPS workshop on cost-sensitive learning*, 2008: Vancouver, CA., pp. 1-10.
- [13] C. Sammut and G. I. Webb, *Encyclopedia of machine learning*. Springer Science & Business Media, 2011.
- [14] X. S. Zhou and T. S. Huang, "Relevance feedback in image retrieval: A comprehensive review," *Multimedia systems*, vol. 8, no. 6, pp. 536-544, 2003.
- [15] P. B. Patil and M. B. Kokare, "Relevance Feedback in Content Based Image Retrieval: A Review," *Journal of Applied Computer Science & Mathematics*, no. 10, 2011.
- [16] D.-p. Tian, "A Review on Relevance Feedback for Content-based Image Retrieval," *J. Inf. Hiding Multim. Signal Process.*, vol. 9, pp. 108-119, 2018.
- [17] S. Jones, L. Shao, and K. Du, "Active learning for human action retrieval using query pool selection," *Neurocomputing*, vol. 124, pp. 89-96, 2014.
- [18] M. Artemi and H. Liu, "Image optimization using improved gray-scale quantization for content-based image retrieval," in *2020 IEEE 6th International Conference on Optimization and Applications (ICOA)*, 2020: IEEE, pp. 1-6.
- [19] P. Vakkari, "A theory of the task-based information retrieval process: a summary and generalisation of a longitudinal study," *Journal of documentation*, vol. 57, no. 1, pp. 44-60, 2001.
- [20] K. Athukorala, A. Oulasvirta, D. Głowacka, J. Vreeken, and G. Jacucci, "Narrow or broad?: Estimating subjective specificity in exploratory search," in *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, 2014: ACM, pp. 819-828.
- [21] M. Hearst, *Search user interfaces*. Cambridge university press, 2009.
- [22] H. C. Huurdeman and J. Kamps, "Designing multistage search systems to support the information seeking process," in *Understanding and Improving Information Search*: Springer, 2020, pp. 113-137.
- [23] R. W. White, I. Ruthven, and J. M. Jose, "A study of factors affecting the utility of implicit relevance feedback," in *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, 2005, pp. 35-42.
- [24] X. Niu and D. Kelly, "The use of query suggestions during information search," *Information Processing & Management*, vol. 50, no. 1, pp. 218-234, 2014.
- [25] B. Kules, R. Capra, M. Banta, and T. Sierra, "What do exploratory searchers look at in a faceted search interface?," in *Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries*, 2009, pp. 313-322.

- [26] H. C. Huurdeman, M. L. Wilson, and J. Kamps, "Active and passive utility of search interface features in different information seeking task stages," in *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval*, 2016, pp. 3-12.
- [27] H. Liu, D. Song, and P. Mulholland, "Exploration of Applying a Theory-Based User Classification Model to Inform Personalised Content-Based Image Retrieval System Design," in *Proceedings of HCI Korea*, 2016: Hanbit Media, Inc., pp. 61-68.
- [28] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.
- [29] S. Amershi, M. Cakmak, W. B. Knox, and T. Kulesza, "Power to the people: The role of humans in interactive machine learning," *AI Magazine*, vol. 35, no. 4, pp. 105-120, 2014.
- [30] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. Van de Weijer, *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford, 2011.