An Interface to Explore Personal Memories

Rui Jesus[1,2] Edgar Santos[1] Rute Frias[1] Nuno Correia[1]
[1]Interactive Multimedia Group, CITI and DI/FCT, New University of Lisbon
[2]Multimedia and Machine Learning Group, Instituto Superior de Engenharia de Lisboa
Quinta da Torre, 2829 Monte da Caparica, Portugal
rjesus@deetc.isel.ipl.pt,nmc@di.fct.unl.pt

Abstract

Query definition and automatic annotation are key issues when searching for images and videos. Due to the richness of the visual information different types of queries (e.g., image examples, keywords or sketches) are required and consequently, automatic annotation of images with keywords or with visual features must be provided. This paper describes a user interface designed to explore personal memories composed by pictures and videos. The application provides the following operations: capture, annotation, visualization and search for media items in a multimedia database. Users can query the media collection in several ways and for this reason a visual language to define queries is proposed. Queries are defined by drag and drop of visual items and operators to a query box. Automatic annotation is also performed using the image content and the audio information obtained when users show their images to others. Experimental results are presented to illustrate the effectiveness of the application.

Keywords

Personal Memories, Multimedia Information Retrieval, User Interfaces.

1. INTRODUCTION

Digital pictures and videos are essential to preserve personal experiences but there are many challenges for digital applications to manage these human memories. Due to the advances in digital technology and the success of the World Wide Web as a way of sharing media (e.g., Flickr or Youtube), people are collecting a vast amount of personal pictures and videos, most of the time in a disorganized way without any type of annotation [Frohlich02]. With this growth of personal collections, more experiences are preserved but it also becomes more difficult to later recover any media information of an event. To retrieve this information, people 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, because images are so valuable and popular, even individuals that are not very familiar with computers are akin to becoming users of digital interfaces for managing personal memories. Hence, these interfaces 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 is an easy 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 to images [Mori1999] 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, low-level visual features automatically extracted and queries using images examples or sketches must be provided.

This paper presents an interface to explore personal memories in a domestic environment, when people retrieve personal experiences for individual use or to share with others. This application was designed to support several types of query and for this reason a query language based on visual items, semantic concepts and logical operators is proposed. The drag and drop of these items to a query box is used to compose and formulate the query. Additionally, the application provides several ways to automatically annotate images. Visual content is used to annotate images by semantic concepts and audio information, captured when users talk or make commentaries about their media materials, is used to annotate images by recognized keywords.

This paper is structured as follows. Next section presents related tools and the state of the art. The subsequent sec-

tion gives an overview of the system and its functionalities. The following sections describe the interface details and the multimedia information retrieval system. This paper ends with experimental results, conclusions and directions for future work.

2. TOOLS AND STATE OF THE ART

Examples of personal memories, such as pictures and videos, are rich vehicles to exchange information across generations and cultures. To manage images and videos everywhere, using several devices and by different users (e.g., low technological level or elder people), tools to access this information in a familiar way are required. Several interfaces have been proposed to explore images and videos, for instance 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 [Bederson01] 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 [Kuchinsky99] 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 [Phlog] 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, however, ethnographic studies [Frohlich2002] show that most of the 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 [Wwmx]. This project uses GPS data information to organize images on a map according to their geographical location. Concerning visual content [Pereira2001] many systems have been proposed [Veltkamp2002]. One of the oldest is QBIC [Flickner95]. 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 [Retrievr] and Imagescape [Buijs99]. 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 [Lew2006], in other words, the gap between semantic concepts and low level features, such as color and texture. To alleviate this problem, MediAssist [O'Hare05] uses both context (GPS, date and time) and content information. MediAssist, also provide 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. At any time one can show memories to his beloved ones or friends. An example of a mobile interface to access personal memories anywhere is MediAssist. As seen before, this interface explores GPS and time information only.

Tangible interfaces on the other hand have the primary objective to make interactions between technology and users seem closer to natural practices. Personal digital historian [Moghaddam04] is a project that includes a tangible interface, a table on which people can sit around and look for images and show them to one another in an ordinary way, using their hands and fingers.

3. SYSTEM OVERVIEW

This paper presents the desktop user interface of the Memoria project. The main goal of this project is to build tools to explore personal memories. The project includes

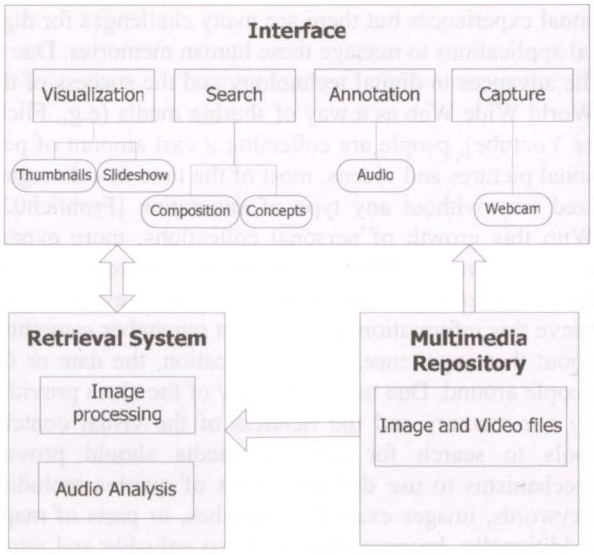

Figure 1. System Architecture.

several 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 proposed application consists of an interface, a multimedia repository and a retrieval system. The interface has four modes of operation (see figure 1):

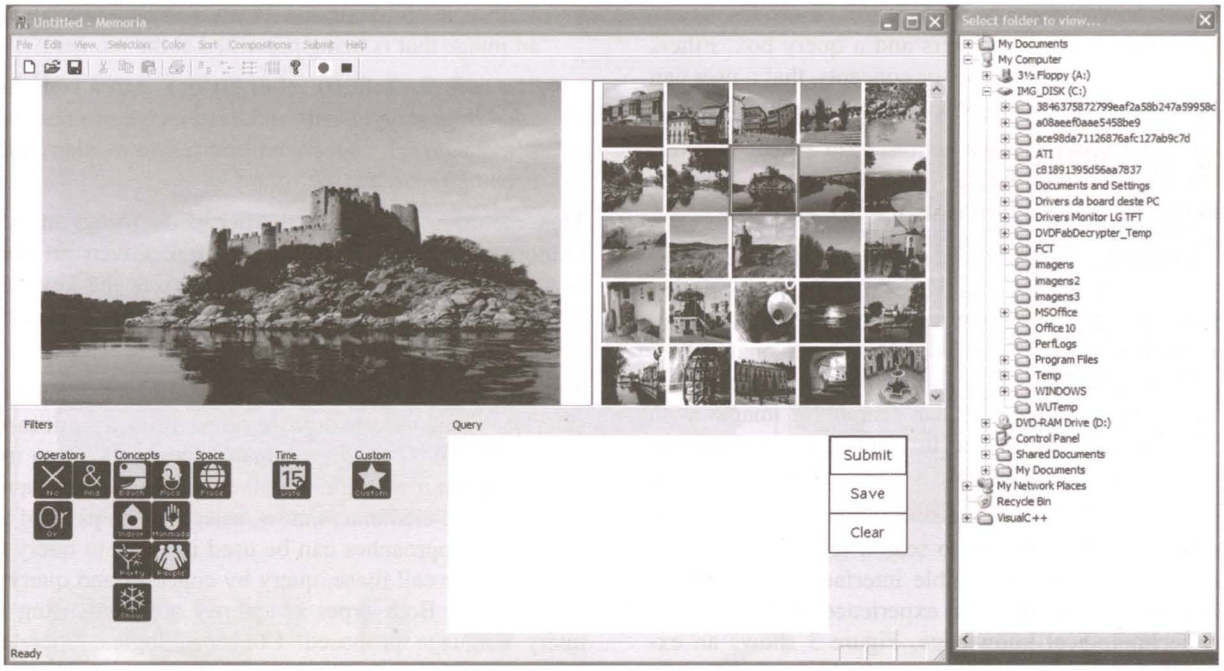

Figure 2. User Interface.

- Capture Mode – this mode is used for capturing personal memories using a webcam. Captured images can be used as input to search for related images;
- Visualization Mode – this mode allows a user to browse and visualize images of the multimedia repository, sort these images using different criteria (e.g., name, date or size) and start a slideshow with the selected images;
- Annotation Mode – this mode is used to annotate images automatically with audio information. Users can make comments that are recorded about their images in two situations: when users annotate by keywords the image content or when they share their images in slideshow mode with friends;
- Search Mode – in this mode a user can produce a query to retrieve memories of a specific type, using the query language. There are two types of queries that can be produced. One consists of a query that uses semantic concepts and the other consists of a query that uses a composed image with visual properties.

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

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 memories using a webcam for this matter.

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

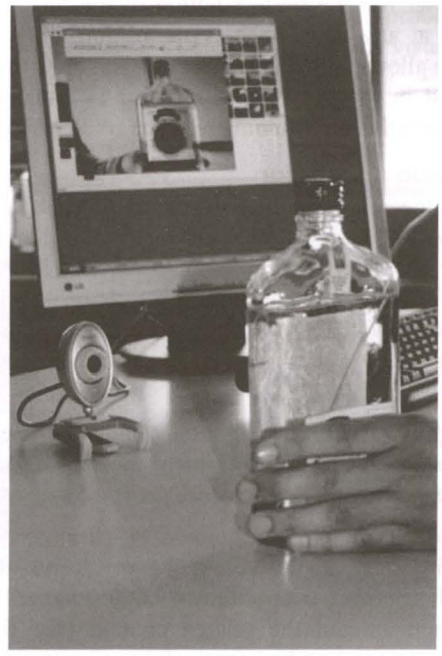

Figure 3. Image capture using a webcam.

mini-slideshow can be run inside this area. The query section 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 the proposed interface puts at the disposal of the user.

4.1 Capture

The interface by means of a webcam and a proper dialog permits the capture of images in a simple way. The acquired picture not only can improve the personal collection but 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 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. Figure 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 appears in these images.

4.2 Visualization

Given a folder selected by the user containing images, 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 chosen a preview is displayed. Both visualizations are integrated on the interface (see Figure 2), so the user has the perfect notion of the image she has 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 as usually in full screen and advanced automatically at fixed periods.

4.3 Audio annotation

One of the recurrent 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 common practice among related applications (see section 2) to the one proposed in this paper.

Our approach simplifies this task, allowing the user to annotate images in an automatic manner, using audio elements and visual features (see section 5).

When people share their photos with friends they use the slideshow and provide comments or stories about the photo that is been visualized to better explain the captured moment and the related context (see figure 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 figure 4).

This recording is analysed 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.

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, query by concepts and query by composition. Both types of queries are built using the query language proposed. Concepts, logical operators, temporal and geographical items, and images parts are

Figure 4. Audio annotation in slideshow visualization mode.

the elements used to build the query using the drag and drop technique to a query box. The following subsection presents the query language and explains how it is used to retrieve images.

4.4.1 Query language

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

$$\textit{Loperator} = \{AND, OR, NOT\}.$$

As normally, 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 her search among a universe of pre-established set of concepts:

Figure 5. Drag and drop to a query box.

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}*.

Visual elements are obtained by cropping sections in 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 have nature in them (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 into more valuable memories.

To retrieve images using the query by concepts, the user indicates what concepts are relevant for her search among the set of predefined concepts. When establishing the concepts she wants to use 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,

our system returns a set of images with a hi-probability to depict those concepts. These images can be visualized or serve as input for the other type of query.

In order to accomplish a query of this kind, concepts and

Figure 6. Query by image composition.

operators are chosen by means of a drag and drop operation into a query box (see Figure 5). As opposed to similar applications or normal database querying there is no

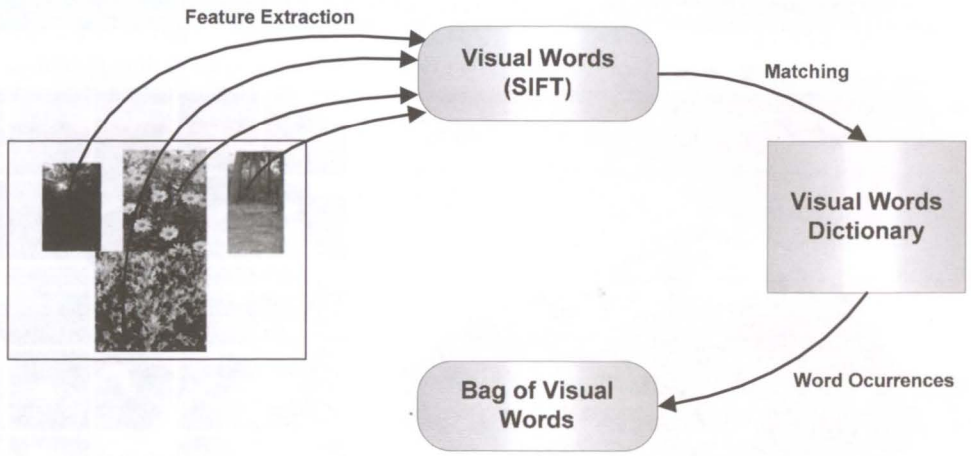

Figure 7. Bag of visual words obtained from de query defined by image composition.

typing at all. The user merely drags operators and concepts she wishes to express 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.

Figure 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 to retrieve the claimed memories is query by composition. This method allows composing an image with other images parts (see Figure 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.

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

5. RETRIEVAL SYSTEM

The user interface described in this paper allows querying the 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 system is weaker.

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 classifier. For more details see [Jesus06].

The image composition query works in a similar manner of the query by sketch [Buijs99] 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 what she really wants to 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) [Deerwester90] is applied and the cosine distance is used to retrieve relevant images.

5.1 Latent Semantic Analysis

The visual content of each picture of the database is represented by a bag of visual "words" (see figure 7) that express the number of times the visual "words" occur in an image. The visual words belong to a visual vocabulary obtained by applying the k-means method to a large set of SIFT (Scale Invariant Feature Transform) [Lowe04] descriptors extracted from all images of the database. Then, the database is represented by the matrix term-document that describes the occurrences of the visual "words" in images. Following this, we apply the term frequency – inverse document frequency (td-idf) weight to each element of the matrix. Finally, we apply the LSA to the matrix term-document. Given a query, the system calculates a new bag of visual words that represents the query and then ranks the database using the cosine distance.

6. RESULTS

To evaluate the application proposed two types of evaluations should be done: tests to evaluate the results obtained by the multimedia retrieval system and tests to evaluate the user interface. At this stage we only present results concerning the evaluation of results of the multimedia retrieval system. However, the Memoria project includes a similar interface that explores personal memories on a mobile device and this interface has already been evaluated. From those results, we can infer some ideas about the option of using the drag and drop technique and a query box to define queries using different items. With that evaluation we conclude that after a period of initial adaptation, users feel comfortable using the

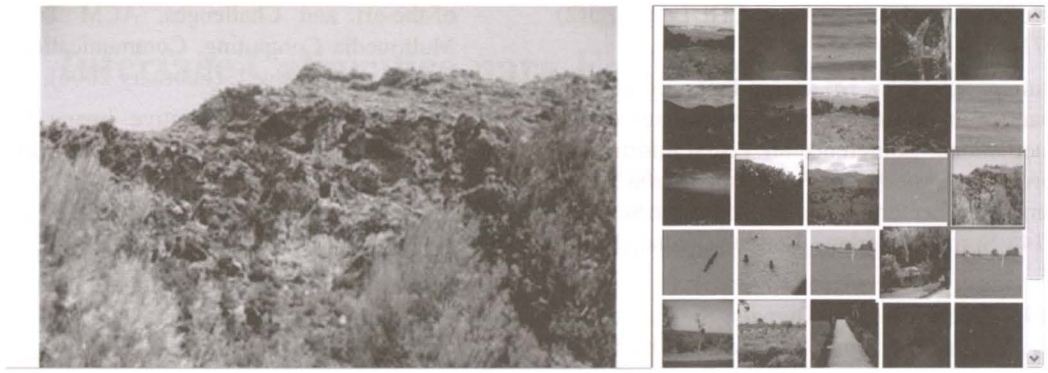

Figure 8. Query by concepts - Results obtained using the query, *Outdoor AND Nature.*

drag and drop technique and the query box (two important features of our interface). However, we are aware that the interface proposed differs from a mobile application, so these conclusions should be taken just as a guiding point.

The desktop interface allows querying the database in two modes: query by concepts and query by image composition. To test these functionalities a personal collection of about 5000 photos was used. These personal memories are essentially composed by pictures of people, nature or urban scenes, holidays and parties. The semantic concepts were evaluated in [Jesus06]. The concepts indoor, outdoor and people presented the best results while party and snow were the worst (see table 1). This happens because party and snow are rare events in the database. Figure 8 shows the results obtained by the query, *Outdoor AND Nature*.

Query	Color	Texture	Color Texture
Indoor	0,61	0,64	0,70
Outdoor (NO Indoor)	0,82	0,79	0,86
People	0,77	0,77	0,78
Party	0,11	0,13	0,12
Outdoor AND Beach	0,41	0,17	0,36
Outdoor AND Snow	0,11	0,03	0,04
Indoor AND People AND Party	0,15	0,18	0,20
MAP	0,42	0,39	0,48

Table 1. Mean Average Precision (MAP) – query by concepts.

To evaluate the query by composition, four cases were tested and compared with the results obtained by each image used to compose the query. The four cases are: to search for buildings; to search for plants; to look for snow images and to find images of a specific house To analyze the performance of this technique, Table 2 presents the precision for both situations. In general, the image composition query obtains results that are very close with one of the images used to build the query (Image 1 column in table 2). However, the query by image composition presents relevant images with more diversity. At this stage, we are using only texture features but we will include more features (e.g., color or shape) to improve the results of this method.

7. CONCLUSIONS AND FUTURE WORK

An interface to explore personal memories was presented. It allows browsing and visualizing memories, capturing new media items, the 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 enables the management of personal memories in a way that is fast and also requires a minimal interaction from the user.

	Image Composition	Image 1	Image 2	Image 3
Buildings	0,56	0,45	0,44	0,33
Plants	0,81	0,85	0,69	0,37
Snow	0,13	0,13	0,05	0,04
House	0,19	0,15	0,11	-
Mean	**0,42**	**0,40**	**0,32**	**0,25**

Table 2. Precision obtained by several queries by image composition considering 100 returned images.

Future work for this project includes a more thorough evaluation for the proposed interface, the construction of a tangible interface and the use of contextual information, such as GPS, date and EXIF metadata.

8. ACKNOWLEDGEMENTS

The authors thank the Microsoft Language Development Center in Portugal for providing the language pack with Speech API used for speech recognition.

9. REFERENCES

[Bederson01] Bederson, B. PhotoMesa: A Zoomable Image Browser Using Quantum Treemaps and Bubblemaps. UIST 2001, ACM Symposium on User In-

terface Software and Technology, CHI Letters, 3(2) (2001), 71-80.

[Buijs99] Buijs, J. and Lew, M. Visual Learning of Simple Semantics in ImageScape. In Proceedings of the Third international Conference on Visual information and information Systems. D. P. Huijsmans and A. W. Smeulders, Eds. Lecture Notes In Computer Science, vol. 1614. Springer-Verlag, London, (June 1999), 131-138.

[Deerwester90] Deerwester, S., Dumais, S., Furnas, G., Landauer, T., and Harshman, R., Indexing by Latent Semantic Analysis. in Journal of the Society for Information Science (1990), 391-407.

[Flickner95] Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D., and Yanker, P. Query by Image and Video Content: The QBIC System. Computer 28, 9 (Sep. 1995), 23-32.

[Frohlich02] Frohlich, D., Kuchinsky, A., Pering, C., Don, A., and Ariss, S., "Requirements for Photoware," Proc. of the ACM conference on Computer supported cooperative work, (2002), 166-175.

[Kuchinsky99] Kuchinsky, A., Pering C., Freeze, D., Creech, M., Serra, B., Gwizdka, J. FotoFile: A Consumer Multimedia Organization & Retrieval System. Proceedings of the ACM Conference on Human Factors in Computing Systems CHI'99, (1999), 496-503.

[Jesus06] Jesus, R., Abrantes, A., Correia, N. Photo Retrieval from Personal Memories using Generic Concepts. Advances in Multimedia Information Processing - PCM 2006, Springer LNCS, 4261. 633-640.

[Lew06] 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, Vol 2, No. 1, (February 2006), 1-19.

[Lowe04] Lowe, D. Distinctive image features from scale-invariant keypoints. In International Journal Computer Vision, (2004), 91-110.

[Moghaddam04] Moghaddam, B., Tian, Q., Lesh, N., Shen, C., and Huang, T. Visualization and User-Modeling for Browsing Personal Photo Libraries. Int. J. Comput. Vision 56, 1-2 (Jan. 2004), 109-130.

[Mori1999] Mori, Y., Takahashi, H., and 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).

[O'Hare05] O'Hare, N., Gurrin, C., Jones, G. and 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, (November 2005), 323-328.

[Pereira2001] Pereira, F., Koenen, R. MPEG-7: A standard for multimedia content description. In International Journal of Image and Graphics 1(3) (2001). 527-546.

[Veltkamp02] Veltkamp, R., Tanase, M. A Survey of Content-Based Image Retrieval Systems. In: O. Marques, B. Furht, Content-Based Image and Video Retrieval, Kluwer, (2002), 47-101.

[Retrievr] http://labs.systemone.at/retrievr/

[Phlog] http://phlog.net/

[Wwmx] http://wwmx.org/