Content-based Image Search System Design for Capturing User Preferences during Query Formulation

Mahmoud Artemi[1][0000-0002-5177-8977] and Haiming Liu[2][0000-0002-0390-3657]

[1]Institute for Research in Applicable Computing, University of Bedfordshire. Luton, UK
[2]School of Computer Science and Technology, University of Bedfordshire. Luton, UK
[1]mahmoud.artemi@study.beds.ac.uk
[2]haiming.liu@beds.ac.uk

Abstract. Most existing studies of content-based image retrieval (CBIR) system design focus on learning users' information needs through relevance feedback at the result assessment stage only. However, in many CBIR systems, the underlying machine learning mechanisms need the users' feedback at query formulation stage for a better training and search performance, which unfortunately is often not supported by the search interface design. The lack of support for the users' query formulation through an effective CBIR interface has been a drawback for system performance and the users' search satisfaction and experiences. We propose a new CBIR system design approach based on Vakkari's three-stage model, which encourages the users to provide feedback at the query formulation stage through a user-centered interface. The interface helps the users to form and express their information needs through enabling the users to participate in the training phase of the machine learning mechanism of the system. A user study with 28 participants shows how the proposed system design supports the users' interaction through the user-centered search interface. The findings of this study highlight the importance for the users to engage in all stages of the search process, especially at the query formulation stage when the considered mechanism requires a training process, through a user-centered interaction design.

Keywords: Content-based image retrieval, interactive machine learning, user interface, relevance feedback, Vakkari's three-stage model, query formulation.

1 Introduction

With the massive growth of the number of digital images online, it is a significant challenge to find required images from the massive image repositories using text descriptions or image example(s). We are familiar with searching images using text, however less familiar with search images using image example(s), namely content-based image search (CBIR). One well-known challenge in CBIR is called Semantic Gap, which is a gap between how users interpret the images (abstract objects, an event) and how computers understand the images (colour, shape, texture). Many research has investigated how to bridge the semantic gap by involving the users in the CBIR search loop [13-15]. Interactive image retrieval approaches have been an effective way to bring the users into the CBIR search loop, which allows the users to provide relevance feedback to obtain improved results. Most of the research on relevance feedback for interactive search focuses on enabling users to provide feedback at the result assessment stage [8, 15]. However, often the underline machine learning mechanisms in many CBIR systems need the users' feedback at query formulation stage for a better training and search performance. There is a need to design an interactive CBIR search system that does not only allows the users to interact with the retrieved image results but also allows the users to visually explore the image collection and facilitates the users to train the underlying search model through a user-centered interactive search interface, therefore to improve the search performance and the users search experiences and satisfaction [13, 15].

In this paper, we introduce an Explicit Searcher Model (ESM) developed based on the concept of Vakkari's three-stage model for an interactive CBIR system design. We design a user-centered search interface to visualize the ESM model. The interface allows the users to provide relevance feedback at query formulation stage to train the underlying search model. We evaluate the ESM model and the interface through a user study. The findings show that the proposed system outperforms the baseline systems, based on pre-selection training data and a system that allows the users to provide relevance feedback at the result assessment stage only. This research enables us to better understand user information needs and the influence of user interaction on the search performance, experience and satisfaction.

2 Related Work

The work presented in this paper is shaped by prior studies in the area of interactive information retrieval, especially user interaction and user interface design for a better CBIR search experience.

2.1 Interactive Search

Machine Learning (ML) algorithms are applied to many CBIR search systems to support the interaction between the users and the system. For example, ML helps the system to learn the users' need based on the users' interaction with the system such as relevance feedback from result assessment [1]. Liu et al. proposed a four-factor user interaction model to improve user interactions with a CBIR system [8]. Zhuang et al. conducted a study of undefined image search task, which uses an image explorer to investigate how user perception is associated with both involvement and attention of user behavior in interactive search settings [19]. Liu et al. developed the uInteract system based on user-centered design for interactive CBIR search [7], which uses the relevance feedback feature to refine the results and formulate a new query. Others have discussed the importance of considering the intent and refinement of the image search activities to understand the user's behavior [18]. This refinement is steered by the user's input, which may be provided in different forms, such as providing image samples. Figure 1 shows the user's role in the ML paradigm when a Support Vector Machine (SVM) with active learning as considered in this study.

Since machine learning needs human intelligence during the refinement process [7], in this study we use an ML with active learning algorithm that allows the individual's engagement in a set–check loop for training the ML model. The ML provides the users with image samples (feedback visualization) that enables the users to supply the model with further information (labelling or demonstration); this is in order to improve the model performance [5]. The workflow of ML-based active learning is driven by the user to shape the search process; this process enables the user to have control over the high-level system interactions. This is distinguished from classic ML, which is used in this study as a baseline system. The baseline system basically applies a classification model based on pre-selection of training data by the system, which means that the users are not involved in the training stage of the model and search process.

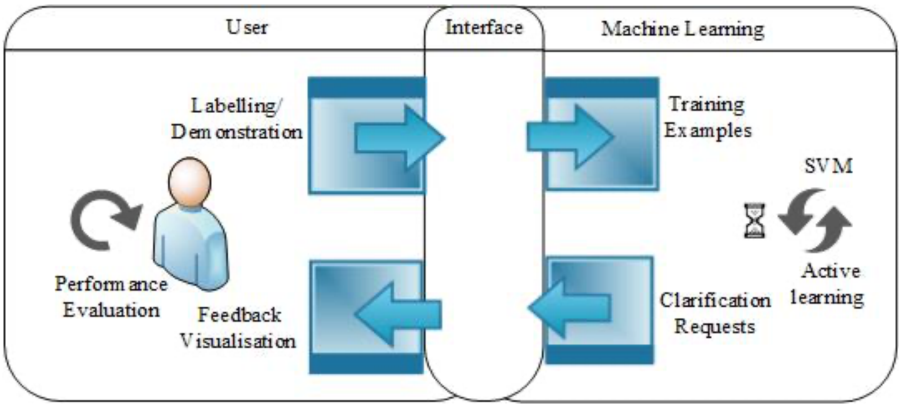


Figure 1. The user's role in Machine Learning

2.2 User Interface

User interfaces are typically designed to submit a query and display a set of search results. In recent literature, this convention is changing and aims more towards offering a better search experience for the users. These user interfaces mainly focus on one or a combination of the following aspects [14]: (1) supporting easy browsing of the image collection; (2) achieving a better presentation of search results; (3) providing relevance feedback on returned results; (4) allowing users to group and move around query images; (5) enabling users to indicate the level of relevance to their information needs.

In this paper, we design a user-centered interface to enable users to engage not only in the result assessment stage but also in the query formulation stage. It is easier to train the search model by incorporating the user's relevance feedback during query the formulation stage. This process can help improve the users' search experience and the system performance to meet the user's information needs.

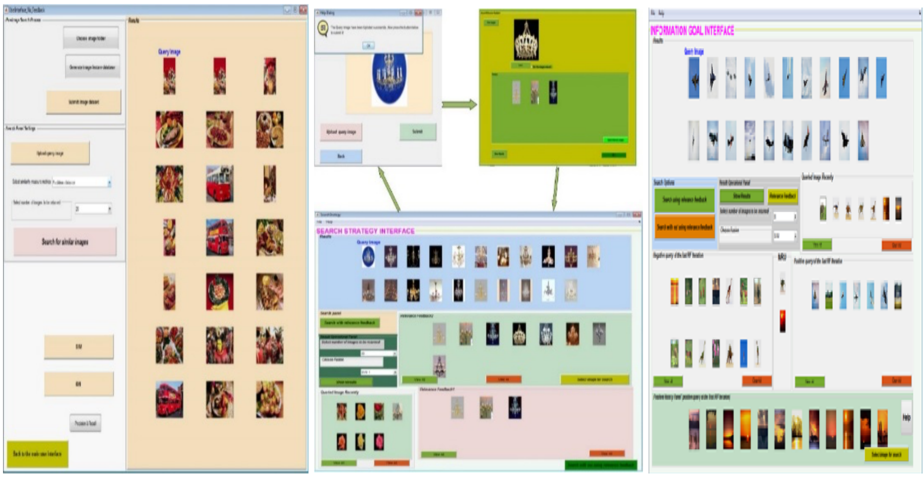
2.3 Vakkari's three-stage model

Vakkari's task-based theory is considered [16], which 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* which comes 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 [4]. Decision-making may occur during the search process and continue to be presented at the assessment stage (*post-focus*). The user assesses a set of returned images to find not only relevant images but rather the best images that fit a given task and have value for the user's needs.

3 The Proposed System

The literature motivates us as we can see many interesting machine learning approaches that can be successfully used to enable effective interaction to happen, one of which is using an SVM-based retrieval with active learning approach. Here, the Search Strategy (SS) interface is presented: it enables our proposed CBIR system (also named SS) to capture user preferences during the query formulation step, where users can provide additional images within the training stage, other than a providing relevance feedback to refine the top of the result list, which the system already knows. The SS interface has three frames (Figure 2-b): the upper left window is for exploring and selecting N random images. The upper right 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 window, the CBIR system returns a diversity of resultant sets considered matching the concept learned, where a user assesses the retrieved image set as being relevant and useful (image *utility*). We also present an Explicit Searcher Model (ESM) that captures the sequence of interactions between a searcher and the CBIR system over the course of a search session. Existing searcher models such as the Complex Searcher Model [9] have not investigated the effectiveness of user engagement in system training as well as the exploratory search process, and why searchers behave as they do has received relatively less attention [10]. The interactive SS system using SVM active learning is developed in this study as it is successful and suitable for situations where data is abundant [17]. The training set-selection algorithm based on SVM active learning selects the most informative images to learn the decision hyperplane and separate unlabeled images to satisfy the user needs [11]. It allows the user to indicate their preferences: in the first place the model requests the user to interactively label a number of images, randomly selected from the image collection, as relevant or irrelevant to their preference. Then, the learner refines the underlying model by choosing samples from a pool of unlabeled images and requests labeling from the user to provide insight into the underlying feature distribution [6]. This method is successful in accelerating learning [2].

The image feature extractors and their parameters are implemented in this study as presented in [3]. Our SS system is compared with two baseline CBIR systems: the Customized system uses SVM without Relevance Feedback (RF) tool ("Cu system") [3] and the Information Goal system which uses RF technique (IG system) [9].



(a) SVM without Relevance Feedback (RF) tool (Customised system,[3]), (b) proposed Search Strategy system, (c) Information Goal system [7].

Figure 2. User interfaces used in this study

The user's perceptions of our Search Strategy (SS) system in real search scenarios are reported, and compared with both baseline CBIR systems. This is in order to find how user engagement could affect search behavior and thus their satisfaction. To simulate and understand user interaction, the three-stage Vakkari model is considered; the search process of the SS system based on the ESM is depicted in Figure 3.

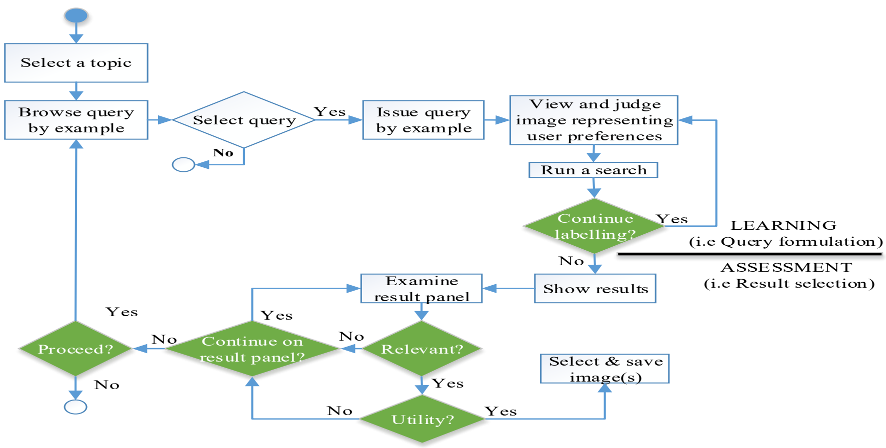


Figure 3. A flowchart of the Explicit Searcher Model (ESM). The ESM is used in this study to simulate the interaction between user and CBIR system.

3.1 Workflow of Explicit Searcher Model (ESM) Through SS Interface

The ESM, guided by the three-stage Vakkari model, consists of several of stages which were taken by a user (Figure 3). The processes are shown in boxes and decision points in diamonds. The flow is divided into logical learning and assessment interactions.

Pre-focus: starts by image exploration to select the search topic where a user gives an image query to the system. Once image searching has been performed, the user can select the query images that reflect their preferences and then the query can be issued through the interface. A user needs to set the number of training image they wish to label as relevant or irrelevant (Figure 2-b, upper left), then the system returns a number of randomly selected images from the image collection. Within the *pre-focus* stage if a user could not find a candidate query, a stopping option point is available (Figure 2).

Focus-formulation: a searcher labels a number of images for further exploration from a pool of unlabeled images as query formulation (see Figure 3) during two iterations. As a result, the system has to re-learn the input features with each new query, where a user uses the interface shown in Figure 2-b, upper right. Our goal for selecting this method is that user can have an opportunity to steer the system as those unlabeled items are more uncertain ones to the system and thus, the user typically can train the model mapping those images to their corresponding vector location in the search space; the user decision in this selection process may impact on result diversity and relevance.

Post-focus: once the image labeling has been established by searcher for training the model as shown in Figure 3, the search system returns a set of images for the searcher to judge. From here, the user is able to view a number of top-ranked returned images, up to 100, that are presented by the Search Strategy interface (Figure 2-b, bottom window), where the default number is set at 20: that is, if the results do not look relevant or promising, the searcher can abandon the result and issue another query by example. If the returned images look relevant, the searcher will then start to examine each image individually. Once the searcher assesses the associated image for relevance/*utility*, if deemed to have value to the information needs, the image is then selected and saved. If it is not useful, the searcher then moves to assess another image, and so on. The searcher will typically stop assessing the returned images and abandon the result panel and then may proceed to further query or stop searching.

In order to obtain useful information during the literature review on how to design a better image search system, understanding the impact of users' engagement on their search perception and behavior is important, which can be investigated during each stage of Vakkari's model: initiate, select the query and explore the image collection in the *pre-focus* stage; labeling the learner-driven point selection to refine the search goal in the *focus-formulation* stage; result assessment and collecting the images in the *post-focus* stage.

4 Evaluation

Although driven-point selections are leveraged by the search system (learner) for labeling by the user, conducting a controlled laboratory evaluation based on user-orientation within Vakkari's three-stage model of the information seeking process can offer additional insight into how the participant's engagement in query formulation at the system training level, as well as at the result assessment step, influences their perceptions of such as usefulness, satisfaction, feelings in comparison with baseline CBIR systems. In order to make a sensible comparison and discover how users develop their interactions in specific settings, two baseline systems are considered, the Information Goal (IG) interface which facilitates with RF technique for three iterations [7] is presented in Figure 2-c, and the Cu interface [3] is depicted in Figure 2-a; The Cu is provided with different distance metrics (Euclidean, Standardized Euclidean, Mahalanobis, City block, Minkowski, Chebychev, Cosine, Correlation, Spearman, and Manhattan) as presented in [5], in addition to SVM-based fusion.

Table 1. 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 diversity 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 leaflet that illuminates customers about the chandelier options in terms of colours and shapes, which can be designed and intended for practical, relaxing use or both combined. Customers do not have knowledge and experience of lighting their homes. Your task is to find diversity of 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.

4.1 Experimental Design

To investigate effectively the interactions that occur between user and CBIR search system, and obtain evidence of what influences a user's behavior when they contribute in all search aspects, we designed a controlled study to obtain explicit feedback on search satisfaction from participants. Each of our participants performed two exploratory-image search tasks on each search system. These tasks are given in Table 1.

Twenty-eight participants were recruited from the Institute for Research in Applicable Computing at the University of Bedfordshire. The duration of the experiment was about 90 minutes. The experiment was conducted in the lab settings within the institute. Data were recorded in different forms: (1) screenshots and video, (2) pre-test for investigating working memory, (3) background survey, (4) post-task questionnaires were provided after performing each task, (5) post-experiment questionnaire, (6) notes taken by structured interview. We imported all these data sources in combined form to SPSS.

Regarding the search tasks, participants were asked to perform two lookup tasks using known-item search to investigate user lookup behavior, and two exploratory tasks using three image search systems. In the lookup search, participants were required to select relevant images, where one image was presented to the participant in each round; at the end, participants were asked to select useful images (image *utility*) to the given task. The procedure of the experiment is depicted in Figure 4. In order to avoid the impact of learning and fatigue, the order of search tasks was rotated by applying Latin square design across the three systems.

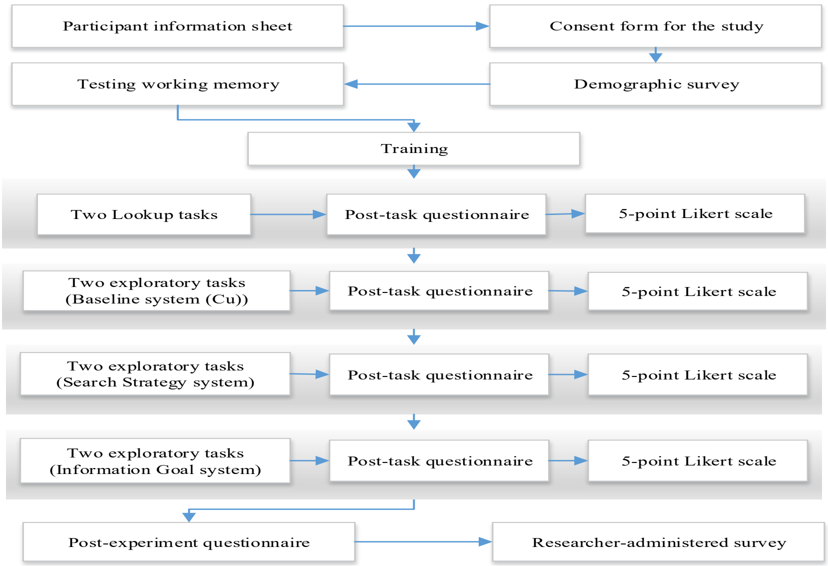


Figure 4. Experimental procedure

As shown in the experimental procedure, each participant was informed of the study objectives and their consent obtained; they completed a background survey and the memory test capacity was performed by using an n-back test "cognitive fun". Before performing any tasks, we provided each participant with sufficient training on each system. Four questions are addressed here:

RQ1: To what extent can using an interactive search system improve user experience in terms of user effectiveness and efficiency?

RQ2: In comparison to the baseline systems, how satisfied were participants with using the Search Strategy system?

RQ3: Which system did participants find easy to use, and more useful?

RQ4: Which system did participants feel in control of, and more confident with?

4.2 Data Analysis

Two types of data were collected, the user perceptions of system evaluation and the user interaction behaviors. Here, we present the data on participants' perceptions of the Search Strategy (SS) system, including satisfaction, usability, usefulness, and system performance in terms of image relevance and image usefulness (image *utility*) with respect to the Cu and IG systems, this is in order to address the benefit of query formulation at the system training level.

In order to investigate the user experience and perceptions of overall task performances (RQ1), a one-way ANOVA with post-hoc Tukey test was conducted. The data were collected using a five-point Likert scale for comparing multiple tasks to address the overall perceived satisfaction (RQ2), the overall usability and the perceived usefulness (RQ3), and confidence and control (RQ4) questions. The participant averages for those high-level constructs are shown in graphically for the three systems (Cu, SS, and IG).

5 Results and Discussion

In this study the participants performed two lookup tasks and two exploratory tasks, each participant has to find the useful images which fit the assigned task. We observed that the overall number of useful images (image *utility*) selected by each participant were less than selected relevant images for both assigned tasks.

RQ1: User overall task performance experience. The two search tasks were performed using Cu, SS, and IG systems; the user perceptions across the three stages of information seeking model were recorded and analyzed as participant success over the whole task process. Figure 5 shows the user success rate in overall task performance.

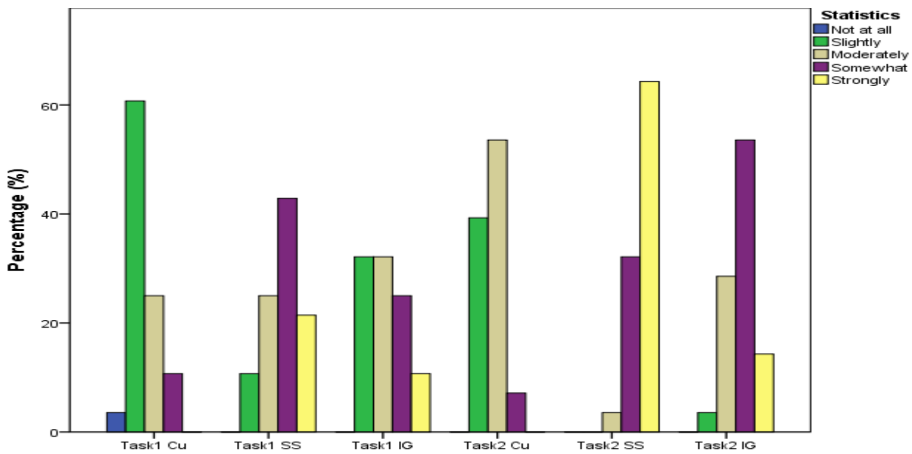


Figure 5. User success in overall task performance

The measures of task success used in the post-task questionnaire are efficiency, which means assigned time to the task, and effectiveness which is task completion. It is clear that in Figure 5, the user success rate in the second task was significantly higher compared to task 1 performed on our search system (SS), unlike other two systems (Cu and IG) where there no significant improvement in overall task performance were observed.

Table 2. Statistical analysis (ANOVA) of overall task success performance

Multiple Comparisons

Dependent variable: *task success performance*

(I) system	(J) system	Mean Difference (I-J)	Sig.
T1 SS	T1 Cu	1.32143*	.000
	T1 IG	0.60714*	.047
T2 SS	T1 Cu	2.17857*	.000
	T1 SS	0.85714*	.001
	T1 IG	1.46429*	.000
	T2 Cu	1.92857*	.000
	T2 IG	0.82143*	.002

* The mean difference is significant at the 0.05 level.

The ANOVA with Tukey multiple comparisons shows significantly increase of task success performance in task 2 compared to task 1 using SS system at [$F (5,162) = 29.85$, $p < 0.005$]. Considering the overall user performance, the participant's experience with the SS system has significantly improved, unlike the other two search systems (Cu and IG). This is likely to be related to the success in query formulation when participants train the search system and gain more search experience over the search course; this process facilitates participants to find the diversity (*utility*) of images within the same image set instead of image relevance. When using the SS system it is worth considering user interest and knowledge in all search task aspects, including system training session (*pre-focus* and *focus-formulation*). From our observations, in addition to issue queries, participant engagement was more at *pre-focus* and *focus-formulation* stages than *post-focus* stage. In the *focus-formulation* stage, some participants increased the number of images to be labeled for system training. Whereas in the *post-focus* stage we noticed that participants were primarily considering image *utility* selection as a sense-making strategy. This is in line with the hypothesis of the adaptive interaction framework [12]. We can assume that the way that participants process the information in their brain plays an important role in evaluating the system outcome when there is a diversity of relevant images, and this potentially makes participants look at the value of images that fit their needs in ways other than relevance. It is obvious that participants reported higher scores for relevance and lower positive scores in perception of image *utility* after assessing the results in the *post-focus* stage using SS system which represents the diversity of image, it still significantly higher than other two search systems, as depicted in Figure 6. On the other hand, we observed very little difference between the tasks performed on both Cu and IG systems (Figure 5), and this may be expected due to lack of user formulation to train the search system based on user preferences.

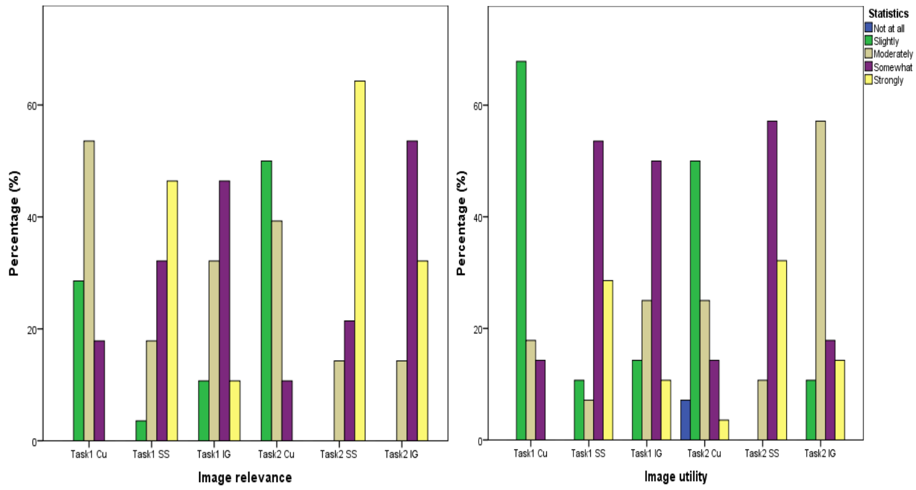


Figure 6. User perceptions of overall image relevance and *utility*

Table 3. Statistical analysis (ANOVA) of user perceptions of image *utility* scores

Multiple Comparisons

Dependent variable: *image utility*

(I) system	(J) system	Mean Difference (I-J)	Sig.
T2 SS	T1 Cu	1.75000*	.000
	T1 SS	0.21429	.931
	T1 IG	0.64286	.052
	T2 Cu	1.64286*	.000
	T2 IG	0.85714*	.003

* The mean difference is significant at the 0.05 level.

Statistically significant differences across user perception of image *utility* scores are shown in Table 3. ANOVA–based Tukey multiple comparisons of image *utility* data resulted in $F\ (5,162) = 20.78$, $p < 0.005$. Generally, a user-oriented approach is crucial to evaluate search system outcomes. There is a highly significant difference between Task1_SS and Task1_Cu for image relevance score, as well as between Task1_SS and Task1_Cu for image *utility*. This emphasizes that our SS system has improved user experience over tasks, unlike the Cu and IG systems where no improvement in user experience was observed.

RQ2: Participant satisfaction: Our results (Figure 7) show that participants gave more positive responses for satisfaction (RQ2), feeling in control and confidence (RQ5), and usefulness after performing tasks on the SS system in comparison to the Cu and IG systems, but a few negative perceptions were reported on the SS system regarding ease of use (RQ3).

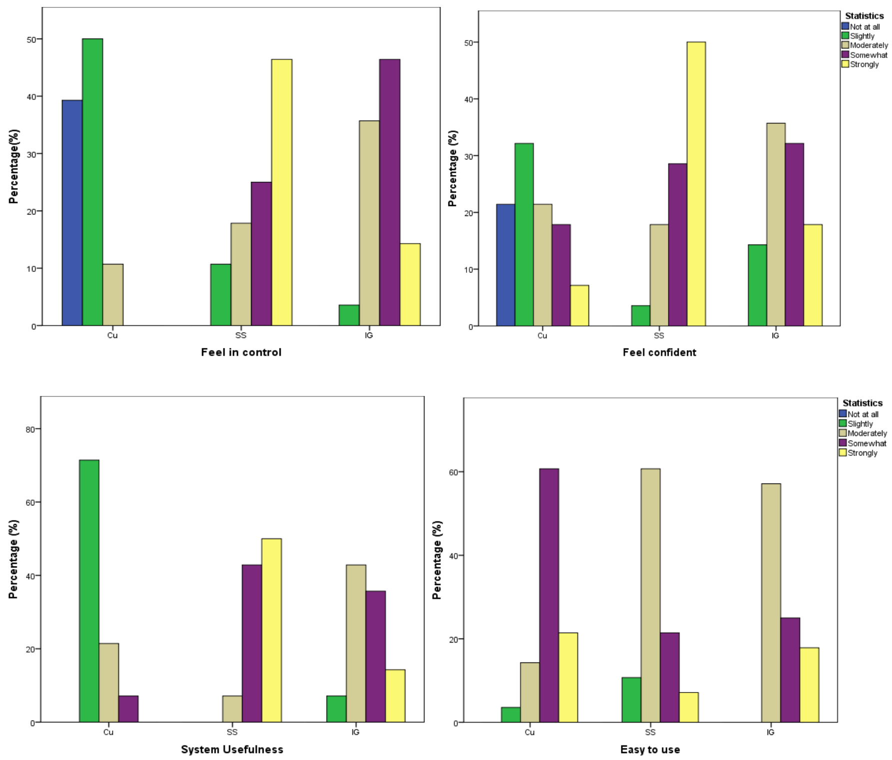


Figure 7. Feedback responding to overall (a) feeling in control, (b) confidence in use (c) ease of use, (d) usefulness, using the SS and Cu interfaces.

6 CONCLUSION

There are three main contributions in this paper: First, we developed an interactive CBIR system-based SVM active learner to consider user-centered design in which participants engage in a set–check loop (*focus-formulation* stage) for training the ML model based on their preferences. The evaluation results show that this procedure helps the overall task performance and user search experience. The proposed CBIR system represents the diversity of image in which users find image *utility* instead of image relevance that fulfill their needs. Second, we proposed an Explicit Searcher Model (ESM) based on Vakkari's three-stage model of information seeking to design a guide for the image seeking process and further. Third, we observed that participants' interactions with interactive CBIR system are adaptive in nature to their knowledge of the search system. It is hoped that this study provides insights into how participants' perceptions are influenced by their understanding of the image data, their interaction with the system through a supportive system design.

References

1. Alzu'bi, A., Amira, A. and Ramzan, N. Semantic content-based image retrieval: A comprehensive study. *Journal of Visual Communication and Image Representation*, *32*. 20-54, 2015.
2. Amershi, S., Cakmak, M., Knox, W.B. and Kulesza, T. Power to the people: The role of humans in interactive machine learning. *AI Magazine*, *35* (4), (2014), 105-120.
3. Artemi, M. and Liu, H., 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, 1-6.
4. Athukorala, K., Oulasvirta, A., Głowacka, D., Vreeken, J. and Jacucci, G., 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, 819-828.
5. Chomboon, K., Chujai, P., Teerarassamee, P., Kerdprasop, K. and Kerdprasop, N., An empirical study of distance metrics for k-nearest neighbor algorithm. in, (2015).
6. Kremer, J., Steenstrup Pedersen, K. and Igel, C. Active learning with support vector machines. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *4* (4). 313-326.
7. Liu, H., Song, D. and Mulholland, P., 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., 61-68.
8. Liu, H., Zagorac, S., Uren, V., Song, D. and Rüger, S., Enabling effective user interactions in content-based image retrieval. in *Asia Information Retrieval Symposium*, (2009), Springer, 265-276.
9. Maxwell, D. and Azzopardi, L., Information Scent, Searching and Stopping. in *European Conference on Information Retrieval*, (2018), Springer, 210-222.
10. Maxwell, D., Azzopardi, L., Järvelin, K. and Keskustalo, H., Searching and stopping: An analysis of stopping rules and strategies. in *Proceedings of the 24th ACM international on conference on information and knowledge management*, (2015), ACM, 313-322.
11. Nalepa, J. and Kawulok, M. Selecting training sets for support vector machines: a review. *Artificial Intelligence Review*, *52* (2). 857-900.
12. Payne, S.J. and Howes, A. Adaptive interaction: A utility maximization approach to understanding human interaction with technology. *Synthesis Lectures on Human-Centered Informatics*, *6* (1). 1-111.

13. Piras, L. and Giacinto, G. Information fusion in content based image retrieval: A comprehensive overview. *Information Fusion*, *37*. 50-60.
14. Thomee, B. and Lew, M.S. Interactive search in image retrieval: a survey. *International Journal of Multimedia Information Retrieval*, *1* (2). 71-86.
15. Tyagi, V. *Content-Based Image Retrieval: Ideas, Influences, and Current Trends*. Springer, 2018.
16. Vakkari, P. A theory of the task-based information retrieval process: a summary and generalisation of a longitudinal study. *Journal of documentation*, *57* (1). 44-60.
17. Witten, I.H., Frank, E., Hall, M.A. and Pal, C.J. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.
18. Wu, Z., Liu, Y., Zhang, Q., Wu, K., Zhang, M. and Ma, S., The influence of image search intents on user behavior and satisfaction. in *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, (2019), ACM, 645-653.
19. Zhuang, M., Toms, E.G. and Demartini, G., The relationship between user perception and user behaviour in interactive information retrieval evaluation. in *European Conference on Information Retrieval*, (2016), Springer, 293-305.