A System to Capture, Share and Access Personal Memories

Rui M. Jesus[1,2], Tiago Martins[2], Rute Frias[2], Arnaldo J. Abrantes[1], Nuno Correia[2]
[1]Multimedia and Machine Learning Group, Instituto Superior de Engenharia de Lisboa
[2]Interactive Multimedia Group, DI/FCT, New University of Lisbon
Quinta da Torre, 2825 Monte da Caparica, Portugal
http://img.di.fct.unl.pt
rjesus@deetc.isel.ipl.pt, nmc@di.fct.unl.pt

ABSTRACT
This paper presents a system to access personal memories composed by digital pictures. The system consists of a retrieval engine, a desktop interface to share personal memories and a mobile user interface that allows capture and automatic annotation of images. The retrieval engine uses Global Positioning System (GPS) location data, low-level visual features and semantic concepts previously trained to retrieve images. With the mobile interface people can capture, share personal pictures and navigate in the physical space when visiting historical sites, museums and other touristic activities using their Personal Digital Assistants (PDA). The visitors can take photos and submit them to the system to receive contextually related photos taken by others or themselves. Experimental results are presented to show the performance of the retrieval mechanisms and the usefulness of the navigation system.

Keywords: Personal Memories, User Interfaces, Multimedia Information Retrieval.

1. INTRODUCTION

Sharing memories and experiences is a major activity of human beings that has been practiced for thousands of years, enabling the exchange of knowledge and information across generations and cultures. Pictures and videos are rich vehicles to transport this information. Recent technological advances of the digital devices have been changing the concepts of capture, share and store of visual information (e.g., camera-phones with integrated GPS and considerable storage capacity). People can take photos or make small video clips of everything, everywhere and, through the World Wide Web share this information with friends. Consequently, a vast amount of personal digital information is being produced and stored.

Stored personal information can play an important role in supporting daily activities. This information can help a doctor in the diagnosis process [10] or people remembering important things about their lives (e.g., meetings, trips and holidays). Some applications (e.g., to find lost objects) requires continuous or passive capture of personal photos or videos [3, 4, 8], but other applications related with personal memories do not need this kind of capture, in particular the applications related with leisure activities like holidays, visiting a museum or a birthday party. Sometimes people want to recall a previous moment or the people they were with while at a given site. This is usually done in a Personal Computer (PC), but when visiting historical sites, museums and doing other touristic activities people may want to share their pictures or to browse for previous photos of the same place, possibly captured by other people. Additionally, these images taken by others can help in choosing the path for a given visit. In both situations, an efficiently multimedia information retrieval is essential in order to access personal information.

Several applications and interfaces for accessing this information in a PC have been proposed. Commercial applications (e.g., Adobe Photoshop Album, Paint Shop Pro and Picasa) and online sites like www.flickr.com or www.phlog.net are available to manage personal memories with pictures. Most of these applications use manual annotations to search photos. The manual annotation is the most effective way but it is time consuming. Ethnographic studies [2] show that people usually do nothing to organize their photographs and only occasionally they create an album of a special event. Automatic systems rely on context metadata or in visual content [1, 6, 9]. Most of these systems use Content Based Image Retrieval (CBIR) techniques combined with context metadata in order to bridge the semantic gap [11], a usual problem in the CBIR systems [7] due to its difficulties in capture semantic concepts e.g., flowers or persons, using the low-level features extracted from the images.
This work, described in more detail in the next sections, is a system to capture, share and access personal memories composed by digital pictures. PhotoNav consists of two user interfaces, a desktop interface and a mobile user interface. Both interfaces are based on a retrieval system that uses visual content and context metadata.

2. SYSTEM OVERVIEW

PhotoNav is based on a desktop interface, in a mobile interface and in a retrieval system that runs in a server. The retrieval system uses Global Positioning System (GPS) location data, low-level visual features and semantic concepts previously trained to retrieve images. All the information and associated metadata is stored in a server that also does the necessary image processing for finding similar images (retrieval system).

The desktop interface allows searching and sharing large image databases using semantic concepts. Semantic concepts are obtained by training binary classifiers (indoor/outdoor) using the Regularized Least Squares Classifier (RLSC) and can be combined to express more complex concepts. To combine several generic concepts the sigmoid function is applied to the output of RLSC. The images are ranked according to their probability of belong to the classes of the chosen concepts. This method was evaluated for several concepts suitable for personal memories (e.g., outdoor, people, nature) in our previous work [5] with good results.

The mobile application is based on an interface where the user can capture pictures and formulate several types of queries to retrieve photos from the server. The personal information returned can help the user navigation in the physical space (e.g., when visiting historical site or museums). This application uses low-level features, semantic concepts and context metadata to query the database. It support three types of queries:

1. Query by image - visitors want to see images similar to the query, which can be a picture taken by them. Low-level

features (color and texture) are extracted from the query. Then, the Manhattan distances between this vector and the feature vectors of each image in the database are calculated. Database pictures are ranked according to this distance and the user can see the top 10 images in the PDA interface.

2. Query using a direction and the GPS data - visitors want the nearest image in a given direction. Only the GPS information is used. The distance between points is calculated using the Great Circle distance. This distance represents the shortest distance between two points on the surface of a sphere and has reasonable accuracy. We use the spherical law of cosines to obtain the distance. The nearest image according to this distance is returned.

3. Query using a direction, the GPS location data and the semantic concepts - visitors want one image in a direction with a given context defined by a set of semantic concepts. We use GPS data to obtain a subset of images that are in the desired directions. Then, these images are ranked according to their probability of belonging to the classes defined by the concepts. The image with higher probability is returned.

3. EXPERIMENTAL RESULTS

The semantic concepts were evaluated in the desktop interface using the personal collection of one person (Rui Jesus) with 818 images and a set of pictures shared by his friends in a total of 2582 photos. Personal memories are essentially composed by pictures of people, nature or urban scenes, holidays and parties. Five binary classifiers for semantic concepts suitable to search in a personal collection were trained: people versus no people; indoor versus outdoor; snow versus no snow; beach versus no beach; party versus no party. The concepts indoor, outdoor and people presented the best results while party was the worst (see [5] for more details).

The mobile interface was tested with a database of 1255 pictures of Quinta da Regaleira, a cultural heritage site in Sintra, Portugal. To evaluate the first type of query, six images were selected to query the database. These images represent several items that can be found in Quinta da Regaleira: people, palace, nature, tower, lake, mosaic. Image retrieval systems for this type of applications in mobile devices should have high precision because the screen is small, only few images can be shown and the user does not want to waste time of her visit navigating over non relevant images. For these reasons this functionality was evaluated calculating the precision in the first 10 relevant images. To look for specific landmarks of Quinta da Regaleira such as “Palace” the system has a good performance.

To test the second type of query, three routes were done starting in different places of Quinta da Regaleira. When no GPS errors occur the system has a good performance.

The third type of query also requires a high precision. Several tests were done using several contexts. The contexts “outdoor” and “outdoor+nature” present the best results because most of the images in the database are outdoor images with nature elements. The worst result was obtained with the concept “people”, because we have few images in the database with people.

4. CONCLUSIONS AND FUTURE WORK

The paper presents a system for navigating and browsing in digital memories in a PC and while at the physical locations (e.g., historical or cultural heritage sites) using mobile devices. It uses image processing techniques to get related images, either by visual similarity, proximity or related concepts. A GPS provides location information which is then combined with queries done using images or semantic concepts, to retrieve related images with the ones the user is capturing.

Additional future work includes adding more concepts that can be used in the queries, which implies evaluating additional features in the image database. Other developments include developing additional interfaces used in other situations, for example, augmented reality settings and desktop versions targeted at elder users.

REFERENCES

[1] Cooper, M., Foote, J., and Girgensohn, A., Automatically organizing digital photographs using time and content. in *Proc. of the ICIP* (2003).

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

[3] Gemmell, J., Williams, L., Wood, K., Bell, G., and Lueder, R., Passive capture and ensuing issues for a personal lifetime store. in *The ACM Workshop on CARPE '04*, (2004), 48-55.

[4] Hori, T., and Aizawa, K., Context-based video retrieval system for the life-log applications. in *Proc. of the ACM SIGMM International Workshop on Multimedia Information Retrieval* (2003), 31-38.

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

[6] Jiebo, L., Boutell, M., and Brown, C., Pictures are not taken in a vacuum - an overview of exploiting context for semantic scene content understanding. in *Signal Processing Magazine, IEEE*, (2006), 101-114.

[7] Lew, M., Sebe, N., Djeraba, C., and Jain R., Content-based Multimedia Information Retrieval: State-of-the-art and Challenges. in *ACM Transactions on Multimedia Computing, Communication, and Applications*, (2006).

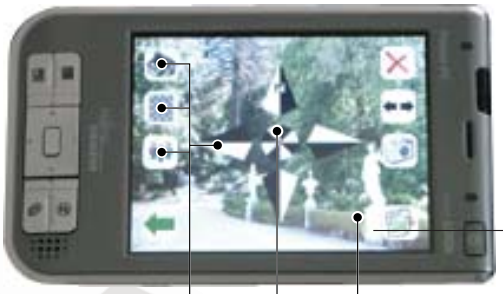
[8] Mann, S., “Smart Clothing”: Wearable Multimedia Computing and Personal Imaging. in *Proc. of ACM Multimedia '96, Boston, Mass., USA, Nov*, (1996), 163-174.

[9] O’Hare, N., Jones, G., Gurrin, C., and Smeaton, A., Combination of content analysis and context features for digital photograph retrieval. in *IEE European Workshop on the Integration of Knowledge, Semantic and Digital Media Technologies*, (2005).

[10] Pratt, W., Unruh, K., Civan, A., and Skeels M. , Personal health information management. in *Communication ACM*, (2006), 51-55.

[11] Smeulders, A., Worring, M., Santini, S., Gupta, A., and Jain, A., Content-based image retrieval at the end of the early years. in *IEEE Trans. Pattern Analysis and Machine Intelligence*, (2000), 1349 -1380.

Mobile interface
Capture, search and share images between people visiting the same place



Query 4 Find images with user-defined semantic concepts

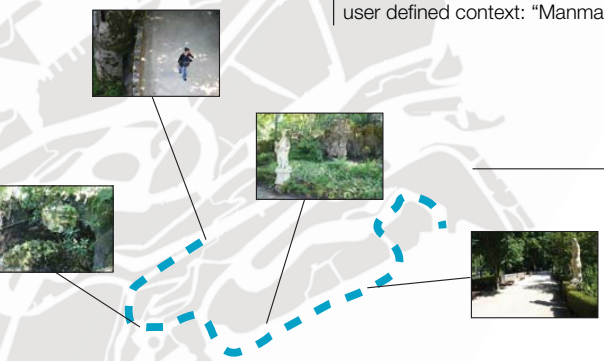
The retrieval system uses visual content and context metadata (GPS), either by visual similarity, proximity or related concepts

PhotoNav Navigation Mode

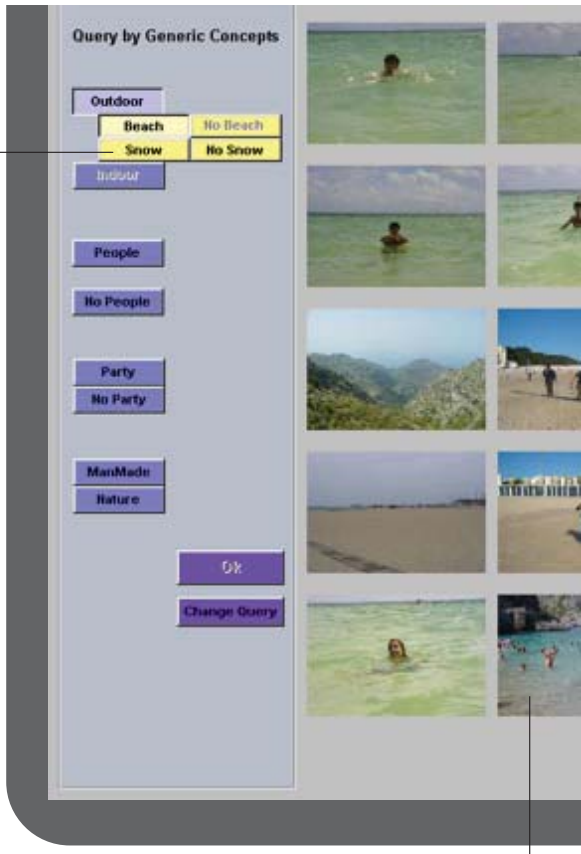
Query 1 Find images of the same place

Query 2 Guide the visit
Retrieve one image of the nearest place (GPS information and direction)

Query 3 Direction and user defined context
Retrieve one photo of a place in a direction according to the user defined context: "Manmade", "Outdoor" and "People"



Personal memories can be accessed based on the user location and path



Desktop interface
Search and share large image databases using semantic concepts previously trained

Results (photos) for a query with the semantic concepts "Outdoor" and "Beach"

PhotoNav is a system to capture, share and access personal memories composed by images and contextual information. PhotoNav has two user interfaces: a desktop and a mobile interface

Challenges

Humans love to collect and store information about their lives and digital images are widely used

Main challenge is the image/video retrieval task

Manual annotation with keywords is not an easy task and it is difficult to describe by keywords the image content

Automatic annotation systems use low level features that are unable to capture semantic concepts (semantic gap)

Difficulty in handling complex mobile interfaces

Improving image visualization in digital screens

Benefits to the user

Memories guide the visit and the recollection

Memories access while visiting a place

Access to personal images based on memoirs like place, visual similarity, concepts, date, events, and also based on the physical context

Users do not have to annotate images

Pleasing recollection experience

Research Stage

Desktop and mobile interfaces first working prototypes

The retrieval system was evaluated resulting in a mean average precision of around 0.50 with 2582 images

Future work includes additional semantic concepts, interfaces for the elderly population, tangible and augmented reality applications

Rui M. Jesus rjesus@deetc.isel.ipl.pt
Multimedia and Machine Learning Group, Instituto Superior de Engenharia de Lisboa
Interactive Multimedia Group, DI/FCT, New University of Lisbon, http://www.img.di.fct.unl.pt

Nuno Correia nmc@di.fct.unl.pt, **Tiago Martins, Rute Frias**
Interactive Multimedia Group, DI/FCT, New University of Lisbon, http://www.img.di.fct.unl.pt

Arnaldo J. Abrantes
Multimedia and Machine Learning Group, Instituto Superior de Engenharia de Lisboa