Evaluating Use of Interfaces for Visual Query Specification

Lars-Jacob Hove

Department of Information Science and Media Studies
University of Bergen

Fosswinckelsgate 6
Pb 7800, 5020 Bergen

Lars-jacob.hove@infomedia.uib.no

www.vq.uib.no

Abstract

We present an empirical study of how users express information needs visually through "Query by sketch" interfaces. The objective of the study has been to gain an understanding of how users translate information needs into visual queries (VQ) and what type of problems they encounter. We have studied the process of translating an information need to a visual query and how a visual query interface assists or hinders the user during this process. The intent is to use this information to both improve user support in the task of formulating visual queries and assist the retrieval system in understanding the user's query.

14 participants were given a set of image retrieval tasks and were asked to perform them in a basic VQ interface. A total of 121 visual queries were performed. The process of formulating these queries was monitored using observation, video and audio logging, and the participants were interviewed after performing the retrieval tasks. We have identified three main categories of tools for visual query formulation, *freehand sketching, use of predefined shapes* and *use of colour*. The overall use of these categories for visual query formulation has been determined. The resulting query images have been analyzed, and are categorized as either *realistic* or *iconic*, and as *complete images* or images with a *single object of interest*, providing 4 distinct categories of visual queries.

We discuss the consequences of these results with regards to interfaces for visual query specification. The results indicate that the way users prefer to express information needs visually differ from the input expected by current CBIR systems. Some possible additions to QBS interfaces are suggested that might help users express their information needs as visual queries.

INTRODUCTION

Research within the field of Content Based Image Retrieval (CBIR) systems has shown that it is possible to perform image retrieval tasks based on feature-based similarity searches. Such systems have proven to be useful in a variety of contexts, such as fingerprint and face identification and searching trademark databases (Smeulders, Worring, Santini, Gupta and Jain 2000). However, in order to be useful for a larger application area, two main challenges must be overcome. First, such systems need to overcome “the semantic gap” and understand the semantic content of a query, not only the low-level visual features. Second, the systems must be adapted to the needs and behaviours of the human user (Lew, Sebe, Djeraba and Jain 2006).

A large share of research in multimedia information retrieval has been aimed at the development and improvement of fast, reliable and working image indexing techniques. The end user and their tasks, needs, requirements and expectations by contrast have received relatively little attention. While it is clear that further improvement of the underlying techniques is fundamental for further advancement in the field, a human-centred approach might provide very important insights on how to support and satisfy the end user’s needs, as well as indicate alternative ways of create and organize CBIR systems to fully utilize the user’s potential.

One area which traditionally has received relatively little attention is the user interface, the tools necessary for visual query specification and how users can successfully express their information needs through these interfaces using their own terminology (Venters, Hartley, Cooper and Hewitt 2001; McDonald and Tait 2003). One approach for visual query specification which provides the user with a high degree of freedom for visual query specification is *Query-By-Sketch*. (For an overview and classification of other visual query types, see for example Smeulders, Worring et al (2000)). QBS was originally developed for IBM’s QBIC System (Niblack, Barber, Equitz, Flickner, Glasman, Petkovic, Yanker, Faloutsos and Taubin 1993), and is primarily based on having the user compose an example image using colour distributions; basic geometric shapes such as circles, squares, triangles, and rectangles; free-hand sketching or combinations of these. The main goal of QBS is to let the user create a template of either a completed object or scene, which is used as a basis for similarity matching against an image collection. According to Venters et al (2001), there is little evidence to support the usability of such query tools, and the these interfaces remain one of the least researched and developed element of CBIR retrieval systems. However, it is generally acknowledged that the main drawback with this approach is that it is highly dependant on the user’s ability to create good example images. Though the actual tools and methods used to draw sketches might be simple and easy to use, creating *good* drawings can be difficult and is likely highly dependent on the artistic abilities of the enquirer. As noted by Venters et al (2001), supporting the construction of image queries by this mode of interaction is a non-trivial problem. The more complex the query, the more difficult it is for the end user to express, and produce a meaningful visual example resembling his information need. This indicates a potential conflict of interest between the needs and skills of the end users and the capabilities of CBIR systems.

Figure 1 presents a simplified view of a CBIR image search. The system analyzes the query image, extracts a set of features, which are compared to the features of the images in the collection. Retrieval is normally based on similarities between the query image and the image collection. Successful comparison is dependent on the query image being either be as detailed and realistic as possible or represented in an abstract way, e.g. a colour distribution. On the other hand, the end user might not have the necessary skill needed to create high-quality query images, or be willing to spend the time required to create these.

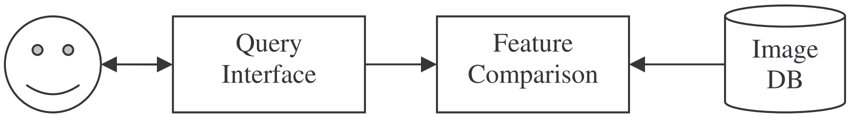

Figure 1 - Simplified view of CBIR query processing

We believe that an examination of this potential conflict of interest might provide important insights towards both how we can create interfaces that support the users' ability to create visual queries, as well as how CBIR system architecture can be adapted to better understand visual queries created by human users. This paper reports on an exploratory study examining how humans behave when expressing information needs as visual queries to a CBIR system. It is believed that the results obtained from this study might provide some insights into how users interact with such systems. These insights might be used as a basis for improving both user interfaces and CBIR query interpretation.

EXPLORATIVE STUDY

The primary goal in this study has been to explore the use of query-by-sketch as a tool for visual query specification: How do users translate their information need into visual queries using a basic interface for query-by-sketch visual queries. We hypothesized that users would prefer to create simple, iconic images over fully detailed, realistic images. Next, we hypothesized that the users would make frequent use supporting tools such as lines and predefined shapes to counter their lack of artistic abilities and drawing skills. In order to evaluate these hypotheses, we devised an explorative study in which a group of volunteers were invited to use a CBIR system to perform a set of tasks. Our main focus was to evaluate and measure the participants use of interface tools, and to attempt to classify the query images according to our own classification scheme.

TASKS, PARTICIPANTS AND MATERIALS

The study was conducted in a laboratory setting. Each study took place in a room with only the participant and the researcher present. The experiment was performed on a portable computer[1] with screen and video capture software. A graphical tablet and pen was used as the primary input device. A stereo audio recorder was used to capture sound. A prototype system, VISI, was used was used as a representative for state-of-the art CBIR system.

The VISI Prototype

The VISI prototype is a prototype web based CBIR system[2]. It was developed as a quick test bed for content-based image retrieval using a Cold Fusion front-end on top of Oracle 9i (O9i) InterMedia CBIR software. Visual Query Specification is provided through a sketch interface, illustrated in Figure 2. The tool is based on a Java Applet, *J-Painter*[3], and provides a set of basic drawing tools: Freehand drawing; basic geometric shapes such lines, rectangles, circles; colour selection and a limited palette of pens.

---

[1] Macbook pro 2.2
[2] Available online at: http://nordbotten.ifi.uib.no/VirtualMuseum/Prototypes/Osdal3/
[3] http://www.izhuk.com/painter/

The image collection used in this study was a limited collection from a maritime domain, depicting maritime animals, people involved in maritime activities and different objects and activities related to this.

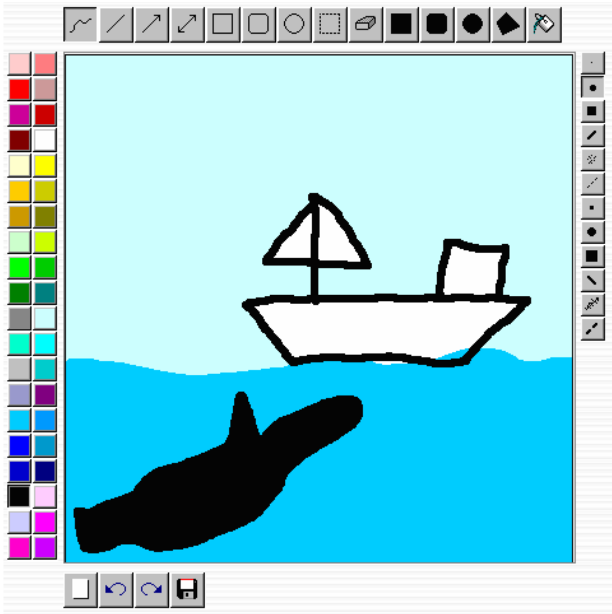

Figure 2 - VISI Sketch tool with an actual query image

Input Device

A graphical tablet with a digital pen was used as the primary input device to the query specification tool. While a mouse might be the most commonplace input device for computers, a mouse might not be the preferred tool for drawing. Recent studies (Barthelmess, Kaiser and Lunsford 2006) has shown that interfaces that depart from existing practices require a higher cognitive load from the users, taking focus away from the task at hand. Furthermore, the results presented by Venters et al (2001) indicate that a graphical tablet might be a more natural and direct input for visual queries. We wanted to give the participants optimal conditions for graphical input, and consequently provided a state-of-the-art graphical tablet as an input device.

Participants

The people who took part in the study were all university students. We feel justified in our use of students as study subjects, as the image retrieval systems we are studying are intended for use by a wide population of users. We believe that students are relatively heterogeneous and representative of a wider population with regards to artistic ability and skill in visual query specification. Independent variables were categorized from a phenomenological perspective, the users deciding on their own level of ability and determined by means of a questionnaire.

14 people participated in the study, 11 males and 3 females, aged 22 through 34. Two of the participants had some prior knowledge of CBIR from previous work. The remaining participants were given a brief introduction to the basics of CBIR functionality, and all claimed a basic understanding of how CBIR similarity search worked. All participants were fairly proficient with image management systems such as Flickr. All participants used text-based image retrieval on a regular basis.

The participants were given a presentation of both the search interface and the input device and a brief overview on capabilities of the VISI prototype through examples and walkthroughs of query-by-sketch searches. Each participant was also given ample time to familiarize themselves with both the tablet and J-painter. Finally, the participants were given an approximate of the number of images in the collection, as well as the topics covered in the image collection.

Tasks

The participants were given two types of tasks representing two different levels of complexity: specific image retrieval tasks and case based scenarios. In the former case, participants were asked to perform a set of basic image retrieval tasks, such as "Find images of one or more sharks" or "Find images depicting a seagull eating". These tasks represent non-complex retrieval tasks, which might be relatively easy to express using linguistic terms. The purpose of these tasks was to give a set of common tasks to each participant, permitting some degree of comparison. However, these tasks were very specific in nature and might not be representative of realistic image retrieval tasks. As noted by McDonald, Tait and Lai (2001), the use of scenarios might simulate how participants might typically interact with image retrieval systems in a non-experimental setting. Furthermore, we wanted to include tasks which were of a more complex nature than the first set of tasks. Consequently, the participants were presented with a 600 word text describing a series of incidents around a bottlenose dolphin visiting the coasts of Norway. They were asked to study the text, and describe at least two images which might be used to illustrate the text. The participants were then asked to attempt to retrieve such images using the VISI system. As the specific image retrieval tasks were mainly based on retrieval of specific entities or entities interacting, the participants were encouraged to find complete scenes based on the scenarios. In the following, the two query types are referred as *task specific queries* and *scenario based queries*.

DATA COLLECTION

Several different types of data were collected in order to obtain a rich description of the process of visual query specification. A large portion of the data was of a qualitative nature, requiring a great deal of effort to analyse. Because of this, we decided to keep the number of participants and tasks relatively small. While this might limit the generality and statistical validity of the results, it provides a solid empirical foundation for later, larger scaled studies.

Video and Audio Logging and Observation

Video capturing software[4] was used to record all actions performed by each participant while using the VISI prototype. A think-aloud protocol was used, and the participants were asked to talk about what they were doing, why they were doing it, difficulties encountered and similar questions. The researcher also observed the participants while performing the search tasks, and asked questions when the participants had difficulties or other breakdowns. Audio was captured using a stereo sound recorder.

The video was first used to capture the time spent on query specification by measuring the time from the frame where the participant initiated a new search to the frame when the query image was submitted. Next, the video stream was also used to identify and classify the different tools used to create the visual queries. Finally, the video was used synchronized with the audio recording and used to determine critical incidents, breakdowns and other interesting observations, and used to supplement other results.

[4] Snapz Pro - http://www.ambrosiasw.com/utilities/snapzprox/

Semi-structured Interview

Immediately following the retrieval tasks, each participant was subject to a semi-structured interview. The interview was transcribed, organized and coded, and subjected to a qualitative analysis (Patton 1980). The participants were asked to elaborate on their choice of tools and the process of creating query images. This was used as an aid for classification of the query images, and to obtain data about the motivation and choices made by the participants.

Query Images

Each query image created by the users were stored on the system and retrieved after each study. A total of 121 images were created, 77 based on specific query tasks and 44 based on scenarios. The task specific queries were all based on the retrieval of a single object or a small number of objects interacting, while the scenario based tasks were all based on retrieval of particular scenes or objects in a particular environment. Consequently, they represent two different query types and have been subject to both separate and combined analysis.

DATA ANALYSIS AND RESULTS

The data gathered in the study was used to analyse the two main research hypotheses, use of iconic images versus realistic images, and the use of different drawing tools.

Query Image Classification

Our first hypothesis was that users prefer to create simple, iconic images over fully detailed, realistic images. In order to evaluate this, a two-dimension classification scheme for visual query images was devised. First, the query images are classified as either *realistic* or *iconic*, where:

- *Realistic* describe images where the objects and scenes are drawn as to resemble the real-world objects they represent.
- *Iconic* describe images where the objects and scenes are drawn as simplified drawings with an iconic representation of their real-world counterparts.

Classification into these categories was performed by the researcher by evaluating each image separately. The criterions used were: *Use of colour* (Saturation, differentiation and modulation), *contextualization* (absence or articulation of background), *level of detail (*use of outlines versus complete objects) and the *level of abstraction* (use of iconic representations such as stick figures). For images with combinations of realistic and iconic representations, the dominating characteristic was used to classify the image. The term "Iconic Representation" in this context represents a sign or representation that stands for its real-world object by virtue of a very simplified depiction or an analogy to it.

Next, images are classified as either *complete representations* or *object-of-interest*, where:

- *Complete* describes images where the images are drawn as to resemble a complete image, with objects in a natural environment
- *Object-of-interest* describes images where only an object of interest, e.g. a dolphin, is depicted.

Classifications into these categories were also performed by the researcher. Images containing depictions of objects directly placed on a white canvas were classified as object-of-interest, while images with objects placed in an environment were classified as complete. The main criterion was the inclusion of other objects than the object of interest, e.g. waves, landscape or the sun in a query after one or more dolphins.

This classification scheme provides 4 distinct classifications: “Realistic and complete” (RC), “realistic and object-of-interest” (RO), “iconic and complete” (IC) and “iconic and object of interest” (IO). Figure 3 presents an illustration of these categories with four actual query images representing the categories.

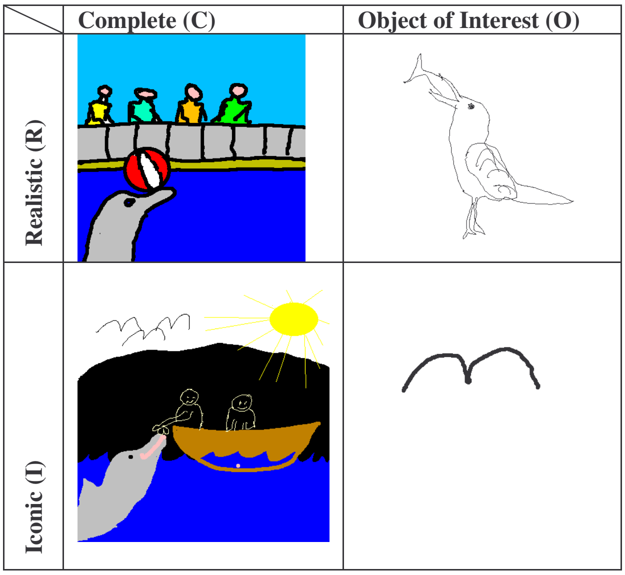

Figure 3 - Image classification categories with examples

Table 1 shows the classification for all images. If we look at the overall distribution, 78 images (65%) of the images were classified as “iconic”, and 78 images (65%) were classified as “object-of-interest”. Only 14 images (12%) were classified as “realistic and complete”, while 49 (40%) were classified as “iconic and object-of interest”. However, distinguishing between the query types reveals some interesting observations.

Table 1

Image query classification

Type	All images		Scenarios		Task specific	
RC	14	11.5%	6	13%	8	10%
RO	29	24%	4	9%	25	33%
IC	29	24%	21	48%	8	10%
IO	49	40.5%	13	30%	36	47%

First, a majority of the scenario based queries are classified as complete (61%), while a large majority (80%) of the task specific queries are classified as object-of-interest. Considering that the scenario based queries are primarily based on complete scenes, it is not surprising that a majority of these are classified as complete. However, it is interesting to note the majority of object-of-interest in the task specific queries. It seems as if when faced with the task of retrieving a particular object, the participants preferred to focus only on the desired object when creating visual queries. When asked, several participants stated they wanted to assist the retrieval system and keep these queries clear of “visual clutter”. Several participants also stated they did not want to spend time on “unnecessary details” such as “surroundings” when their retrieval task was clearly defined. This might be understandable from the users’ point of view. However, it is highly unlikely that a retrieval system based on pixel-by-pixel comparison with real images will achieve a high precision from such queries. In relation to this, it is also with interesting that the 6 scenario based queries and 4 the task specific queries classified as RC, were created by the two participants with prior CBIR experience. During the interview session, both participants stated that they attempted to create query images as realistic and complete as possible in order to achieve good results, based on prior experience and knowledge of CBIR systems.

Another interesting observation is the number of queries classified as RO. For the scenarios, this is the smallest category (9%), while it is the second largest category (33%) for the task specific queries. This observation might indicate that the participants were more likely to attempt realistic queries when retrieving specific objects. When asked, several participants claimed they attempted to create such queries by drawing as realistically as possible, particularly from objects they believed were simple to draw. However, as the complexity of the objects and scenes being retrieved increased, the participants' inclination to use iconic representations grew. Similarly, the more knowledge the participant had of the object in question, the more likely they were to use an iconic representation. For example, most queries involving humans were expressed using iconic representations.

It is quite possible that these observations are related to the users' drawing skills and the complexity of the tasks. Complex objects are naturally more difficult to draw. Furthermore, as the knowledge of an object grows, it is possible that the user will attempt to include minute details that might appear important to them, but might be of little value for the retrieval system, e.g. the shape of a human's eye, or spending time creating a realistic depiction of a beach ball. This is further reinforced by looking at the video/audio log of the query specification process. Several participants attempted to create realistic representations of complex objects, but changed to iconic representations. When questioned during the interview, several participants stated that realistic representations were either too difficult or too time consuming to create. Finally, the time spent during query specification for the different categories give further reinforcement. If we look at the time spent on the different main categories, the median time spent on realistic images were 199 seconds, 140 seconds for iconic images, 205 seconds for complete images and 155 seconds for objects-of-interest. A two sample t-test revealed that the difference between these scores is significant ($t = 3.07$, $p < 0.01$ for iconic versus realistic images, and $t = 3.89$, $p < 0.01$ for complete scenes versus objects of interest).

Based on this, there appears to be some evidence to partially support the hypothesis that users prefer to create simple, iconic images over fully detailed, realistic images, particularly when querying after images of particular objects. There are also several indications that the use of iconic queries is somewhat related to the complexity of the retrieval task.

Use of Tools

Our second hypothesis was that users would prefer making frequent use of drawing tools, such as lines and predefined shapes. In order to evaluate this, we decided to distinguish between three main categories of drawing tools:

A. *Freehand sketching*: the use of the freehand tool
B. *Predefined shapes*: The use of predefined shapes such as lines, arrows, circles and squares.
C. *Colour*: The use of other colours than black on a white canvas.

Supporting tools, such as the colour filler, the duplicator and the eraser were included in all three categories. White was considered to be a part of the canvas and not a colour unless it was an actual part of the image, e.g. a polar bear or objects on ice or snow. Using this classification scheme, we measured overall use of each category and the combinations used in each query image. We classified the tool use for each of the 121 query images based on both video of the query specification process and the final images. We also measured the average and median time used to create query images using the different combinations.

Table 2 shows the overall tool use for query types, while Figure 4 shows the distribution of the various combinations used to create the query images, e.g. 22% of the query images were created using only freehand drawing (A), while 27% of the images were created using a combination of all three categories (ABC). Similar measurements were created separately for both query types.

Table 2

Overall tool usage for all queries

Tool	# Images	%
A: Freehand Sketching	117	97 %
B: Predefined shapes	39	32 %
C: Colour	90	74 %

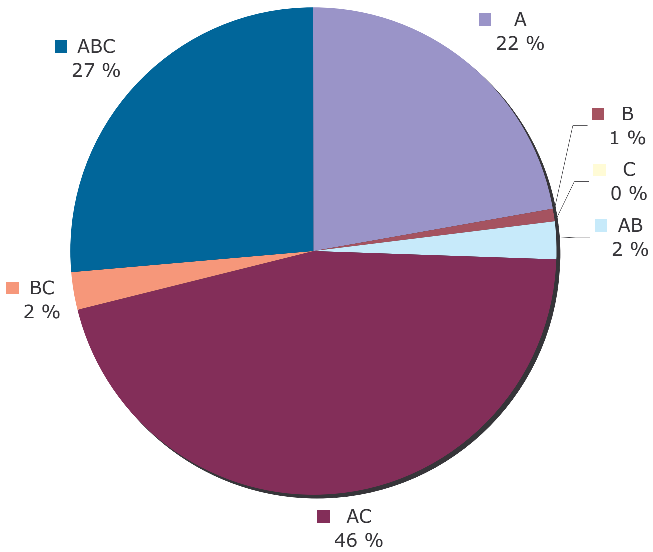

Figure 4 - Distribution of tool use for all queries

The first impression from this is that contrary to the first hypothesis, freehand drawing was clearly the preferred tool. Looking at the tool combinations (Figure 4), we see that freehand alone (A) was used alone in 27 images, in combination with shape (AB) in 3 images and combined with colour (AC) in 55 images. Predefined shapes alone (B) were used in 1 image and in combination with colour (BC) in 3 images. Colour (C) was not used alone in any images. The combination of all three tools (ABC) was used in 32 images. The predefined shapes were the least used category, both alone and in combination with other tools.

If we distinguish between the two types of tasks, the results are similar with two major exceptions. For the scenario based queries, the combination of all tools (ABC) was used in

45% of the queries and the combination of freehand and colour (AC) was used in 43%. Freehand alone (A) were only used in 4.5% of the queries. Next, for the task specific queries, A was used in 32% of the queries, AC in 46% and ABC in 15%.

These results raise two interesting issues. First, concerning the considerable use of freehand sketching, there is evidence to support that it requires less effort and time than the use of predefined shapes. If we look at the median time spent on the different categories, we see that the participants spent considerably less time using A (68 seconds), compared to AC (186 seconds) and ABC (199 seconds). The differences were found to be significant ($p < 0.01$). When asked, several participants claimed that freehand sketching was faster and more efficient than using the predefined shapes. Most participants claimed that their own drawing skills represented the greatest problem, and that the predefined shapes might alleviate some of the difficulties of freehand drawing such as drawing a straight line or a perfect circle. However, they felt that the increased accuracy could not justify the time and effort needed to adapt the predefined shapes to the image they were creating. In several cases, participants first attempted to use predefined shapes, but reverted to using freehand sketching after spending time trying to adapt the shapes to their drawing. Several participants also stated that the shapes provided by the interface did not match up with the images they tried to retrieve. Further evidence towards this is found if we study *how* the predefined shapes were used during query specification. In 14 of the 32 queries where all three tool groups were used, the predefined shapes were used to create objects that are defined by such shapes in the real world, e.g. a round ball or the sun. In 9 of the queries the shapes were used to define large areas of the image, e.g. using a box or a line to denote either the sky or the ocean. In 3 images, shapes were used as abstract representations of objects, e.g. an oval with a small circle representing a shark. Other than this, shapes were used to create minor details such as eyes or windows. Put together, these results indicate that a selection of basic geometric shapes does not provide users with the means to express a wide range of real world objects, making freehand drawing a more viable alternative.

Next, considering the differences in tool use between the two query types, there is some indications that this might be related to the complexity of the queries. The specific retrieval tasks were mostly based on retrieval of a certain object or objects, or objects involved in a particular activity, such as “one or more seagulls eating”. The retrieval tasks defined by the scenario often reflected a certain scene, e.g. “a dolphin swimming in front of a boat, with people”, representing more complex retrieval tasks. The query image classification revealed that queries after particular objects were often created as a drawing of the object of interest on a white canvas, while queries after complete scenes were created by including details such as the sky, the ocean or the shoreline. These were usually represented by coloured areas, often created using shapes such as rectangles or lines. If we compare the use of colour, we see that it is used in 93% of the scenario based queries, against 63% in the task specific queries. Similarly, shapes were used in 52% of the scenario based queries, and 20% in task specific queries. Finally, freehand sketching alone (A) was used in only 4.5% of the scenario based queries, and in 32% of task specific queries. All differences were found to be significant ($p < 0.01$). Furthermore, several participants claimed they felt the simple retrieval tasks could be expressed using only freehand sketching, while more complex tasks, such as complete scenes or complex objects such as humans, required more detail. More than half of the participants explicitly stated that coloured shapes, such as rectangles, were particularly suited to create backgrounds such as shoreline, sky or the ocean. Several participants also stated the shapes could be used to add details to complex objects, such as a human’s eye or a shark’s tooth.

The above results indicate two important observations. First, it seems likely that basic geometric shapes are too rigid to be of direct use when creating visual queries for real-world objects. A certain degree of flexibility and freedom is required to express these objects, and freehand sketching appears to be better suited for this. Next, as the complexity of the queries increase, the need for supplemental tools, particularly colour, is required to fully express information needs visually.

Discussion, conclusions and future work

This work has implications for both the design of visual query interfaces, and the way CBIR systems should handle image queries. First of all, we have seen that tools such as freehand drawing and a set of basic predefined geometric shapes are insufficient for supporting visual query specification. First, freehand sketching seems to be sufficient for expressing simple queries, allowing flexible and efficient query expression. However, as the complexity of the queries increase, limitations in user skill make this insufficient. A set of predefined, basic geographic shapes provide too little flexibility and freedom to be fully usable for most expressing real world objects, but might be useful in certain circumstances. Colour appears to be very important when searching for complete scenes, but not very important when searching for particular objects. Next, there are indications that users prefer to create iconic representations of objects and scenes, particularly with increased complexity in their information needs and query images.

For visual query interfaces, these observations indicate that a more varied set of tools could be of use. Two possible additions were suggested by a majority of the participants. First, all participants would like the ability to deform or customize the basic shapes included in the tool. This would make the basic shapes more useful, and might possibly reduce the need to use freehand drawing to create some objects. Further studies should be performed to evaluate the usability and actual implementation of this suggestion, but it is highly likely that this would be very beneficial to the end user.

Next, several participants claimed that inclusion of more domain specific tools such as representations of domain specific objects might be very helpful, e.g. sharks or dolphins for a maritime image collection. Such representations might be real images of objects for realistic representations, or a set of user-customisable shape prototypes for iconic representations. A similar approach to this is used in the InterScape search engine (Lew 2000). In this system, the user could make direct queries using spatial positioning of a limited set of iconic representations of visual objects such as sky, trees, water and so on. However, while he results reported in this study suggests that such domain tailored tools might be very beneficial to the user, the number of possible real-world domains in image collections make the development of customized tools for each domain a very challenging and labour intensive work. However, the suggested usefulness of such tools for a human user indicates that further studies should be performed to evaluate if it is possible to balance these considerations and make use of this potentially powerful tool.

Regarding CBIR system architecture, this study implies it might be fruitful to reconsider the use of similarity between a query image and an image collection as the basis for image retrieval. We have seen that users have a tendency to use iconic representations as the query complexity grows, or create query images depicting the object of interest isolated from their natural surroundings. We cannot reasonably expect that a retrieval system based on direct comparison between such images and an image collection will provide results with any degree of recall or precision. One possible direction might be to introduce a *visual query interpreter* between the visual query interface and the retrieval system, as shown in Figure 5.

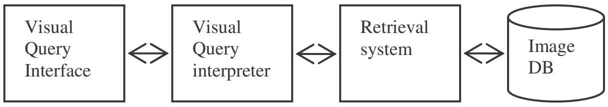

Figure 5 – Possible architecture for a CBIR system

While further studies needs to be performed, it is believed that the task of recognizing the components in a query image is easier than bridging the semantic gap directly between a query image and an image collection. This might be particularly the case if visual query specification is based on iconic representations. It should also be examined if it is possible to use machine learning to recognize freehand user sketches. While it has proven difficult to create systems capable of visual-concept detection of real objects, it might be possible that detection of *user drawn sketches* representing such objects represents might be easier. Variances in these might be smaller than in their real-world counterparts, which might produce better results. A visual query interpreter might provide the retrieval system with the capacity to interpret visual queries, and translate these into terms more suited for machine-based image retrieval.

The main benefits of such an approach would be to better support users in expressing visual information needs. It could alleviate some of the main challenges to visual query specification, such as limitations in drawing skills and the time and effort needed to create such queries, while retaining the strengths, such as the possibility to describe spatial relationships and other visual structures difficult to express linguistically.

The main objection to this approach is that while it is palpable that users might benefit, it will not solve the problems of automatic image indexing. However, some recent initiatives involving community-based indexing methods, such as The ESP Game (von Ahn and Dabbish 2004) and Peekaboom (von Ahn, Ruoran and Blum 2006) has shown that it is possible to semi-automatically detect and index objects present in an image, as well as the spatial distribution of these objects. It is believed that by combining techniques based on human computation with the suggested visual query interpreter and a visual query language, it is possible to achieve results surpassing both traditional text based retrieval and feature-based CBIR. Further studies will be necessary to evaluate this.

LIMITATIONS AND GENERALITY

First of all, it should be noted that this study only evaluates the users' preferences when expressing queries in a given context, not the actual recall and precision achieved by the retrieval system. Further studies should be done to measure the impact of visual query interfaces on retrieval efficiency.

Next, the methodology of this study has both strengths and drawbacks. The major drawback is that the limited number of participants and query images clearly limits the statistical validity and generality of the observed results and classifications. Some of the results, particularly pertaining to the choice of tool use and time measurements could be better evaluated using a larger quantitative study.

Furthermore, both the image collection and retrieval tasks are limited to objects, activities and entities related to maritime domain. It is possible that this has influenced the results, but we believe that both the images and the tasks are representative for a general visual image retrieval application.

The main strength of the chosen methodology is that the small number of participants provided us with the possibility to do a detailed interview with each participant, allowing for a richer description of user behaviour than pure quantitative measurements would provide. Next, the detailed evaluation and classification of query images is a very labour demanding process, and it is not clear how this process could be automated. The limited number of query images allowed us to spend time analysing each image, hopefully providing a thorough classification.

We believe that while the limitations in the reported study somewhat limits the generality, but the results obtained may both provide a foundation on which to base further studies, and provide some important insights into how users behave when creating visual queries.

FURTHER STUDIES

We are currently working on establishing a classification of possible visual query language elements, aimed at identifying possible objects and elements which might be beneficial for users creating visual queries. Our aim is to combine theory from the field of multimedia retrieval with theories of visual literacy, grammar for visual design and human visual interpretation to create a human-centred interface for visual queries. A prototype visual query interface and visual query interpreter based on this is planned as a deliverable from this project. Additionally, we are planning a larger scaled study following up some of the observations from the current study in order to establish statistically sound results.

About the author

Lars-Jacob Hove is a research fellow at the Department of Information Science and Media Studies, University of Bergen. His Ph.D. project, *Visual Queries*, is a study of the use and usability of Visual Queries in digital image retrieval. He received a Master’s degree (Cand. Polit.) in Information Science at the University of Bergen in 2004. His research interests include information- and database management, multimedia retrieval and content based retrieval.

REFERENCES

Barthelmess, P., E. Kaiser, et al. (2006). Human Centered Collaborative Interaction. HCM'06, Santa Barbara, California, UAS, ACM.

Lew, M. (2000). "Next Generation Web searches for visual content." IEEE Computer **33**(11): 46-53.

Lew, M., N. Sebe, et al. (2006). "Content-Based Multimedia Information Retrieval: State of the Art and Challenges." ACM Transactions on Multimedia Computing, Communications and Applications **2**(1): 1-19.

McDonald, S. and J. Tait (2003). Search Strategies in Content-Based Image Retrieval. SIGIR '03, Toronto, Canada, ACM.

McDonald, S., J. Tait, et al. (2001). Evaluating a Content Based Image Retrieval System. SIGIR '01, New Orleans, Louisiana, USA, ACM.

Niblack, W. R., R. Barber, et al. (1993). The QBIC Project: Quering Images By Copntent Using color, Texture and Shape. Storage and Retrieval for Image and Video Databases, San Jose, California, USA, SPIE.

Patton, M. Q. (1980). Qualitative evaluation methods. London, Sage.

Smeulders, A., M. Worring, et al. (2000). "Content Based Image Retreival at the End of the Early Years." IEEE Transactions on Pattern Analysis and Machine Intelligence **22**(12): 1349-1380.

Venters, C. C., R. Hartley, et al. (2001). Query by Visual Example: Assessing the Usability of Content-Based Image Retrieval System User Interfaces. Second IEEE Pacific-Rim Conference on Multimedia, Beijing, Republic of China, Springer-Verlag Berlin.

von Ahn, L. and L. Dabbish (2004). Labelling Images with a Computer Game. ACM Conference on Human Factors in Computing Systems, Vienna, Austria, ACM.

von Ahn, L., L. Ruoran, et al. (2006). Peekaboom: A Game for Locating Objects in Images. ACM Conference on Human Factors in Computing Systems, Montréal, Québec, Canada, ACM.