

Research Article

A Picture is Worth a Thousand Keywords: Exploring Mobile Image-Based Web Searching

Konrad Tollmar,¹ Ted Möller,² and Björn Nilsved³

¹Department of Design Sciences, Lund University, Sölvegatan 26, 221 00 Lund, Sweden

²Sony Ericsson Mobile Communications, Scheelevägen 16, 221 88 Lund, Sweden

³Epineer, Stora Varvsgatan 1, 211 19 Malmö, Sweden

Correspondence should be addressed to Konrad Tollmar, konrad.tollmar@design.lth.se

Received 3 April 2008; Accepted 29 August 2008

Recommended by Regina Bernhaupt

Using images of objects as queries is a new approach to search for information on the Web. Image-based information retrieval goes beyond only matching images, as information in other modalities also can be extracted from data collections using an image search. We have developed a new system that uses images to search for web-based information. This paper has a particular focus on exploring users' experience of general mobile image-based web searches to find what issues and phenomena it contains. This was achieved in a multipart study by creating and letting respondents test prototypes of mobile image-based search systems and collect data using interviews, observations, video observations, and questionnaires. We observed that searching for information based only on visual similarity and without any assistance is sometimes difficult, especially on mobile devices with limited interaction bandwidth. Most of our subjects preferred a search tool that guides the users through the search result based on contextual information, compared to presenting the search result as a plain ranked list.

Copyright © 2008 Konrad Tollmar et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Mobile Internet applications have been in public use for a few years. However, compared with a desktop computer, mobile devices such as smart phones and PDAs with the ability to connect to the Internet differ quite considerably. First, the size of the display is small. Even if a PDA has larger display than a mobile phone, it is still far from the size of an average desktop computer. Second, text input is slower than with a full desktop keyboard and in many mobile situations attention is divided which makes text input even harder. Third, the data transfer on mobile Internet is still far slower than wired Internet. In short, handheld web environments are still severely limited, and it has been suggested that mobile Internet applications should be designed to support more task-specific uses [1].

We have developed a mobile content-based search system that matches camera-phone images against images found by web crawling servers [2, 3]. What makes this interesting is the context of wireless connectivity and inexpensive digital

imaging hardware. What makes it potentially useful is the Web's vast database of imagery for comparison against what the camera sees (see Figure 1).

Image-based searching has been the subject of much research, and systems for desktop searching of media libraries have been deployed (e.g., ImageRover [4], Webseek [5]). But for the majority of these applications, image searching has been less popular than traditional keyword search methods. It has hence been suggested that an image search could be used instead of, or as a complement to, a text-based search when users' search needs are vague, and/or text as input may be hard to define [6].

However, creating a general image-based web search engine and opening it to the general public on a larger scale might not yet be feasible. The challenges are clear when one compares a normal-sized image database of a hundred thousand images with Google's 25 billion indexed web pages, as well as 1.3 billion images [7]. While we and other researchers continue the work of refining and optimizing image-matching algorithms to match these kinds



FIGURE 1: Mobile image-based web search.

of numbers, this paper explores instead the interaction sides of image-based searches. This area is ripe for exploration as we approach the inevitable availability of the technology.

The concept of a mobile image-based web search, though, brings many questions to life. How would users perceive such a search system and what would they use it for? Would mobile availability change users' web searching behavior? How will mobile users interact with such a system? What could its interface look like?

The purpose of this paper is to explore users' experience of using image-based web searches on a mobile device [8]. The particular focus is on understanding the search process when using images of objects as queries and how that search process can be improved for the mobile case. The empirical data of this study has been collected through a multipart study by creating and letting respondents test prototypes of mobile image-based search systems and collecting data using interviews, observations, video observations, and questionnaires [9].

In Section 2, we start by exploring how and for what purpose subjects in our study might use images to search for information in mobile situations. Then we depict a user study where we test and compare two different search strategies using a prototype on a mobile device. Our findings indicate that active guidance directing that search was preferred even if it did not always cut down the search time. At the end of the paper, we discuss these findings and talk about how to apply different kinds of search strategies when searching for information on various categories of objects.

2. Related Work

Text-based image searching is the most common method. However, and also noted in our study, images contain multiple dimensions of information where it might be difficult to express some features in language or by a keyword. Just imagine the example where you simply do not know the word for a particular object, or you would like to find similar images such as images with a large proportion of red (could be a sunset), images with typical orientations (e.g., buildings), or images with typical shapes (e.g., specific objects).

The problem of searching for digital images based on content, content-based image retrieval (CBIR), in large databases has been widely researched in the last decade. "Content-based" means that the search makes use of the

contents of the images themselves, rather than relying on human-inputted metadata such as captions or keywords. A vast amount of systems based on CBIR have been developed for desktop searching of media libraries (e.g., ImageRover [4] and QBIC [10]); yet another domain is image copyright protection and sorting images in personal image databases. However, for the majority of these applications, CBIR has turned out to be less popular than traditional keyword search methods [6]. So in current approaches for very large image databases, like Google and Yahoo!, images are typically pre-labeled with keywords or matching is performed based on image captions or filenames. Image matching based on appearances of very large image databases, though, is a very active domain where we and many others continuously contribute with new methods and approaches [11].

The uses for these fixed systems are, however, rather different compared to the mobile case. A fundamental problem for visual searching is the difficulty of getting a query image that is naturally much easier when using a mobile device. More recently, work has also considered the more specific case of mobile image-based web searches using, for example, a PDA or camera phone [3, 12], while some other research has also demonstrated the utility of mobile solutions in specific domains (e.g., location-based information [13]).

Finding web-based information on a mobile device is yet another challenge. Kim and Albers study of search strategies in small displays found a high data spread across all search times and conditions and claimed that designing information differently does not play a major role for some users [14]. They suggested that future work should focus on other factors such as user motivation for finding the information, user knowledge of the types of information, and different types of information. Studies of various techniques that streamline web content to small screens also indicate that this causes confusion and that models that are more closely related to the mental model of full screen are preferred [15].

3. Mobile Images as Search Queries

In the first part of the study, ten subjects were asked to take at least five photos each during the course of 1-2 days. The group consisted of three female and seven male respondents aged 16 to 35, among these subjects were five students.

They were given instructions to use their own digital camera or mobile phone to take photos of objects or places that they were interested in learning about and write down any thoughts regarding what information they hoped to retrieve using the image (see Figure 2).

These pictures were then used as the basis of a semistructured interview. We directed the respondents to ask themselves what questions they wanted answers to and gave suggestions of the kind of information elements they might need in the result list to be able to evaluate links that appeared there. This follows the approach to improving the presentation of the search result introduced by Woodruff:

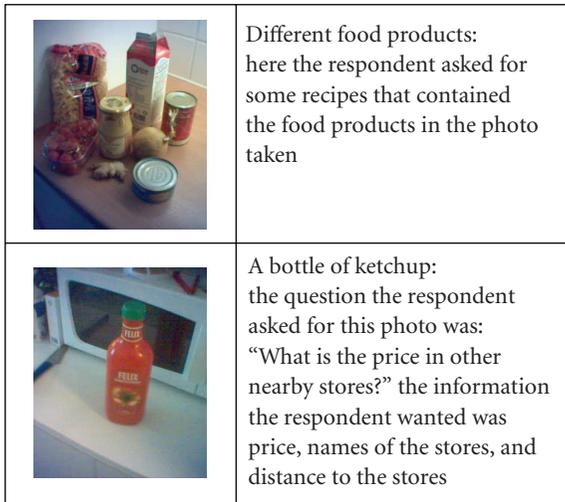


FIGURE 2: Examples of image queries.

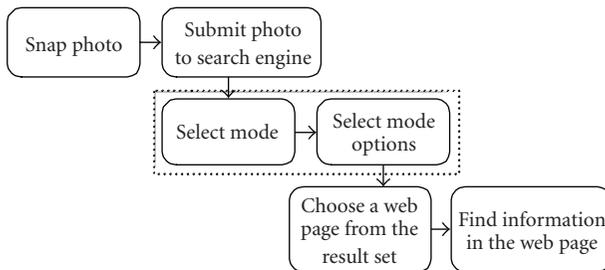


FIGURE 3: Interaction sequence of an image-based search using search modes. The two steps marked in the middle of the sequence are what differentiate it from a straight image-based search.

“The search engine can increase user efficiency by either (1) returning higher-quality document lists (e.g., through better index coverage and ranking algorithms) or (2) providing information that allows the user to evaluate the results more quickly” [16]. Based on these interviews, 50 miniscenarios—small, everyday situations—were created and clustered (see Table 1) in a similar way as in McDonald and Tait’s study of search strategies in content-based image retrieval [17]. Scenarios that were unlikely to generate a specific answer were sorted out. These patterns in search questions cannot be generalized to any larger population but they give us better insight into what search questions could arise when people image search using a mobile system.

In the interview, we also presented a list of information elements that can be presented in a result list and asked the respondents to rate them for each photo they had taken and the search question they related to that photo. Several respondents believed that for all searches the most valuable information elements in the result page would be to categorize the hits and statistics about the total number of hits. Other useful information elements were the matching pictures, the URL, and the thumbnail image of the contained web page. For the bottle of ketchup, for instance, the information element “when was the web page last updated”

was essential because prices of such products may change from day to day.

We found that there are many situations where an image is considered to be very useful to capture the interest for a particular object, such as products, places, people, and unknown objects (e.g., plants and flowers and nonlabeled artifacts). It seems like information about unmarked objects (i.e., objects without a text) and objects that are not known to the respondent are hard to find on the Web using text-based searches. By letting our subjects freely take pictures and put questions in relation to these images, we also found some other interesting issues. First, it was not always possible for the subjects to translate an image query into a simple search question that could be used on Google, for example. Second, local and location information was often contextually embedded in the image (i.e., the image expressed both a general as well as a specific question). Third, our subjects suggested that categories might be helpful in the search process (i.e., presenting some kind of category for each hyperlink in a search result would help them evaluate if the links could answer their questions). For example, shopping-related pages in the search result could be tagged in some way.

Last, we also noted that several search questions were so specific; they would probably need a very specialized service or agent to answer the question. An example of this was after taking a photo of a bunch of food products, one respondent asked for recipes for possible dishes to cook. A system might use techniques derived from semantic web, ontology, and knowledge representation to handle such questions, but they are not yet available on a larger scale and hence, at this stage, we have primarily selected to study the user interaction with mobile image-based search [18].

4. Mobile Web Search

The first part of the study provided evidence that image-based searching for information could be useful. However, as previously mentioned, related work has also provided evidence that using Internet-based services and applications on a mobile device is still problematic [1, 19].

Studies of navigation techniques for small displays suggest that mobile search tools should develop new practices that are more efficient rather than copy “traditional” web search concepts [20]. The work by Labrinidis and Rousopoulos [14] has shown that search tools need to present search material in a suitable way for the user to be able to categorize and compare data and to make conclusions and overviews. They suggest that more targeted search strategies such as focused search, preference-driven search, and metadata labeling might help users find good results even when limited by the speed and display restrictions of mobile devices. Thus the user should be provided with both detail and contextual information, which is a very challenging task considering the limited interaction given by mobile devices.

As noted, our subjects also suggested that categories might be helpful in the search process, such as shopping-related pages in the search result being tagged in some

TABLE 1: Most frequent scenarios.

Type of scenario	Number similar scenarios	Most common questions
Products, such as cars, groceries, etc.	14	Where can I buy this? What does it cost?
Places	8	What is this? Is this the place I'm looking for?
Unknown objects	7	What is this? Where to find more information?
People	5	Who is this person? Where have we met?

way. Based on this, we hypothesize that a user of a mobile device can be helped by actively narrowing the search result information. This would be particularly useful given open search queries such as images. Take for example an image of a hat; this image can naturally be connected to many different questions. A user may desire different information, for example, what kind of hat is this? Where can I find one? What famous person was seen wearing such a hat? This issue could be solved by adding a text input field to an image search for the user to further specify the information need, but it would require typing text on the limited keyboard of the mobile device.

To test this hypothesis we have designed new search strategies, or as we call them “search modes”. This concept was inspired not only by how computer wizards guide a user through a series of selections, but also by Aslandogan and Yu’s proposal to combine different query expressions and the results of different retrieval algorithms for text retrieval [21]. The basic idea of search modes is based on combining the photo search with some type of categories that help a user filter a search result. It does not require any text input: a user selects parameters in a linear series of steps by simply choosing between options generated by the system.

A search using search modes starts with the user taking an image and submitting it to the search engine. A selection of search modes is presented and the user selects a mode, for example, “Shopping mode” that filters search results based on certain preferences, such as location and price. The search modes then present additional options for that mode, for example, availability and open hours of the store. At the completion of the wizard, a filtered search result list is displayed. A schematic model of how search modes could work can be found in Figure 3.

Special search modes can be created to support many different types of image searches (see Figure 4). An example would be a Shopping mode that filters prices and places to shop. This would be appropriate for the ketchup search that one respondent wanted to carry out.

5. User Study

The purpose of this paper has been to explore users’ experience of general mobile image-based web searches.

We will now highlight in particular one part of the study where we tested and compared two different search strategies that would facilitate both general and more specific image-based searches in mobile situations. In this case we used two objects, one unlabeled glass tray and a CD (see Figure 5).

5.1. Prototype

Almost all respondents in the first part of the study gave examples of search questions related to local shopping, like those concerning the ketchup bottle (Table 1). This indicates that at least some of users’ information needs adhere to local shopping. Shoppers using mobile devices have also earlier been identified as a plausible area for mobile application. One central need found in a study of grocery shopping was to assist the user in finding products [22]. Based on these observations, we designed a prototype and implemented a Shopping mode search mode, as well as a general search technique that does not use mode, that is, Nonmode (see Figure 6).

(A) “Nonmode”

Search for information only based on visual similarity. (1) The user takes a picture and (2) chooses to view all matches, and (3) the search result is presented as a ranked link with links and text summaries (e.g., <http://www.google.com/>).

(B) “Shopping Mode”

Search for information based on visual similarity and filter the search result based on price information and location. (1) The user takes a picture and (2) chooses to use Shopping mode, (3) the search is narrowed through a couple of contextual questions, and (4) the filtered search result is presented as a ranked list with links and text summaries.

For the prototypes used in the user study, some design elements from the semistructured interview, described above, were considered. For example, respondents thought some information elements in the search list were more critical than others. They were URL, snippet (e.g., Google search results), amount of hits, and category. Respondents also thought that it would be useful to have the matched

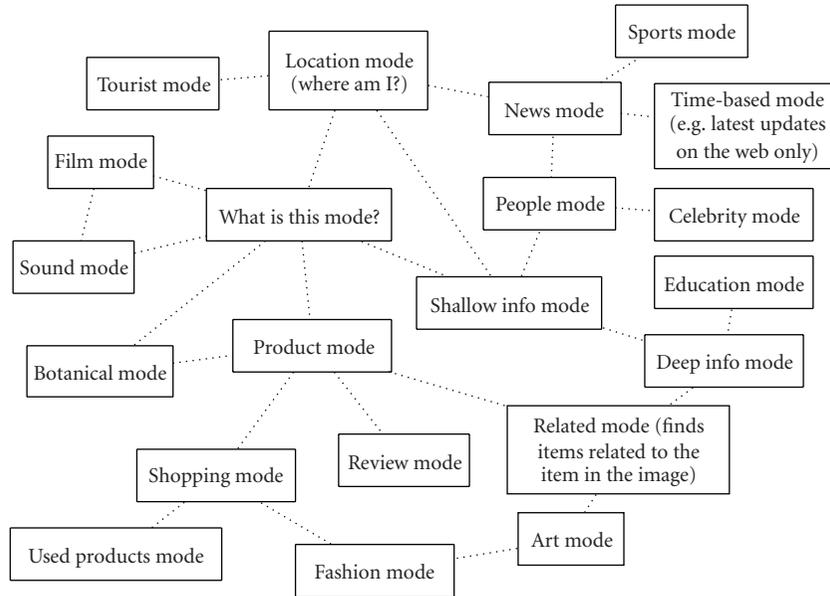


FIGURE 4: Suggestions for search modes. The edges in the graph give an idea of which search modes were inspired by each other.



FIGURE 5: Objects used in the test.

image in the search list so that they could see if the system matched images correctly. These information elements were implemented to see if they actually could help users find the information sought.

These two search techniques were implemented on a PDA with a camera and Wifi card. All search material was precomputed with an error rate of approximately 30% for the CD (one of the shopping items) and 50% for the glass tray (the second shopping item), that is, the number of items in the search result that were not correct. The error rate is higher than our current image matching algorithms but given other search factors and errors, we claim that this provides a realistic test situation [11]. To create search result lists for the prototype, we created an automatic prototype generator written in PHP. After supplying it with parameters and options, it creates a set of static HTML files for both Nonmode and Shopping mode and these files are then stored on to the handheld device to be viewed in its browser. The prototype generator creates the search result list by

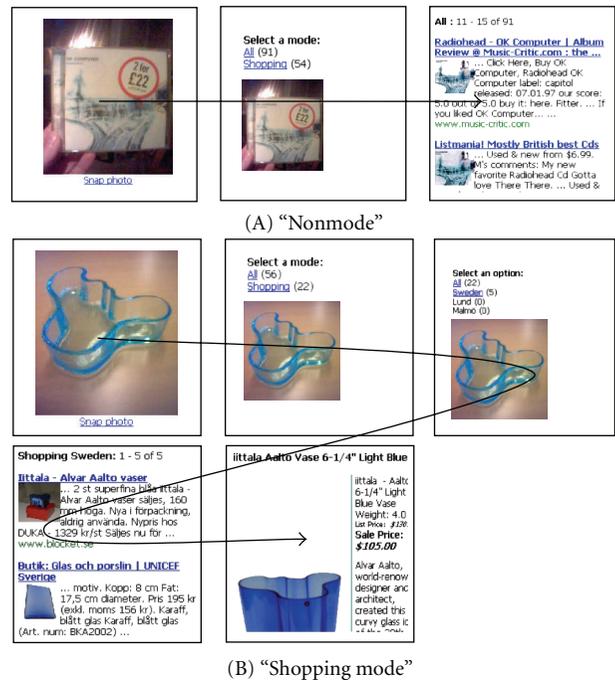


FIGURE 6: Screen shoots from the prototype used in the test.

connecting to Google’s text search functionality via their API using the SOAP protocol.

Method

A user study with ten participants was performed to study the two search modes. The participants who took part in the experiments were all students at Lund University.

As McDonald et al. point out [6], the use of students as experiment respondents is justified because the system was intended for use by a wide population of general users which will certainly include students. The group consisted of four female and six male respondents aged 23 to 26. However, we also realized that using only 10 respondents makes it hard to statistically generalize our results but as mentioned before, this was not the primary goal at this stage of the project.

Half of the subjects used the general “Nonmode” and the second half the “Shopping mode”. The subjects were randomized. The test not only was video recorded but also included a questionnaire and a one-on-one discussion. The subjects were given two different kinds of objects, a labeled CD record and a nonlabeled glass tray (see Figure 6). The following procedure was used:

- (1) test the device and try the device web application;
- (2) take a picture of the object;
- (3) search for information using either the “Nonmode” (five subjects) or the “Shopping mode” (five subjects);
- (4) fill out a questionnaire, and round-off with a short discussion.

Two quantities factors were measured in the test: (1) time to find relevant information and (2) the number of pages that were visited before the subject was satisfied with the search. In the questionnaire and the discussion, we also asked them to compare their experience of using image-based searching with their previous experience from text-based searching.

5.2. Results

5.2.1. Measured Time

The time it took for the two groups of participants to find the requested information using the prototypes showed a significant difference (see Figure 7). All respondents found the information sought about the glass tray. In general, the respondents in the group using Shopping mode (B) seemed to take a longer time finishing the task than those using Nonmode (A).

The fastest search time was just above 2 minutes and the slowest just above 8 minutes. The average time it took for both groups to finish the task was about five and a half minutes. For the (A) group the average time was about four minutes, and for the (B) group the average time was just under seven minutes.

5.2.2. Experienced Time

An interesting aspect about these results is to compare the times it took respondents to complete the search task. The group using Shopping mode (B) took significantly longer time to finish the task and yet they tended to rate the image search as equally easy to use as text search.

The time difference between the two groups can to some extent be explained by the group using Shopping mode (B)

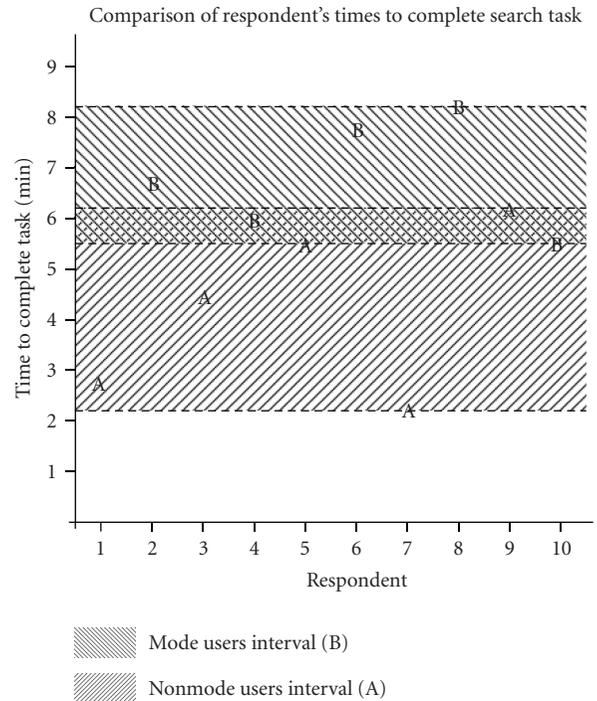


FIGURE 7: Diagram showing how long it took the respondents to find the information sought. Group A used “Nonmode”, while Group B used “Shopping Mode”.

being able to change mode setting compared to group (A) that just had one list to go through. The members of the group using Shopping mode (B) first looked through the results in one mode (e.g., shopping in Sweden) and then switched to another mode (e.g., shopping in Lund) to start all over again. This could account for some of the difference in time between the two groups. However it should also be noted that comparing the time to complete between the tasks is difficult in this small test sample.

5.3. Observations

5.3.1. The Mobile Device

In general, all the subjects tended to handle the PDA with more ease than expected (all were novel PDA users—but experienced mobile phone users). Two of the respondents thought that they had some problems with the device, they were not used to it or they thought that it limited them a bit. Yet another respondent thought that technical features of the device could limit good search results. Particular scrolling through long list of search results could be very time consuming.

“Its hard to find specific information, the screen is very small and limiting.”

5.3.2. Searching with Text Queries and Image Queries

Most respondents indicated that image-based searching for unlabeled objects could be very useful. One respondent commented as follows.

“It would have been difficult to find the unlabeled object without being able to image-based search. The alternative, talking to someone, going to a shop, would be very time consuming.”

However, that lack of refining an image-based search compared to a text-based search could be a hinder:

“It’s harder to modify an image-based search compared to a text-based search.”

Moreover, three of the respondents believed that an image had to be of good quality to get sufficient matches. The conception was that a bad image is like a misspelled text query or a bad query. The opposite was also true, the users thought that the image was sufficient, but if used in a real system (as compared opposed to the prototype used in the test), it would present bad matches.

5.3.3. Verifying

The image was found to be useful to verify the search results. Most of the respondents used the thumbnail of the matched image as a way to verify the hits of the web pages. One respondent expressed that

“searching with an image might take some longer time but once you get the search results it makes you more confident that you really found the right information ... its very easy to verify that its really the right object ... when I search with text I am sometimes uncertain that I found the right information.”

However, a couple of the respondents did not seem to reflect upon the fact that the image was just taken seconds ago. Instead, their focus seemed to be on the information to be found.

5.4. Questionnaire

For each respondent, an interview was set up after the test was conducted. The questions asked regarded the respondent’s perception of the utility of the prototype. We had observed some usability issues in the test and wanted to verify how this affected the perceived value of different kinds of image-based searches. We asked “how well do you think the system helped you find information” (1) about unknown objects in the vicinity, (2) about objects of the same type as one in the vicinity, (3) about known objects, for example, labeled with logotype or text, (4) about shopping information, for a specific item, (5) within a reasonable time to complete the task?

To get a better overview of the results, the data collected in the questionnaires is arranged as rating diagrams (see

Figure 8). In these diagrams, each arrow represents one respondent’s answer to a question in the questionnaire. The arrow starts at the respondent’s rating of text-based search systems and ends at the rating of image-based systems. This means that an arrow pointing upward indicates that the respondent rated the image-based search higher than the text-based one for that question and vice versa. The length of the arrow indicates how much the rating differed. In addition, the respondents are sorted by group with the results of the group using the Nonmode to the left and the group using Shopping mode to the right. We decided to present the data in this way because the groups are too small to properly analyze statistically.

5.4.1. Question 1: Unknown Objects in the Vicinity

All respondents that used Nonmode (A) rated the text-based search low. The same group of respondents rated the image-based search high. Group (B) who used the Shopping mode prototype was not as homogenous. Three of the five were more positive to the image-based search than the text-based one, and the other two rated the image-based search lower than the text-based one but their difference in rating is just one step on the scale. Overall, eight out of ten were very positive to image-based searching compared to text-based searching when it came to an object in the vicinity.

5.4.2. Question 2: Objects of the Same Type as One in the Vicinity

Six out of ten were more positive about using the image-based search compared to the text-based search when it came to searching for information about objects of the same type as one in a user’s vicinity. However, this is not as significant as in the first question.

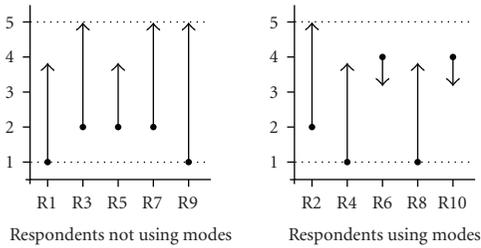
5.4.3. Question 3: Known Objects

None of the respondents believed that the image-based search would satisfy their needs better than the text-based search when it comes to searching for information about known objects such as objects marked with text. The ratings could be an indication that the error rates for the prototypes were too high, or at least higher than if one knew the name of an object and used a text-based search. It is easier to match words and therefore it is more likely that the hits would be more accurate.

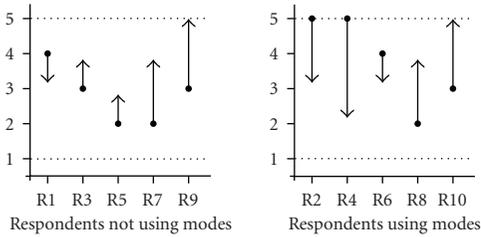
5.4.4. Question 4: Shopping Information

All respondents tended to rate both text- and image-based search systems quite high when it comes to finding shopping-related information. This could be because the web generally contains such information. Six out of ten respondents thought that both systems satisfied their needs equally when it came to finding prices and places to buy an object. In the

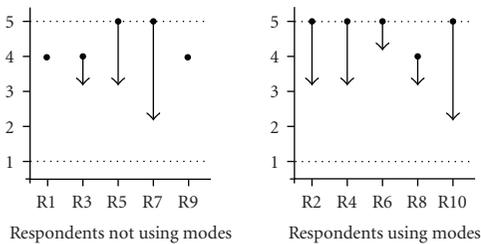
Question 1: Respondent's rating of how well text and imagebased search systems can help finding information about objects in the vicinity



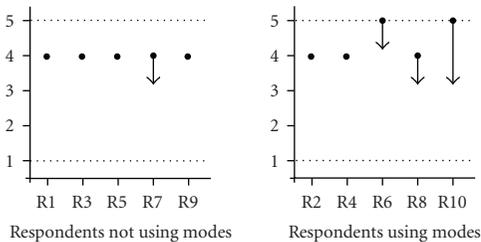
Question 2: Respondent's rating of how well text and imagebased search systems can help finding information about objects of the same type as one in the vicinity



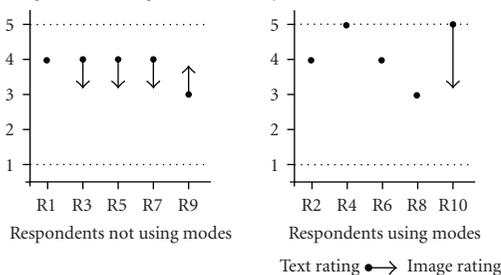
Question 3: Respondent's rating of how well text and imagebased search systems can help finding information about known objects



Question 4: Respondent's rating of how well text and imagebased search systems can help finding price information and places to buy an object



Question 5: Respondent's rating of how fast information can be found using text- and image-based search systems



Text rating ● → Image rating

FIGURE 8: Rating diagrams from questionnaires.

group that use the Nonmode prototype, the consensus of equally good systems was more significant.

5.4.5. Question 5: Experienced Time to Complete Task

Five out of ten respondents think that image and text searches would need the same amount of time to find the information sought, and the others think it would not be much different. An interesting aspect about these results is when you compare them to the times it took respondents to complete the search task. The group that use Nonmode tended to rate the text-based search as a faster way of searching than the image-search. The group using Shopping modes took a significantly longer time to finish the task and yet they tended to rate the image-search as equal to the text-search. Perhaps the search modes (such as the Shopping mode) make users feel more time efficient.

5.5. Summary

We found that image-based searching as such is very satisfying and enjoyable to use. The results indicate that using an image-based search on unknown, unmarked objects is more rewarding than using a text-based search. With marked objects, it seems that users prefer text-based searches to some extent. In conclusion: the respondents using Shopping mode spent more time on the task but still felt that image-searching was a little more time efficient.

A couple of respondents pointed out that the use of the Shopping modes resulted in less workload. Based on the observations, it also seems that the users made fewer decisions when using the Shopping modes. Five out of ten respondents thought that the two search methods would need the same amount of time to find the sought information, and the others thought there would not be much difference. If an image search is more suitable in a mobile situation because it demands less attention, a tradeoff of a more time-demanding search process is probably acceptable to the users. Future research should evaluate how many steps are appropriate: is there a tradeoff between perfect filtering using many steps and the workload of evaluating hits of a large unfiltered set? Some observations indicate that even a more precise Shopping mode that just returns hits of prices could be appropriate, but this would call for even more mode options steps to choose from.

6. Discussion

6.1. Text versus Image-Based Search

This study showed that the search performance was dependent on the search type. We agree with McDonald et al. findings that in some cases it is more effective and faster to use a text-based search engine [1]. However, our findings indicate that image-based searching is more valuable when the search is about an unknown object. Future studies could

examine in what contexts and what classes of objects people prefer a certain search strategy.

It has been confirmed in previous studies of image-based searches that users' search needs may be less precise or even vague [17]. Users may be looking for an image to illustrate a general theme, not exactly the object in front of them. This suggests that in some cases a less detailed image would be better to search with. During the construction of the prototype we thought of a slide bar that could iconize the image to retrieve a more conceptual image. This could be three buttons as well, one at a detailed level, one iconized, and one in between. We suggest future research to study this subject, and to study the possibility of letting the user disconnect the color algorithm in the system and only go for grayscale, or let the user define priorities among shape, texture, color, and size.

6.2. Mode-Based Search

The respondents using Shopping mode spent more time on the task but still felt that image-search was a little more time efficient than text-based search. However, the Shopping mode introduces extra steps to the process and the interface in our prototypes does not communicate how to navigate back up through those levels very well. Some of our respondents got stuck browsing a level that they might have left earlier if the interface would have given better hints on how to do this.

Special search modes can, however, be created to support many different types of image searches. They can either be based on algorithms filtering on image features, on textual information, on manually created directories, or a combination of these. Investigating the design and usability of special search modes is definitely still a hot candidate for future research. For example, one respondent requested a language mode. One can realize the demand for this since a considerable amount of the web's content is in different languages. Would the systems be localized based on languages, or have a functionality of choosing languages? Google is using a language and a country filter that filters the results based on the language the text is written in, or the country where the computer is connected to the Internet. Would a language mode be a necessary step as a mode, or should the system always provide the ability to filter more as Google does? This could be done automatically if there are personalized preferences in the system.

6.3. Mobile Images

In the context of everyday life, we believe that a good image is a subjective judgment. Applying this to the image matching system, a user might believe that an image is of bad quality, but it may be good enough for the system to find and match other images. The opposite can also be true: the user thinks the image is sufficient, but the system cannot handle it, or presents bad matches. Factors affecting image quality include the camera (lens quality, resolution, etc.) and the image-matching algorithm. The most interesting is how users perceive what a good image is. When is an image good

enough? Is there a need for trial and error? It would be interesting to study what the general population believes to be a good image. This could be very useful when optimizing and studying the algorithms for image matching.

One issue when taking a photo is scale. Let us say that you are in Paris and take an image of the Eiffel Tower, and then you get a lot of matched images depicting souvenirs modeled from the tower. As we found, some of the respondents estimated the size of the tray object and then compared this with the printed size information on websites. A suggestion for future work is to examine if and how a system can measure the size of an object when taking the picture. Another issue that was revealed is the necessity to define the color of an object. Let us say that the user wants a tray in another color, but the system retrieves only trays of the same color as the one in the image.

6.4. Mobile Use (Interaction)

In general, the respondents seem to think that text-based searching is more accurate and of more use than image-based searching, but that it is faster to snap an image than it is to enter a text query. It is interesting to compare this to the answers in the questionnaires where the text-based search was experienced as a little bit faster. There could be an issue of what we are measuring here, is it the time it takes to enter the query or is it the time it takes to find the sought information?

A problem with several search engines today is that high-ranked pages seldom contain such specific information. If a web page that contains the sought information is not ranked high, a user has to browse several result pages before finding it (such as software for an appliance or opening hours for a particular store). Here we think that image-based search modes could make a difference. An assumption is that search modes where the resulting web pages probably have a high ranking, like the Shopping mode, are not as rewarding as search modes where the sought information can only be found in low-ranked web pages such as personal ones.

Finally, as respondents had some difficulties in navigating (scrolling horizontally and vertically) on a web page, a suggestion that may ease this is to let the browser automatically focus on the part of the web page where the matching image is located. A proposal for future research is to see if this actually helps users to find needed information faster.

7. Summary

This paper describes a study of a novel approach to searching for information on the Web using images of objects as queries. The study was divided into two parts. In an exploratory phase, we observed how people use images as information queries. We found that there are many situations where a picture is considered to be very useful, especially when it comes to products, places, people, and unknown objects (e.g., plants and flowers and nonlabeled artifacts). With marked objects it seems that users prefer text-based searches to some extent. Based on this outcome we designed and evaluated two approaches for image-based searching in a small user study (10 participants). We observed that

searching for information based only on visual similarity and without any assistance is at times difficult, especially on mobile devices. We found that respondents felt that image-based searching was more time efficient even if the actual time to complete the tasks varied significantly and that a search tool that guides users through the search result has less workload compared to presenting the outcome of the search as a plain ranked list. This approach might be especially useful in browsing a large result set of low-ranked web content on a mobile device. Our study also provided many suggestions for an image-based search system in order to match the efficiency of text-based search methods.

Acknowledgment

The authors are very grateful to their colleagues Tom Yeh and Trevor Darrell at MIT where this project once started and with whom they share ideas and inspiration.

References

- [1] Y. Maarek, A. Soffer, and B. Chang, Working Notes of the WWW2002 Workshop On Mobile Search, 2002.
- [2] K. Tollmar, T. Yeh, and T. Darrell, "IDeixis—searching the web with mobile images for location-based information," in *Proceedings of the 6th International Symposium on Mobile Human-Computer Interaction (MobileHCI '04)*, vol. 2, pp. 288–299, Glasgow, UK, September 2004.
- [3] T. Yeh, K. Tollmar, K. Grauman, and T. Darrell, "A picture is worth a thousand keywords: image-based object search on a mobile platform," in *Proceedings of CHI Extended Abstracts on Human Factors in Computing Systems (CHI '05)*, pp. 2025–2028, Portland, Ore, USA, April 2005.
- [4] S. Sclaroff, L. Taycher, and M. La Cascia, "ImageRover: a content-based image browser for the World Wide Web," in *Proceedings of IEEE Workshop on Content-Based Access of Image and Video Libraries*, pp. 2–9, San Juan, Puerto Rico, USA, June 1997.
- [5] J. R. Smith and S.-F. Chang, "Image and video search engine for the World Wide Web," in *Storage and Retrieval for Image and Video Databases V*, vol. 3022 of *Proceedings of SPIE*, pp. 84–95, San Jose, Calif, USA, February 1997.
- [6] S. McDonald, T. Lai, and J. Tait, "Evaluating a content based image retrieval system," in *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '01)*, pp. 232–240, New Orleans, La, USA, September 2001.
- [7] C. Biever, "See it, snap it, and let your phone find it on the web," *New Scientist*, pp. 20–21, 2006.
- [8] J. Fleming and R. Koman, Eds., *Navigation: Designing the User Experience*, O'Reilly, Sebastopol, Calif, USA, 1998.
- [9] M. B. Miles and A. M. Huberman, *Qualitative Data Analysis*, New Bury, Calif, USA, Sage, 1994.
- [10] C. W. Niblack, R. Barber, W. Equitz, et al., "QBIC project: querying images by content, using color, texture, and shape," in *Storage and Retrieval for Image and Video Databases*, vol. 1908 of *Proceedings of SPIE*, pp. 173–187, San Jose, Calif, USA, February 1993.
- [11] T. Yeh, K. Tollmar, and T. Darrell, "Searching the web with mobile images for location recognition," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '04)*, vol. 2, pp. 76–81, Washington, DC, USA, June-July 2004.
- [12] X. Fan, X. Xie, Z. Li, M. Li, and W. Ma, "Photo-to-search: using multimodal queries to search the web from mobile devices," in *Proceedings of the 7th ACM SIGMM International Workshop on Multimedia Information Retrieval (MIR '05)*, pp. 143–150, Hilton, Singapore, November 2005.
- [13] N. Davies, K. Cheverst, A. Dix, and A. Hesse, "Understanding the role of image recognition in mobile tour guides," in *Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices & Services (Mobile-HCI '05)*, pp. 191–198, Salzburg, Austria, September 2005.
- [14] A. Labrinidis and N. Roussopoulos, "WebView materialization," in *Proceedings of the 19th ACM SIGMOD International Conference on Management of Data (SIGMOD '00)*, pp. 367–378, Dallas, Tex, USA, May 2000.
- [15] O. Buyukkokten, G. Molina, G. Hector, A. Paepcke, and T. Winograd, "Power browser: efficient web browsing for PDAs," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '00)*, pp. 430–437, The Hague, The Netherlands, April 2000.
- [16] A. Woodruff, A. Faulring, R. Rosenholtz, J. Morrison, and P. Pirolli, "Using thumbnails to search the web," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '01)*, pp. 198–205, Seattle, Wash, USA, March-April 2001.
- [17] S. McDonald and J. Tait, "Search strategies in content-based image retrieval," in *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '03)*, pp. 80–87, Toronto, Canada, July-August 2003.
- [18] T. Berners-Lee, J. Hendler, and O. Lassila, "The semantic web," *Scientific American*, vol. 284, no. 5, pp. 34–43, 2001.
- [19] A. Kaikkonen and V. Roto, "Navigating in a mobile XHTML application," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03)*, pp. 329–336, Ft. Lauderdale, Fla, USA, April 2003.
- [20] L. Kim and M. Albers, "Web design issues when searching for information in a small screen display," in *Proceedings of the 19th Annual International Conference on Computer Documentation (SIGDOC '01)*, pp. 193–200, Sante Fe, NM, USA, October 2001.
- [21] Y. A. Aslandogan and C. T. Yu, "Evaluating strategies and systems for content-based indexing of person images on the web," in *Proceedings of the 8th ACM International Conference on Multimedia*, pp. 313–321, Marina del Rey, Calif, USA, October-November 2000.
- [22] E. Newcomb, T. Pashley, and J. Stasko, "Mobile computing in the retail arena," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03)*, pp. 337–344, Ft. Lauderdale, Fla, USA, April 2003.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

