

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/220094639>

# An Interface to Retrieve Personal Memories Using an Iconic Visual Language

Article in *Lecture Notes in Computer Science* · January 2011

DOI: 10.1007/978-3-642-18452-9\_18 · Source: DBLP

CITATIONS

0

READS

92

3 authors:



**Rui M. Jesus**

Instituto Superior de Engenharia de Lisboa - IPL

57 PUBLICATIONS 162 CITATIONS

[SEE PROFILE](#)



**Teresa Romão**

Universidade NOVA de Lisboa

93 PUBLICATIONS 443 CITATIONS

[SEE PROFILE](#)



**Nuno Correia**

Universidade NOVA de Lisboa

235 PUBLICATIONS 1,241 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Inspædia [View project](#)



myCrohn disease [View project](#)

# An Interface to Retrieve Personal Memories Using an Iconic Visual Language

Rui Jesus<sup>2</sup>, Teresa Romão<sup>1</sup>, and Nuno Correia<sup>1</sup>

<sup>1</sup> CITI, Departamento de Informática, Faculdade de Ciências e Tecnologia, FCT, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal

<sup>2</sup> Multimedia and Machine Learning Group, Instituto Superior de Engenharia de Lisboa  
Rua Conselheiro Emdio Navarro nº1, Lisboa, Portugal  
rjesus@deetc.isel.ipl.pt, {tir,nmc}@di.fct.unl.pt

**Abstract.** Relevant past events can be remembered when visualizing related pictures. The main difficulty is how to find these photos in a large personal collection. Query definition and image annotation are key issues to overcome this problem. The former is relevant due to the diversity of the clues provided by our memory when recovering a past moment and the later because images need to be annotated with information regarding those clues to be retrieved. Consequently, tools to recover past memories should deal carefully with these two tasks. This paper describes a user interface designed to explore pictures from personal memories. Users can query the media collection in several ways and for this reason an iconic visual language to define queries is proposed. Automatic and semi-automatic annotation is also performed using the image content and the audio information obtained when users show their images to others. The paper also presents the user interface evaluation based on tests with 58 participants.

**Keywords:** Personal Memories, User Interfaces, Visual Languages, Image Retrieval.

## 1 Introduction

Pictures are one of the richest ways to register and to preserve personal experiences. Currently, due to the advances in digital technology, their capture and storage are easier and very popular, even among people that are not very familiar with technology. Consequently, more experiences are being preserved but it also becomes more difficult to later retrieve media information of an event. Besides, these images are collected in a disorganized way without any type of annotation [5]. To recall this information, humans must remember something about that experience, e.g., the location, the date or the people around. Due to the diversity of the clues provided by our memory and the richness of the visual content, tools to search for personal media should provide mechanisms to use different types of queries including, keywords, images examples, sketches, or parts of maps. Additionally, these tools should not provide difficulties for users with low technological knowledge.

Query definition requires different annotation methods according to the type of query used. If a query is defined by words then media must be annotated with a set of keywords. In this way, a query by image requires media to be annotated using visual features. If the media items are annotated with keywords describing their content, searching for an event becomes an easier task. Unfortunately, inserting those keywords manually is rather time consuming. As a consequence, in a world pervaded by image production and consumption, ways to achieve automatic annotation are essential. Extracting visual content to train semantic concepts (keywords) and associate them with images [12] is one way to obtain automatic annotation. Indeed, keywords may be an intuitive way to search for information but there are some occasions where keywords are not appropriate to express the desired query. When words are not enough, automatically extracted low-level visual features and queries using images examples or sketches must be provided.

This paper presents an interface to explore digital memories to be used when people retrieve their past experiences in domestic environments. This application was designed to support several types of query and for this reason a query language based on icons describing different types of items is proposed. Additionally, the application provides several ways to annotate images. Visual content is used to annotate images by semantic concepts, and audio information, captured when users talk about their media materials, is used to annotate images by recognized keywords.

The paper is structured as follows: the next section presents related tools and the state of the art; the subsequent section gives an overview of the system and its functionalities; the following sections describe the interface details and the image retrieval system. The paper ends with experimental results, conclusions and directions for future work.

## 2 Tools and State of the Art

To manage digital pictures everywhere, using several devices and by different users (e.g., with low technological expertise or elder people), tools to access this information in a familiar way are required. Several interfaces have been proposed to explore images, including desktop, web, mobile and tangible interfaces.

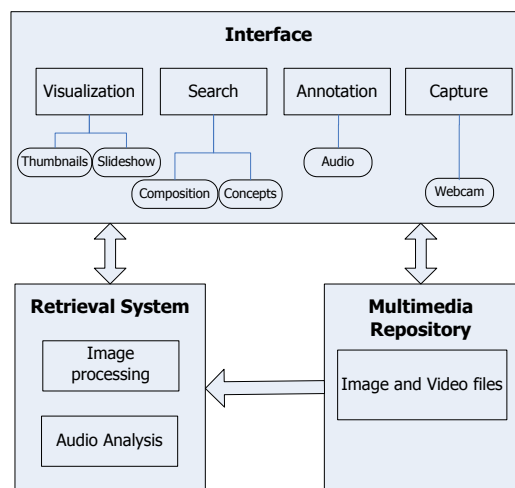
Commercial applications (e.g, Picasa, Photofinder and Adobe Photoshop Album) provide ways to explore this information by means of directory navigation. They require manual annotation as a mechanism to support searching for pictures and allow visualizing those pictures chronologically. Photomesa [1] is an example of another interface for image browsing. It employs Treemap layout to view hierarchies of image directories and provides zoomable interfaces for navigation. Fotofile [6] provides a mechanism for quick annotation, which requires a minimal amount of user interaction. They also integrate both manual and automatic annotation based on image content (face recognition).

Web applications like Flickr and Phlog [17] also use manual annotation, but they have different concerns, such as enabling media items to be shared among friends. Flickr makes use of the manual annotation provided by the community, in order to reduce, for each user, the time consumed by the individual annotation.

Applications described so far require manual annotation, but ethnographic studies [5] show that most people do not spend time annotating images. For this reason, it is important to use a method that simplifies this task. In order to automatically annotate images, content or context metadata should be used. For instance, context information like temporal or geographical data obtained at capture time and recorded on the EXIF (Exchangeable Image File). One interface that explores automatic annotation using GPS (Global Positioning System) information is WWMX [18]. This project uses GPS data information to organize images on a map according to their geographical location. Concerning visual content [14] many systems have been proposed [15]. One of the oldest is QBIC [4]. This system uses color, texture and shape to represent an image and allows defining queries by image examples or sketch using color or texture patterns. Other systems that use sketches to define queries are Retrievr [16] and Imagescape [2]. Retrievr is a Web application where the user produces a drawing to query the database whereas Imagescape in addition, permits creating visual queries using images that represent previously trained concepts.

Content-based systems have one major problem, the semantic gap [9], in other words, the gap between semantic concepts and low level features, such as color and texture. To alleviate this problem, MediAssist [13] uses both context (GPS, date and time) and content information. MediAssist, also provides a mobile interface to explore personal pictures. However, this mobile version uses context data exclusively. Mobile applications are other types of interfaces used to explore personal memories. They allow temporal and spatial independency.

Tangible interfaces, on the other hand, have the primary objective of making interactions between technology and users seem closer to natural practices. Personal digital historian [11] is a project that includes a tangible interface on a table. With this application, people can sit around and look for images, as well as show them to one another, using their hands and fingers.



**Fig. 1.** System Architecture

### 3 System Overview

This paper presents the desktop application of the Memoria project. The main goal of this project is to build tools to explore personal memories. The project includes several applications with interfaces adapted to different contexts of use, including an interface for desktop PC, a tangible interface and a user interface for mobile devices to explore personal collections related with physical locations. The project also includes a multimedia retrieval system.

The application presented in this paper consists of an interface, a multimedia repository and a retrieval system. The interface has four modes of operation (see Fig. 1):

- Capture – to capture personal memories using a webcam;
- Visualization – to browse and visualize images of the personal collection;
- Annotation – to annotate images with audio information;
- Search – to retrieve memories of a specific type, using the visual query language.

To handle these queries a multimedia retrieval system is included. Image processing, classification and audio analysis are among the operations that are supported by this system.

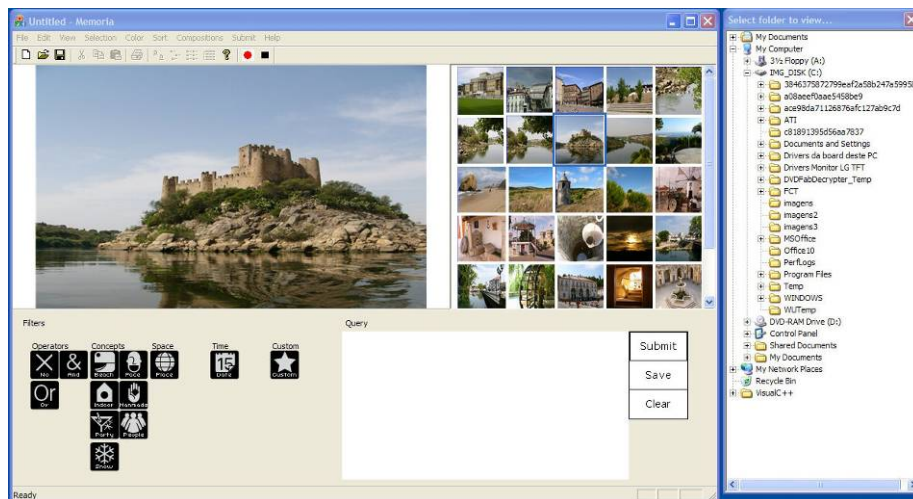


Fig. 2. User interface

### 4 User Interface

The proposed interface makes possible for a user to manage personal memories with pictures. In this way, the interface supports the search and the visualization of a set of images from a multimedia repository, the annotation of these memories with audio elements and the capture of new images using a webcam.

The proposed interface is organized in two main sections (see Fig. 2): the results section and the query section. The results section is located in the upper part of the

screen and is used essentially as a display and selection area, showing the query results or the images that belong to a particular directory. These images are organized in a variable number of thumbnails and in one larger image that exhibits the preview of the selected thumbnail or the preview of an image captured by the webcam. Also, a mini-slideshow can be run inside this area. The query section, in the lower part of the screen, contains a list of filters and a query box. Filters are criteria, such as operators or concepts, that a user can drag into the query box in order to specify a query.

Within these areas various actions take place. The following sections describe the main actions that are available using the proposed interface.

#### 4.1 Capture

The interface by means of a webcam and a dialog permits the capture of images in a simple way. The acquired picture can enrich the personal collection, but it can also be used as input to look for similar images. For instance, the user can search for images with objects alike the one placed in front of the webcam or quickly search for photos of a friend that just arrived without typing names or browsing folders.

This mode allows the user to search for pictures using physical objects, in the spirit of a tangible interface, and can be a way to easy recover personal experiences of people with lower technological knowledge. Fig. 3 shows an example of this mode of interaction with the multimedia repository. For example, if the user wants to find images of parties, by placing a bottle in front of the webcam some party images will be retrieved because bottles usually appear in these images.



**Fig. 3.** Image capture using a webcam

## 4.2 Visualization

Given an image folder selected by the user, thumbnails of these pictures are generated in real-time and presented to the user in the results section. These thumbnails can be sorted using different criteria such as name, time or size of the original file. When one of these thumbnails is selected a preview is displayed. Both visualizations are integrated on the interface (see Fig. 2), so the user has the perfect notion of the image that was selected at any time. Other visualization mode that the interface allows is the slideshow. In this mode, images are displayed in the preview window or in a full screen slideshow and advanced automatically at fixed periods of time.

## 4.3 Annotation

One of the current methods used to classify images uses a set of words (tags) to describe an image. Usually, the user is responsible for manual insertion of these tags and many applications use this information as a basis for image retrieval. This is a regular practice in applications for managing personal photos.

Our approach simplifies this task, allowing the user to annotate images in an automatic manner, using audio elements and visual features.

When people share their photos with friends they use the slideshow and provide comments or stories about the photo that is being visualized to better explain the captured moment and the related context (see Fig. 4). These comments can be very useful to annotate images.

Audio annotation is performed in two different ways:

- With a microphone the user can describe by words an image that is being presented at the moment;
- The user can select a set of images, start a slideshow and while she presents and describes those photos to someone, such as a friend or a fellow worker, audio is being recorded (see Fig. 4).

Recorded audio is analyzed afterwards using an ASR (Automatic Speech Recognition) tool. Given an audio file and a dictionary, this tool recognizes the keywords that were said and annotates images using these words.



**Fig. 4.** Audio recording when commenting images

#### 4.4 Search

The proposed interface offers two approaches when considering visual information retrieval from a multimedia database. One is based on semantic concepts, often used to describe an image and the other consists of a composition of visual elements, that is, selections of parts of images. Both approaches can be used in order to query the database. We call these queries, query by concepts and query by composition. Both types of queries are built using the visual query language proposed. Concepts, logical operators, temporal and geographical items, and images parts are the elements used to build the query using the drag and drop technique to a query box. The following subsection presents the visual query language and explains how it is used to retrieve images.

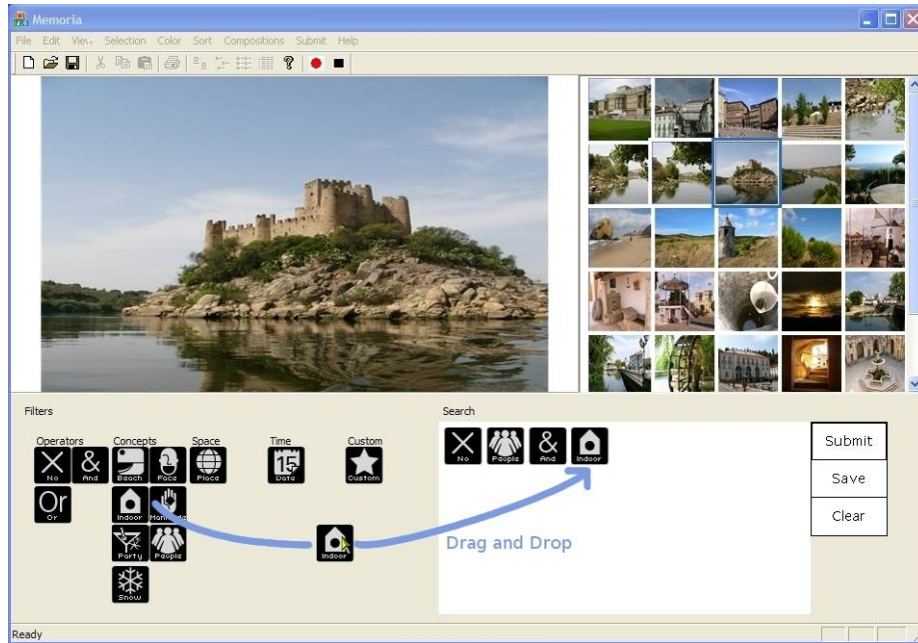


Fig. 5. Drag and drop to a query box

##### Visual Query Language

The visual query language uses the following elements to build queries: image parts, contextual items and concepts. To combine these elements, logical operators are used. At this stage, the operators used are defined by the set,

$$\mathbf{Loperator} = \{\text{AND}, \text{OR}, \text{NOT}\}.$$

The AND operator expresses conjunction or intersection whereas the OR operator expresses disjunction or union. The NOT operator permits to address to all concepts counterparts (e.g., NOT face = No Face).



The user can indicate what concepts are relevant for the search among a universe of pre-established set of concepts:

**Concepts** = {Beach, Face, Indoor, ManMade, Party, Snow, People}.

Using the operator NOT more concepts are defined:

**NOT Concepts** = {No Beach, No Face, Outdoor, Nature, No Party, No Snow, No People}.

Because contextual information is important to recover personal media, a set of contextual elements are also defined:

**Contextual** = {Location, Time}.

Image elements are obtained by cropping sections of different images. Recurring to these elements and operators we can produce queries like,

NOT Indoor AND NOT ManMade.

Such query can be translated to images of the exterior that include nature elements (Note: NOT ManMade = NATURE and NOT Indoor = Outdoor). The integration of these operators can be an asset for retrieving a specific subset from the multimedia database, turning it into more valuable memories.

To retrieve images using the query by concepts method, the user indicates what concepts are relevant for the search among the set of predefined concepts. When establishing the concepts needed to formulate a query, the user denotes that she is looking for images that are somehow related to her choice. For instance, if the selection consists of,

Outdoor AND No People AND Nature,

the system returns a set of images with a high probability to simultaneously depict those concepts. These images can be visualized or serve as input for the other type of query (query by composition).

In order to accomplish a query of this kind, concepts and operators are chosen by means of a drag and drop operation into a query box (see Fig. 5). As opposed to similar applications or standard database querying there is no typing at all. The user merely drags operators and concepts to the query box and presses the submit button. Also, elements present in the query box can be reordered or deleted at any time. This solution is easier and faster than having a user typing keywords.

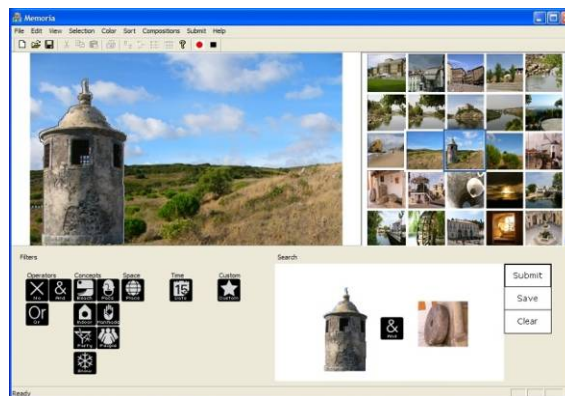
Fig. 5 shows an example of a query by concepts. The concepts People and Indoor, and the operators NOT and AND are in the query box to search for indoor images without people.

Another method that this application offers is query by composition. This method allows composing an image with other images parts (see Fig. 6).

Using queries by composition the user indicates that she wants images with visual properties similar to the objects that were selected. Rectangular or freehand selections

can be done to construct the composition and their location and size is defined by the user. Compositions can be submitted or saved for reutilization. A list of existing compositions is presented on the Compositions menu and the user can select one of them at anytime to resend that query to the database.

Fig. 6 shows an example of a query by image composition. Two image parts and the operator AND are in the query box to define the query and search for similar images.



**Fig. 6.** Query by image composition

## 5 Image Retrieval System

As mentioned before, the user interface described in this paper allows querying a multimedia database using two methods: query by concepts and query by image composition. The first method gives more information to the system because concepts, in general, are trained using hundreds of images and the second gives more freedom to personalize the query, but the information given to the system is weaker.

The query by concepts is based on a set of semantic concepts obtained by training a binary classifier. To combine several concepts in a query, the sigmoid function is applied to the output of the classifier, as described in earlier work [7, 8].

The image composition query works in a similar manner to that of the query by sketch [2], but the query is defined using parts of other images. When the user navigates in the personal collection, sometimes she does not find images that express the intended search. However, there are parts of several images that together can form the desired query (see Figure 6). After the query is formed, the system extracts from each of these parts a set of visual “words” to represent the query, in a way that is similar to text retrieval techniques. Then, Latent Semantic Analysis (LSA) [3] is applied and the cosine distance is used to retrieve relevant images (see [7, 8]).

## 6 Evaluation

To evaluate the proposed application two different types of evaluations should be accomplished: performance evaluation of the multimedia information retrieval system

and usability evaluation of the user interface. The Memoria Desktop interface was built using an iterative design model based on a cyclic process of prototyping, testing, brainstorming and then refining the work developed. This paper presents the usability tests of the working prototype described above and some results obtained by the retrieval system. The evaluation of the retrieval system can be consulted in [7, 8].

### 6.1 Image Retrieval System Evaluation

This section presents the experiments performed to evaluate the multimodal image retrieval system based on semantic analysis. A subset of concepts selected from the set of the 449 LSCOM concepts, suitable for personal collections, were trained using a data set obtained from the Corel Stock Photo CDs, from the TRECVID2005 database and from Flickr. These concepts were tested in the three applications described above, with two different picture collections. The applications Memoria Desktop was evaluated using a personal collection of one of the authors and a database of pictures taken by several visitors of Quinta da Regaleira, a cultural heritage site in Sintra, Portugal. Table 1 presents the results obtained when applying the image retrieval method using only time and visual information to the personal collection. Table 2 shows the performance of the retrieval system in the Quinta da Regaleira database using audio, visual and spatial information.

**Table 1.** Mean average precision (MAP) obtained for a set of concepts combining color moments with the Gabor filter features and a bag of color regions with SIFT descriptors

Concepts	Color Moments + Gabor	Bag color + SIFT
People	0.69	0.75
Face	0.58	0.44
Outdoor	0.91	0.87
Indoor	0.59	0.57
Nature	0.45	0.57
Manmade	0.61	0.71
Snow	0.17	0.09
Beach	0.26	0.34
Party	0.14	0.22
MAP	<b>0.49</b>	<b>0.51</b>

Color moments, Gabor filter features and bags of color regions and SIFT descriptors [8] were the visual features used in the experiments per-formed. When using only visual information (color moments combined with Gabor filter features), the concept “Outdoor” presents good results in both databases, “People” concept is the best in the personal collection and “Nature” concept presents a better performance in the second database.

When time information is included in the visual model, the MAP increases to 0.53 which is better than the values obtained with only visual information (see table 1). Audio information also performs better than visual information (see table 2).

**Table 2.** MAP obtained by a set concepts combining audio, visual and geographic information to select a region or a direction in relation to a point

Concepts	GPS 60m, Visual and Audio	GPS dir, Visual and Audio
Outdoor	0.97	0.64
Indoor	0.09	0.63
Nature	0.86	0.46
Manmade	0.71	0.86
People	0.20	0.50
Indoor + Manmade	0.16	0.48
Outdoor + Nature	0.86	0.51
MAP	0.55	0.58

## 6.2 Usability Evaluation

To analyze the latest system's functionality and assess its usability, we conducted several user tests.

The main goals of this study are:

- To produce detailed user feedback about the Memoria Desktop application that will guide future improvements of the system;
- To refine understanding of the user's needs concerning the tools to manage personal memories;
- To evaluate our proposals to define queries and to visualize the corresponding results.

### 6.2.1 Methodology

The tests were accomplished individually by each user, but several users were testing the application simultaneously. After being briefed about the objectives, users were given a questionnaire that guided them throughout the test. The questionnaire, as explained later in this paper, describes the tasks the users should accomplish and presents related questions that the users should answer after completing the corresponding task. Two facilitators/observers were supervising the tests, encouraging users to "think aloud", helping them when it was essential, taking notes on the users' performances and recording every problem users explicitly mention.

Tests lasted for a minimum of 20 minutes and a maximum of 30 minutes each, depending on each user's strategy to explore the application.

We were mainly interested in:

- Finding out if it is easy to learn how to use the interface;
- Analyzing the aesthetics of the user interface;
- Evaluating the utility of the interface;
- Finding out if it is easy to use the different search techniques;
- Estimating the icons intuitiveness;
- Evaluating the user's satisfaction with the search results.

The information was gathered with the ultimate goal of further refining the Memoria Desktop interface according to the users' feedback.

### 6.2.2 Participants

The tests were performed on a population of 58 voluntary participants, who were graduate students. Ten of them were female. The participants in this experiment ranged in age from 21 to 31 years old with a mean age of 23.5.

Most of the participants own a digital camera and, on average, they claimed they used it 2.7 times per month and 24 times per year (mode = 1 and 12 respectively). When asked about the number of photos in their personal collection answers differ: average = 2592, maximum value = 10000 and mode = 1000. Only fourteen participants (24.1%) declared they usually annotate their photos.

The majority of participants (67.2%) reported that, when they need to search their personal photo collections, they explore the folders in their file system, which are named after the events they correspond to. Some users (13.7%) also referred they use the date as a search criteria. Other users (8.6%) said they use specific software to manage their photo collections.

The clues most used by participants to search for photos related to a past event are: date (60.3%), place (48.3%), event type (20.7%) and people involved (12.1%).

All participants had their first contact with the application during the test and used it under similar conditions.

### 6.2.3 Questionnaire

The questionnaire was composed of three different parts. The first one captured personal data. Besides age and gender, personal data included digital camera usage, dimension of photo collections and photo search methods. The second part guided the users through the interface exploration, explaining the tasks they should accomplish and capturing their experimental feedback. The questionnaire presented four different tasks the users should carry out:

1. Navigate the photo collection using the folder's tree to analyze, in more detail, some images.
2. Visualize a group of images in slideshow mode (normal and full screen);
3. Search images by dragging & dropping concepts and logic operators to the query box;
4. Search images by composing a sketch with parts of other images in the database.

After completing each task, the users had to answer several questions concerning their experience when using the interface. This part of the questionnaire is composed of two types of questions: open-ended questions and Likert scale questions.

The last part of the questionnaire relates to the global evaluation of the interface. Each participant was presented with four statements related to their experience while using the interface. Each user then indicated her level of agreement with each statement by circling a response on a 5-point Likert-type scale. On this scale a response of 1 (one) means disagreement and a response of 5 (five) means agreement.

### 6.2.4 Results

As stated before, the second part of the questionnaire guided the users through four different tasks. In the first task, after using the folders tree to navigate through the photo collection, the users were asked to point out the difficulties they faced while executing the task. Most participants reported no difficulty (74.1%).

Concerning the second task, which consisted in visualizing a group of images in slideshow mode, 36.2% of the subjects described no difficulty; 48.3% considered that the icons were not appropriate, making it difficult to find the right button to start the slideshow. Users suggested the use of labels or tooltips to reveal the function of each button. Several users also claimed for methods to control the slideshow, namely to stop and to move forward or backward.

The third task consisted in building a search image query by dragging & dropping concepts and logic operators into the query box. To complete this task, users had to accomplish several image searches with different goals:

1. Photos with people;
2. Photos showing natural landscapes;
3. Outdoor photos, excluding beaches;
4. Photos showing people or natural landscapes.

After each search, users had to classify the results of the created query using a 5-point scale, where 1 (one) means bad and 5 (five) means excellent. The users' classification is summarized in table 3. In general, the results obtained by using this search method were good.

Participants were then asked about their difficulties to build the queries that drove the searches. This was an open-ended question, so we got many different opinions. While 24% of the participants had no difficulties, the remaining ones pointed out some problems. The most referred one was the need to build the queries sequentially, which makes it difficult to change a query: if users need to insert a new concept in the middle of a query they need to rebuild the whole query (no undo command). They also reported some difficulties in using the operators, specially, because they do not know their priorities. Some of the participants would like to have more concepts available and they reported some difficulty in using the concept "nature" in the queries, requiring a specific icon for it.

**Table 3.** Users classification of the searches results

Queries	Mean	Standard Deviation	Mode
People	3.9	1	3
Nature	3.8	1	4
Outdoor AND No Beach	4	1	4
People Or Nature	3.5	1	4

We wanted to know if users considered the drag & drop method appropriate to define queries for image search. In a 5-point scale, where 1 (one) means inappropriate and 5 (five) means very appropriate, most users considered it appropriate (Mode = 4; Mean = 3.54, SD = 1.1).

We also asked users if the icons appropriately represent the concepts: 81% of the participants considered the icons perceivable and 17.2% had the opposite opinion. The icon representing "People" was elected as the most comprehensible one (chosen by 23 participants), followed by the icon representing "Snow" (chosen by 20 participants). Most unintelligible icons were the ones representing "Manmade" items (selected by 40 participants) and "Beach" (selected by 38 participants).

Moreover, we would like to know if the users consider the method to combine the concept icons to build a query and obtain the corresponding results appropriate. In a 5-point scale, where 1 (one) means inappropriate and 5 (five) means very appropriate, most users considered it appropriate (Mode = 4; Mean = 3.43, SD = 0.82).

Users (34) choose “People” concept as the most useful one when searching images in personal collections. They also suggested new concepts, such as “Sports”, “Sea”, “Day” and “Night”, and seasons (e.g., “Summer” or “Winter”).

The fourth task, searching images by composing a sketch with parts of other images stored in the database, comprises two sub-tasks. In the first subtask, to query the image database users had to make a sketch based on rectangular parts of others images. In the second subtask, participants used the freehand mode to select parts of other images to build a sketch. Image searches were then based on these sketches. After completing these tasks, users were demanded to classify the obtained results. In a 5-point scale, where 1 (one) means bad and 5 (five) means excellent, most participants choose 3 for the first subtask (Mode = 3; Mean = 2.59, SD = 0.97) and 2 for the second subtask (Mode = 2, Mean = 2.2; SD = 0.93). We also wish to figure out if users find this functionality useful when searching for an image in an image database. The majority of participants (67.2%) considered it useful, since it allows them to visually query the image database. Five users (8.6%) think it is particularly appropriate when searching for faces or people. However, 15.5% of the participants considered it useless.

The last part of the questionnaire aims at capturing the participants overall opinions about the interface. Users were asked to evaluate the interface in what concerns its learnability, aesthetics and usefulness. Most participants considered that the information provided by the system is useful, but not with very strong feelings (Mode = 4, Mean = 3.53, SD = 0.79). The results were similar for the statement “it is easy to learn how to use the application” (Mode = 3, Mean = 3.34, SD = 0.83).

Once again, the results were positive, but centered around a medium position, with most participants agreeing with the statement “I like the application aesthetics” (Mode = 3, Mean = 3.19, SD = 0.91) and also the statement “I would use the application to manage my personal photos” (Mode = 3, Mean = 3.12, SD = 1.08).

Overall, the results of the usability tests were quite positive and provided useful feedback to inspire future improvements.

## 7 Conclusions and Future Work

An interface to explore personal memories was presented. It allows browsing and visualizing memories, capturing new media items, the semi-automatic annotation of images using audio information, and the construction and reuse of queries using a visual query language to search for images on a multimedia repository. The described capabilities, in conjunction with a drag and drop operation into a query box, enable the management of personal memories in a way that is fast and also requires a minimal interaction from the user. The interface was evaluated with 58 users that confirm our main options. Most of the users considered appropriate the proposed technique to define queries based on an iconic visual language. They also like the search results (see Table 3). In [7, 8] we evaluated the semantic concepts with a Mean

Average Precision (MAP) of about 50%. As can be seen in Table 3, the users classified this result as “good” which means our image retrieval method performed well in this case. Future and ongoing work includes integrating the users suggestions for the proposed interface and providing more usability tests to analyze in depth some parts of the interface that were less explored, namely our strategies for image annotation with audio. Finally, we are also developing a tangible interface where physical objects are used to create queries.

## References

1. Bederson, B.: PhotoMesa: A Zoomable Image Browser Using Quantum Treemaps and Bubblemaps. In: *UIST 2001, ACM Symposium on User Interface Software and Technology*, CHI Letters, vol. 3(2), pp. 71–80 (2001)
2. Buijs, J., Lew, M.: Visual Learning of Simple Semantics in ImageScape. In: Huijsmans, D.P., Smeulders, A.W.M. (eds.) *VISUAL 1999. LNCS*, vol. 1614, pp. 131–138. Springer, Heidelberg (1999)
3. Deerwester, S., Dumais, S., Furnas, G., Landauer, T., Harshman, R.: Indexing by Latent Semantic Analysis. *Journal of the Society for Information Science*, 391–407 (1990)
4. Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D., Yanker, P.: Query by Image and Video Content: The QBIC System. *Computer* 28(9), 23–32 (1995)
5. Frohlich, D., Kuchinsky, A., Pering, C., Don, A., Ariss, S.: Requirements for Photoware. In: *Proc. of the ACM Conference on Computer Supported Cooperative Work*, pp. 166–175 (2002)
6. Kuchinsky, A., Pering, C., Freeze, D., Creech, M., Serra, B., Gwizdka, J.: FFotoFile: A Consumer Multimedia Organization & Retrieval System. In: *Proceedings of the ACM Conference on Human Factors in Computing Systems CHI 1999*, pp. 496–503 (1999)
7. Jesus, R., Santos, E., Frias, R., Correia, N.: An interface to explore personal memories. In: *15th Portuguese Computer Graphics Group Conference*, Porto Salvo, Portugal (2007)
8. Jesus, R., Dias, R., Frias, R., Abrantes, A., Correia, N.: Sharing personal experiences while navigating in physical spaces. In: *ACM SIGIR Conference on Research and Development in Information Retrieval, Multimedia Information Retrieval Workshop*, Amsterdam, The Netherlands (July 2007)
9. Lew, M., Sebe, N., Djeraba, C., Jain, R.: Content-based Multimedia Information Retrieval: State-of-the-art and Challenges. *ACM Transactions on Multimedia Computing, Communication, and Applications* 2(1), 1–19 (2006)
10. Lowe, D.: Distinctive image features from scale-invariant keypoints. *International Journal Computer Vision*, 91–110 (2004)
11. Moghaddam, B., Tian, Q., Lesh, N., Shen, C., Huang, T.: Visualization and User-Modeling for Browsing Personal Photo Libraries. *Int. J. Comput. Vision* 56(1-2), 109–130 (2004)
12. Mori, Y., Takahashi, H., Oka, R.: Image-to-word transformation based on dividing and vector quantizing images with words. In: *Proceedings of the International Workshop on Multimedia Intelligent Storage and Retrieval Management* (1999)
13. O’Hare, N., Gurrin, C., Jones, G., Smeaton, A.: Combination of Content Analysis and Context Features for Digital Photograph Retrieval. In: *Proceedings of the 2nd European Workshop on the Integration of Knowledge, Semantic and Digital Media Technologies*, pp. 323–328 (2005)



14. Pereira, F., Koenen, R.: MPEG-7: A standard for multimedia content description. *International Journal of Image and Graphics* 1(3), 527–546 (2001)
15. Veltkamp, R., Tanase, M.: A Survey of Content-Based Image Retrieval Systems. In: Marques, O., Furht, B. (eds.) *Content-Based Image and Video Retrieval*, pp. 47–101. Kluwer, Dordrecht (2002)
16. <http://labs.systemone.at/retrievr/>
17. <http://phlog.net/>
18. <http://wwmx.org/>