An Eye-Tracking Study of User Behavior in Web Image Search

Wanxuan Lu and Yunde Jia

Beijing Laboratory of Intelligent Information Technology, School of Computer Science, Beijing Institute of Technology, Beijing 100081, China
{luwanxuan,jiayunde}@bit.edu.cn

Abstract. Studies of Web search have mostly examined user behavior in text search. Recent studies begin to explore user's Web image search behavior through survey, questionnaire, interview, etc. This study investigates user behavior on a Web image search engine using eye-tracking. The goal is to get insight into how users view search results and whether search task type and results presentation order influence their behavior. We found that search results at certain locations, e.g., the top-center area in a grid layout, were more attractive than others. The search task type significantly influenced user behavior while the results presentation order didn't. In addition, we looked at the question of why a particular search result was selected, which showed variety of reasons. User behavior researchers and search engine developers can take advantage of these findings in order to create better search experiences to the users.

Keywords: Web Image Search, Eye-Tracking, User Study, User Behavior, Search Task Type, Results Presentation Order.

1 Introduction

Internet is one of the most important information resources in people's lives. People seek for online information mainly by search activities, which have gained lots of attentions. Knowing how users search on the Web can help improving both search engine performances and user experiences.

Researchers have done much work studying user's Web search behavior (Hsieh-Yee 2001, Jansen and Pooch 2001), but most of them focus on text search (i.e., searching for textual information or Web pages). People search the Web for various purposes, not only textual information but also multimedia objects such as images. As more and more people load images onto the Web, the availability of online images increases rapidly. Searching for images on the Web has become an everyday activity, which needs to be treated as important as text search.

Recently, there are studies focusing on user's Web image search behavior (Goodrum and Spink 2001, Pu 2003, Choi and Rasmussen 2003, Jörgensen and Jörgensen 2005, Cunningham and Masoodian 2006, Choi 2010, Huang and Kelly 2013). These studies employ technologies including survey, questionnaire, interview, log analysis, and user experiment. The goals are to investigate the information needs, the queries submitted to search engines, etc. However, none of them

D.-N. Pham and S.-B. Park (Eds.): PRICAI 2014, LNAI 8862, pp. 170–182, 2014.

have explored how users view the results returned by image search engines. Do they scan the results from left to right and top to bottom, i.e., following the Western reading habit? How many results do users view before clicking on one? The answers to these questions are beneficial to not only user behavior researchers, but also search engine developers.

In this paper, we describe a study using eye-tracking technologies to observe user behavior in Web image search tasks. We are interested in general characteristics as well as the influence of two factors: search task type (*Task Type*) and results presentation order (*Result Order*). Although there are lots of eye-tracking studies about user's Web search behavior, few focus on image search.

2 Related Work

2.1 User Behavior of Web Image Search

The research on user behavior of Web image search has become active over a decade. Goodrum and Spink (Goodrum and Spink 2001) analyzed Web image queries submitted to Excite, an Internet search service. The Excite did not provide specific mechanisms for image search at that time, so users needed to include image request terms such as "jpg" or "photo". Results showed that users input short terms for image queries, and most terms were used infrequently. Pu (Pu 2003) examined the difference between Web image and textual queries, and found that image queries had higher specificity and were refined more frequently.

Choi and Rasmussen (Choi and Rasmussen 2003) studied search queries for images in American history by questionnaire and interview. More than half of the queries in their study were "general or nameable" queries which could be expressed in key words (e.g., "a ruined castle"). In addition, types of event, action, and individual names were the most popular words. Jörgensen and Jörgensen (Jörgensen and Jörgensen 2005) reported an analysis of search logs made by image professionals. They observed that unique search terms were less frequent in their study than in other studies, boolean operators were heavily employed but were ineffective, and the tactics of query modification were largely experimental.

Cunningham and Masoodian (Cunningham and Masoodian 2006) focused on the understanding of casual image seeking behavior. In the study, 70% of the information needs were specific (referring to a specific person, event, or activity) and browsing was the primary strategy instead of searching. Choi (Choi 2010) investigated college student's process of Web image search. A variety of techniques such as questionnaire, interview, video capturing, and thinking-aloud were used to collect both qualitative and quantitative data, which indicated that the search process was influenced by contextual factors such as task goal, searching expertise, and topic familiarity. Huang and Kelly (Huang and Kelly 2013) conducted a survey exploring the image information need and seeking behavior of Chinese undergraduates in which the top searching motivation was "to make PPT document", the most images being searched were wallpapers, and the most favorite searching tools were search engines.

The present study extends above studies by showing how users view the search engine results page (SERP) during Web image search tasks. Unlike previous studies that used questionnaire, interview, or log analysis, we conducted an eye-tracking experiment to collect users' visual behavior data.

2.2 Eye-Tracking of Web Search Behavior

Eye-tracking technologies have been widely used to study user behavior of Web search, most of which are about text search. Klöckner et al. (Klöckner et al. 2004) identified three categories of search results processing strategy: "depth-first", "breadth-first", and "partially breadth-first". The "depth-first" strategy refers to examining each result in turn (starting from the top) and deciding immediately whether to click on the result or not. The "breadth-first" strategy refers to looking ahead of several results and clicking on the most promising one. The "partially breadth-first" strategy is a mix of the two aforementioned strategies. Granka et al. (Granka et al. 2004) explored how users interact with the SERPs of a search engine. They reported that users tended to scan search results from top to bottom, spent more time on higher ranked search results, and click on them more often.

Lorigo et al. (Lorigo et al. 2006) found that search task type influenced the task completing time and pupil dilation, while gender influenced the pattern for evaluating search results. Cutrell and Guan (Cutrell and Guan 2007) changed the snippet length and observed that increasing the length would improve the performance for informational queries but worsen it for navigational queries. In another study (Guan and Cutrell 2007), they examined the effect of target rank. Results showed that when targets were placed relatively low in the page, people spent more time searching and were less successful in finding them. Buscher et al. (Buscher et al. 2010) focused on the impact of advertisement (ad) quality, in which good quality ads received much more attention than bad ones and there was a strong bias against all ads.

Dumais et al. (Dumais et al. 2010) investigated individual differences in gaze patterns. By utilizing clustering algorithms, participants were divided into three categories: "Exhaustive", "Economic-Results", and "Economic-Ads". "Exhaustive" participants explored search results broadly, "Economic-Results" participants explored more narrowly by looking at some additional results, and "Economic-Ads" participants regularly looked at ads. González-Caro and Marcos (González-Caro and Marcos 2011) presented a study analyzing user's browsing behavior under different query intents. Users with different intents preferred different types of search results, paid attention to different parts of SERPs, and focused on search results with different ranks. Jiang et al. (Jiang et al. 2014) observed the changes of user behavior for relatively long search sessions in different types of search task. Users shifted their attention from top results to results at lower positions as time going on, and results became less and less attractive.

Our work differs from above studies in that we focus on Web image search. There are some obvious differences between image search and text search. For example, the results of image search are often presented in a grid layout, comparing

to the list layout commonly used in text search. And instead of using text snippets, image search uses image thumbnails which can describe the search results more precisely. Therefore, image search may differ from text search in user behavior which is worth investigating.

3 Experiment

In order to investigate how users search the Web for images, we designed an eye-tracking experiment to capture their eye movements in Web image search tasks. Eye-tracking technology has been widely used as a proxy for users' attention. By analyzing the eye movement data, we can get valuable insights about where users pay attention and in what order.

3.1 Design

In the experiment, we asked participants to conduct 10 Web image search tasks on a commercial search engine. The design of the experiment was *Task Type (2)* x *Result Order (2)*. The *Task Type* was a with-subject factor and the *Result Order* was a between-subject factor.

As described by Batley (Batley 1988), the information needs of image search can be categorized as "specific", "general or nameable", "general or abstract", and "general or subjective". We integrated the three "general" categories so that there were two types of search task in our experiment: specific and general. The "specific" task refers to searching with a specific information need and the "general" task refers to searching with a general information need. Table 1 shows the 10 search tasks (5 for each type) used in the experiment.

Table 1. Search tasks in the experiment

Task Type	Search Goal
Specific	AUDI R8
	The map of Beijing subway
	Scientist Thomas Edison
	The Eiffel Tower
	The emblem of the Beijing Olympic Games
General	A large crowd of people
	A quiet street
	Funny animals
	The England during the Industrial Revolution
	The image reflecting team work

The SERPs returned by the search engine typically consisted of 15 search results, which were presented in a grid of 3 rows × 5 columns. For each result, a thumbnail image and some meta data (filename and resolution of the original image) were presented. Fig. 1 shows an example. We defined each result as

an area of interest (AOI) and labeled them according to the common Western reading habit (from left to right and top to bottom). The query words were predefined, and all the SERPs were cashed beforehand to make sure that all participants in the same results presentation order (detailed below) would see exactly the same SERPs.

Fig. 1. An example of the SERPs and the defined AOIs

Pan et al. (Pan et al. 2007) reported that in Web text search, college students were heavily influenced by the order of search results. To investigate if this behavior still exists in Web image search, we manipulated the results presentation order. There were two orders (conditions): normal and reversed. In "normal" condition, search results were presented in the original order as the search engine returned. In "reversed" condition, search results were presented in a reversed order in which the 1st result was swapped with the 15th result, the 2nd result was swapped with the 14th result, and so on. The manipulation was undetectable to the participants.

3.2 Participants

Sixty-five undergraduate or graduate participants were recruited for this experiment through bulletin boards of a university campus. Among them, 7 participants were excluded due to calibration problems or low quality of data, leaving us with 58 participants (30 males and 28 females). These 58 participants were from a variety of disciplines with an age range from 18 to 28 (the mean was 22.7 and the standard deviation was 2.7). Participants reported that they were familiar or moderate familiar with Web image search (one female participant reported that she was unfamiliar with Web image search), and most of them

(47 out of 58) searched the Web for images at least once a day. Half of the participants (15 males and 14 females) were assigned to "normal" condition and the rest were assigned to "reversed" condition.

3.3 Apparatus

Eye-tracking was performed using a Tobii T120 eye-tracker which has a data rate of 120 Hz, a visual angle accuracy of 0.5°, and a 17" TFT monitor. We set the screen resolution to be 1024×768 pixels and used the browser Internet Explorer 6 in full screen mode. With this setting, all the 15 images were above the page fold so that they could be seen by the participants in the first place. Logging of click and gaze data was done by the software Tobii Studio.

3.4 Procedure

At the beginning of the experiment, the eye-tracker was calibrated for each participant using 9-point calibration. Then, participants started with a practice to get familiar with the procedure. After the practice we recalibrated the eye-tracker and let participants continue for the 10 search tasks. For each task, a text description of task goal was shown on the screen. Participants read the description and clicked on a "start" button when they were ready, which would triggered the prepared SERP to show up. A task was considered completed when the participant click on one of the search results that he/she thought to be most satisfied the task goal. After completing all the tasks, participants answered a questionnaire about their experiences in the study (see Section 4.3 for details), and provided demographic information. At the end of the experiment, participants were paid for their efforts. The experiment took about 30 minutes per participant.

4 Results

Gaze data derived from eye-trackers are commonly analyzed in form of fixations. In our study, a fixation is defined with a minimum of 100 ms and a radius of 50 pixels. From the 58 participants, we obtained usable eye-tracking data for 576 of the 580 tasks. Some tasks were filtered out because of incomplete data or misoperations by mistake.

4.1 General Characteristics

Before describing the influence of the two factors, we present some general characteristics of user behavior across all the search tasks and conditions. On average, participants took about 6 seconds and viewed 6 to 7 search results to finish a search task. Fig. 2 shows the fixation count, time to first fixation, and click rate of each search result, which indicate participants' gaze distribution, viewing order, and their final choice (i.e., clicks) respectively.

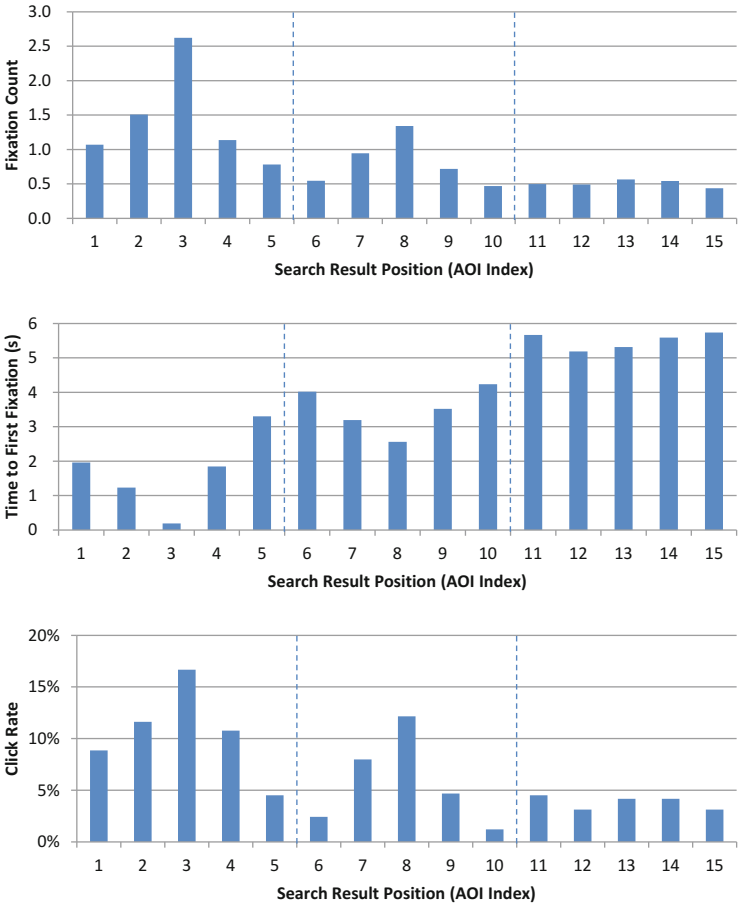

Fig. 2. The fixation count (*top*), time to first fixation (*middle*), and click rate (*bottom*) of each search result in general. *Dash lines* denote gaps between two rows.

As shown in the figure, search results in the first row (results 1 to 5) and the middle of the second row (results 7 to 9) received more attention than others, and these results were viewed earlier. More specifically, the first and most viewed result was the result in the middle of the first row (result 3), then the results adjacent to it (results 2 and 3) and the leftmost result in the first row (result 1) were viewed. After that, participants viewed the results in the middle of the second row (results 7 to 9) and the rightmost result in the first row (result 5). The rest results were viewed last and least. Besides, participants were more likely to click on search results in the first row (results 1 to 4) and the middle of the second row (results 7 and 8). Interestingly, results 5 and 6 received less attention, had longer arrival time, and were less selected than result 8. This indicates that for Web image search, "center-center" results are more important than "top-right"

results and "center-left" results, which is against the left-to-right and top-down reading habit. As participants viewed 6 to 7 results on average, we can conclude that search results in "top-left", "top-center", and "center-center" areas were viewed early and were the most viewed and selected ones.

4.2 Influence of *Task Type* and *Result Order*

To investigate the influence of the two factors, we analyzed the following measures: task completing time, number of search results viewed, number of regression, total fixation duration, total fixation count, average fixation duration (the average duration of a single fixation), and pupil diameter. For these measures, we performed the repeated measures ANOVA test (*Task Type* × *Result Order*).

For *Task Type*, we found significant main effects for number of search results viewed, number of regression, and pupil diameter (Table 2). Participants in "general" tasks viewed more results, made more regressions, and had larger pupil diameters. This indicates that participants made more efforts completing "general" tasks than "specific" tasks. For *Result Order*, we found no main effect for any of the measures. And there was no significant interaction between *Task Type* and *Result Order*.

Table 2. The main effects of *Task Type*

Measure	Specific	General	F(1,56)	Significance
Number of Search Results Viewed	6.04	7.13	52.11	$p \ll 0.001$
Number of Regression	2.47	3.29	9.72	$p < 0.01$
Pupil Diameter (mm)	3.28	3.33	68.50	$p \ll 0.001$

Fig. 3 shows the influence of the two factors on participants' gaze distribution. For each result, we compared the fixation count using ANOVA tests. For *Task Type*, participants paid significantly more attention on top results in "specific" tasks. On the contrary, they paid significantly more attention on results at lower positions in "general" tasks. This indicates that results at lower positions have better chance of being viewed in "general" search tasks. For *Result Order*, results at lower positions received relatively more attention in "reversed" condition than in "normal" condition. However, neither *Task Type* nor *Result Order* changed the "big picture". Results in "top-left", "top-center", and "center-center" areas received more attention than others.

Both *Task Type* and *Result Order* did not have significant influence on the viewing order (measured by time to first fixation, graphs are not presented due to page limitations). As described in Section 4.1, participants tended to firstly view the "top-center" results, then the "top-left" results, the "center-center" results, and other results.

The influence of the two factors on click rate is shown in Fig. 4. The chi-square analysis were used to compare them. For *Task Type*, participants in "general" tasks clicked on top results less and were more likely to click on results at lower

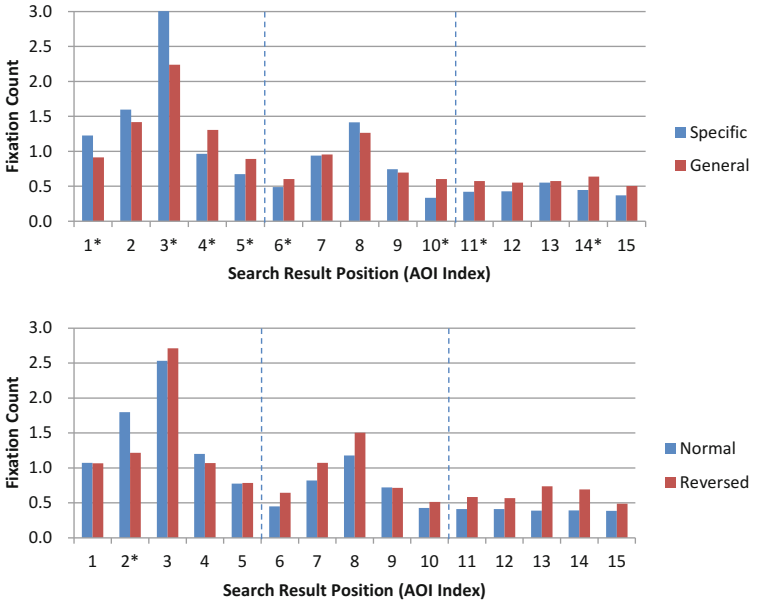

Fig. 3. The influence of *Task Type* (*top*) and *Result Order* (*bottom*) on gaze distribution. *Dash lines* denote gaps between two rows and ∗ denotes a significant difference.

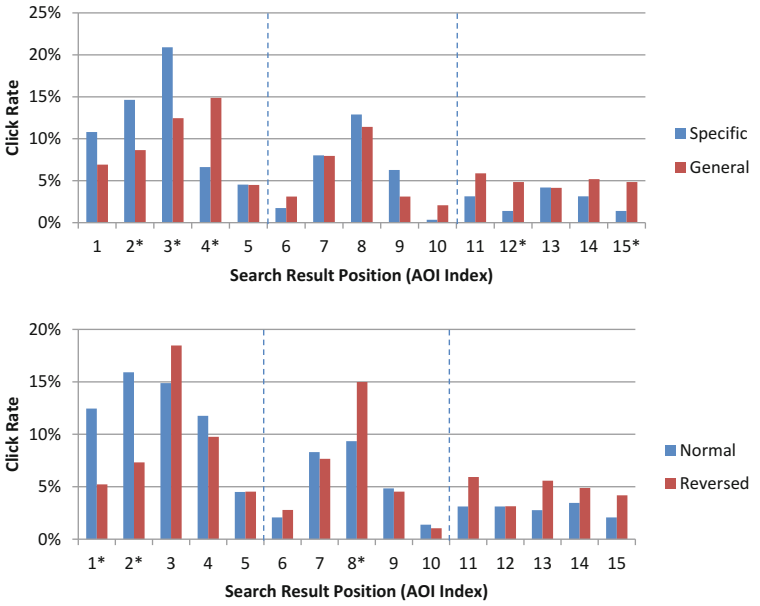

Fig. 4. The influence of *Task Type* (*top*) and *Result Order* (*bottom*) on click rate. *Dash lines* denote gaps between two rows and ∗ denotes a significant difference.

positions. For *Result Order*, participants clicked on top results significantly more often in "normal" condition. In "reversed" condition, they were more likely to click on results at lower positions, especially the result in the middle of the second row (result 8). Again, the "big picture" did not change. Results in "top-center", "top-left", and "center-center" areas were the most selected ones.

4.3 Post-experimental Questionnaire

In the questionnaire, participants rated their satisfaction with the search results in each SERP (the page-level satisfaction) with a 5-point Likert scale (1 means "completely dissatisfied" and 5 means "completely satisfied"), and wrote down why they selected the particular search result in each task. During this process, participants could review the SERPs and check their selections via log data.

On average, participants were satisfied by the search results with a mean of 4.49 (the standard deviation was 0.75). ANOVA tests showed a significant main effect of *Task Type*, $F(1,56)=18.26$, $p \ll 0.001$. Participants were more satisfied in "specific" tasks (4.61) than in "general" tasks (4.36). These was no main effect of *Result Order*, and no significant interaction between *Task Type* and *Result Order*.

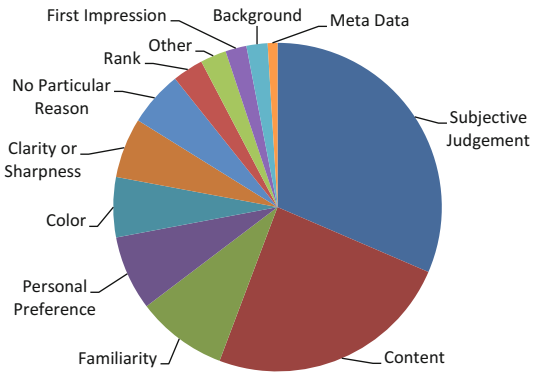

Fig. 5. Categories of the reason why a particular search result was selected

We manually analyzed the reason of participants' selections and grouped them into categories (Fig. 5). About one-third of the answers were subjective judgements, which contained subjective or emotional expressions. For instance, a participant wrote that the result being chosen was "more real". A quarter of the answers referred to the content of the image. Some participants' selections (9%) were based on the familiarity of the image, e.g., "I've seen it (the image) in my history books". There were also a proportion of answers concerning personal preferences (7%), colors (6%), and clarity or sharpness of the image (6%). In 5% of the cases, participants randomly clicked on a result with no particular reason.

The rest of the answers fell into categories such as "rank", "first impression", "background", "meta data", and "others".

5 Discussion and Design Implications

The experiment described above investigate user behavior in Web image search, including general characteristics, the influence of search task type and results presentation order, and the reason of search result selections.

In general, users paid more attention on results in "top-left", "top-center", and "center-center" areas, and were more likely to click on them. When considering the viewing sequence, users viewed the "top-center" area first, the "top left" area second, and the "center-center" area third. Other areas were viewed last and least. It is interesting that results in "top-right" and "center-left" areas were less important than results in "center-center" area, which is against the left-to-right and top-down reading habit. This indicates that in Web image search, users do not follow the common Western reading habit, which is different from Web text search. In addition, the distributions of fixations and clicks were similar in Web image search (Fig. 2). The most clicked results were the most viewed ones. But in Web text search, people tent to click on the first result substantially more often though they spent a non-negligible proportion of attention on other results (Granka et al. 2004, Pan et al. 2007)

For the two factors, we found that the *Task Type* has a strong influence on user behavior but the *Result Order* doesn't. Although on average users spent equal time completing "general" and "specific" search tasks, they were more "busy" in "general" tasks: they viewed more results, made more regressions, and had larger pupil diameters. As pupil diameters can reflect arousal, the larger pupil diameter indicates that users need to pay more cognitive efforts in "general" tasks. We think this is the reason why users rated the satisfaction of "general" tasks lower than "specific" tasks. Pan et al. (Pan et al. 2007) reported that in Web text search users tend to click on top ranked results even if the results presentation order is reversed. In our experiment, there is a similar trend. In "normal" condition, users tended to click on the "top-left", "top-center", and "center-center" results. In "reversed" condition, users still clicked on these results more than others. This can be explained in two ways: (1) users have a "position preference" in which they prefer search results at certain locations; and (2) the results in the page are all highly relative to the search goal so that in "reversed" condition the top results still satisfy the need, which leads to the insignificant influence. Further studies are needed to investigate this phenomenon by varying the relevance of the top results to see if this preference still exists.

In addition, we looked at the reason why users selected a particular search result. It seems that the selections were based on their subjective or own judgements and image contents. Only a small proportion of users stated that the rank and meta data were useful and were considered. However, this should be interpreted with cautions, as it is the self-reported statement.

Our results suggest that the rank of search results in Web image search should not be left-to-right and top-to-bottom. For example, the "rank 1" result should

be the "top-center" result (result 3), and the "center-center" result (result 8) should be considered more important than "top-right" and "center-left" results. Our results also suggest that there is a closer alignment between user's attention (fixations) and decision (clicks) in Web image search, which means that using clicks as implicit feedbacks may be more accurate in Web image search than in Web text search.

6 Conclusion

We presented a study of user behavior in Web image search using eye-tracking. In particular, we looked at the general characteristics of user behavior and the influence of two factors: search task type and results presentation order. We found that results located at "top-left", "top-center", and "center-center" areas were more important, and the search task type had a strong influence on user behavior but the results presentation order didn't. We also investigated the reason of search result selections which fell into a number of categories.

Our study is only a prelude to studies in Web image search using eye-tracking. There are lots of questions that need to be answered such as the gender differences and individual differences as in Web text search (Lorigo et al. 2006, Dumais et al. 2010). Analysis of these questions will be the topic of future work.

Acknowledgments. This work was supported in part by the Natural Science Foundation of China under Grant No.61375044, the specialized Research Fund for the Doctoral Program of Chinese Higher Education under Grant No. 20121101110035, and the Specialized Fund for Joint Building Program of Beijing Municipal Education Commission. We would like to thank Prof. Shengfu Lu and Dr. Mi Li of International WIC Institute for their advices and all the participants for their efforts.

References

1. Batley, S.: Visual Information Retrieval: Browsing Strategies in Pictorial Databases. In: 12th International Online Information Meeting, pp. 373–381. Learned Information Ltd., Oxford (1988)
2. Buscher, G., Dumais, S., Cutrell, E.: The Good, the Bad, and the Random: An Eye-Tracking Study of Ad Quality in Web Search. In: 33rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 42–49. ACM, New York (2010)
3. Choi, Y.: Effects of Contextual Factors on Image Searching on the Web. Journal of the American Society for Information Science and Technology 61(10), 2011–2028 (2010)
4. Choi, Y., Rasmussen, E.M.: Searching for Images: the Analysis of Users' Queries for Image Retrieval in American History. Journal of the American Society for Information Science and Technology 54(6), 498–511 (2003)

5. Cunningham, S.J., Masoodian, M.: Looking for a Picture: an Analysis of Everyday Image Information Searching. In: 6th ACM/IEEE-CS Joint Conference on Digital Libraries, pp. 198–199. IEEE, New York (2006)
6. Cutrell, E., Guan, Z.: What Are You Looking For? An Eye-tracking Study of Information Usage in Web Search. In: 25th SIGCHI Conference on Human Factors in Computing Systems, pp. 407–416. ACM, New York (2007)
7. Dumais, S., Buscher, G., Cutrell, E.: Individual Differences in Gaze Patterns for Web Search. In: 3rd Symposium on Information Interaction in Context, pp. 185–194. ACM, New York (2010)
8. González-Caro, C., Marcos, M.C.: Different Users and Intents: An Eye-tracking Analysis of Web Search. In: 4th ACM International Conference on Web Search and Data Mining, pp. 9–12. ACM, New York (2011)
9. Goodrum, A., Spink, A.: Image Searching on the Excite Web Search Engine. Information Processing and Management 37(2), 295–311 (2001)
10. Granka, L.A., Joachims, T., Gay, G.: Eye-Tracking Analysis of User Behavior in WWW Search. In: 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 478–479. ACM, New York (2004)
11. Guan, Z., Cutrell, E.: An Eye Tracking Study of the Effect of Target Rank on Web Search. In: 25th SIGCHI Conference on Human Factors in Computing Systems, pp. 417–420. ACM, New York (2007)
12. Hsieh-Yee, I.: Research on Web Search Behavior. Library and Information Science Research 23(2), 167–185 (2001)
13. Huang, K., Kelly, D.: The Daily Image Information Needs and Seeking Behaivor of Chinese Undergraduate Students. College and Research Libraries 74(3), 243–261 (2013)
14. Jansen, B.J., Pooch, U.: A Review of Web Searching Studies and a Framework for Future Research. Journal of the American Society for Information Science and Technology 52(3), 235–246 (2001)
15. Jiang, J., He, D., Allan, J.: Searching, Browsing, and Clicking in a Search Session: Changes in User Behavior by Task and Over Time. In: 37th Annual International ACM SIGIR Conference on Rearch and Development in Information Retrieval, pp. 607–616. ACM, New York (2014)
16. Jörgensen, C., Jörgensen, P.: Image Querying by Image Professionals. Journal of the American Society for Information Science and Technology 56(12), 1346–1359 (2005)
17. Klökner, K., Wirschum, N., Jameson, A.: Depth- and Breadth-First Processing of Search Result Lists. In: CHI 2004 Extened Abstracts on Human Factors in Computing Systems, p. 1539. ACM, New York (2004)
18. Lorigo, L., Pan, B., Hembrooke, H., Joachims, T., Granka, L., Gay, G.: The Influence of Task and Gender on Search and Evaluation Behavior Using Google. Information Processing and Management 42(4), 1123–1131 (2006)
19. Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G., Granka, L.: In Google We Trust: Users' Decisions on Rank, Position, and Relevance. Journal of Computer-Mediated Communication 12(3), 801–823 (2007)
20. Pu, H.T.: An Analysis of Web Image Queries for Search. Proceedings of the American Society for Information Science and Technology 40(1), 340–348 (2003)