Sharing Personal Experiences while Navigating in Physical Spaces

Rui Jesus[1,2], Ricardo Dias[1], Rute Frias[1], Arnaldo J. Abrantes[2], 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

ABSTRACT
Social networks popularity has been increasing on the Web over the last years. These web sites provide an easy way to share digital data between people with common interests. With this way of exchanging memories of personal experiences, large repositories of data are being created. Indeed personal media, like photos and videos, are excellent vehicles to exchange experiences, however to manage large multimedia databases suitable tools are needed. This paper describes a multimedia retrieval system to access the personal memories shared by the community that visits a point of interest. The multimedia retrieval system uses multimodal information: visual content, GPS data and audio information annotated at capture time. This system can be used via a Web site or when people are visiting the place. The multimedia retrieval system was evaluated using a mobile interface in a cultural heritage site where the personal media can be shared by visitors and can be used to guide the visit. Experimental results are presented to illustrate the effectiveness of the system.

Categories and Subject Descriptors
H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval; H.5.1 [**Information interfaces and presentation**]: Multimedia Information Systems – Evaluation/methodology.

General Terms
Algorithms, Performance, Design, Experimentation, Human Factors, Standardization, Languages.

Keywords
Social Media Applications, Personal Memories, Multimedia Information Retrieval

1. INTRODUCTION
Currently, due to the success of the World Wide Web, forming communities of people that have common interests is an accessible goal. Many people, that almost certainly would never know each other without the advent of the Internet, are now exchanging media using photoblogs, photo sharing websites or other online networks. Flickr or YouTube are examples of sites where users share personal experiences by publishing media that attracts other people.

When visiting historical sites, museums or other points of interest people usually produce numerous images and can also be engaged to form a virtual community to share those images. Within this community people can increase their personal collection as well as build an interesting repository of shared memories related to that physical space. This sharing can be done in the usual manner, by taking the time at home to upload the media items for a web site. But it can also be done when people are visiting the place using current mobile technology and introducing a form of communication that shapes the experience of the visit.

Sharing of pictures in a mobile scenario and the amount of information to be managed introduces new challenges. As this information grows more knowledge about the place is shared but also more difficulties in managing this information arise.

Software tools to manage this information must be developed in order to help the task of accessing shared memories using a mobile device or a Web site.

This paper describes the multimedia retrieval system used in the Memoria project to access personal memories composed by pictures or videos shared by visitors when visiting a site of interest. The main goal of this project is to build a system to explore personal memories. Currently, the project includes a mobile user interface, a desktop Personal Computer user interface (a web interface will be developed) and a multimedia retrieval system that runs in a server. This system uses multimodal information: visual content, GPS data and audio information annotated at capture time (using the mobile interface). The sharing of pictures can occur when the user requests information using the multimedia retrieval system or when the system returns pictures taken by others. The shared pictures can be used to guide the visit.

The novel aspect of this work is related with the multimedia retrieval system that integrates audiovisual information with GPS to retrieve images from a collective repository built by visitors of a point of interest (e.g., a museum or a historical site).

The paper is structured as follows. Next section presents the related work and the following gives an overview of the system. The subsequent section describes the multimedia retrieval system used in this application. After that, the experimental results are

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
SIGIR'07, July 23–27, 2007, Amsterdam, The Netherlands.
Copyright 2004 ACM 1-58113-000-0/00/0004…$5.00.

presented and discussed. The paper ends with the conclusions and directions for future work.

2. RELATED WORK

The work proposed in this paper addresses the problem of accessing personal memories composed by images and videos. Thus, this section presents related proposals for managing personal memories in desktop personal computers and in mobile devices.

Usually, people use desktop interfaces to navigate and share their digital photos with their family or friends. Nevertheless, the available commercial applications (e.g., Adobe Photoshop Album, Picasa and Photofinder) as well as the online sites (e.g., www.flickr.com, www.phlog.net, www.youtube.com) only use manual annotation. Generally, they use directories to organize pictures and some of them allow visualizing the directories chronologically.

Automatic systems rely on visual content or context metadata. Content Based Image Retrieval (CBIR) [1] systems use low-level information extracted from the image. Many visual features have been proposed. Color histograms, color moments, Tamura features and Gabor filter features are examples of the most used features in CBIR (see [1, 2] for more details). Additionally a set of well tested visual descriptors were proposed for inclusion in the MPEG-7 standard [3]. Some of the features are global features but these representations are unable to detect local details. Scale and rotation invariant local descriptors based on the detection of interest points were proposed by several authors [4-6]. These methods can be divided in two steps: detection of keypoints and representation of the regions around them. The SIFT descriptor is one of the most used for region representation mainly due to its invariant properties. This low-level information provides some degree of automatic annotation however it is unable to capture semantic concepts [2]. These difficulties can be overcome by the previous training of semantic models and then, automatically associate textual descriptions to the images [7]. Other systems [8-10] attempt to solve the semantic gap by combining visual content with context metadata (e.g., Time or GPS information).

When people are in places where computers are not available, they can share their pictures with friends using mobile devices. There are also available commercial applications (e.g., HP Image Zone and ACDSee) to manage images in mobile devices. Pocket PhotoMesa [11] is an interface to browse images in small displays. It employs Treemap layout to view hierarchies of image directories and provides zoomable interfaces for navigation. To search for personal pictures, using context information several systems were proposed, including the MediAssist [12] which uses GPS information and time. In [13] it was proposed other system that uses content and context metadata. This system automatically stores location, time, and user data in addition to the picture taken in a server and uses this information shared by several users to annotate images in a semi-automatic way. It also uses image features to recognize faces and to identify the picture location.

Our work has similarities with some of these systems, however we also use audio information, different visual characteristics, and the application interface has different features, including the ability to share images when visiting sites of interest.

3. SYSTEM OVERVIEW

This paper presents the multimedia information retrieval system used in the Memoria project to explore information stored in a shared repository of images and videos built by the community that visits a site of interest (e.g., a museum or a cultural heritage site). The main goal of this project is to build a system to explore personal memories (see figure 1). The project includes several interfaces adapted to different contexts of use, including a Web interface and another for mobile devices (see figure 2) to explore personal collections related with physical locations. The project also includes a multimedia retrieval system that uses visual information as well as GPS data and audio information, automatically captured when the picture was taken.

The Memoria project is based on a client/server architecture. Users can access personal pictures by means of a client program running on a mobile device. This application can be used as a regular camera, with the additional advantage that it captures GPS data (labeling the images with location information) and audio information when the picture is taken. The multimedia retrieval system runs in a server accessed via wireless connection (WLAN).

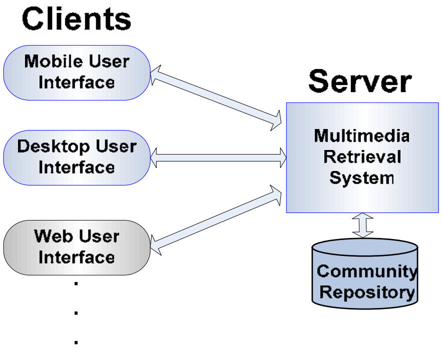

Figure 1. Memoria Project Architecture.

The proposed system was evaluated in a cultural heritage site.

Figure 2. Layout of theMemoria mobile interface

After taking a picture the visitor may desire to see similar images, nearby pictures or even close images according to a user-defined context (e.g., images with people or with buildings). The visitor can use the mobile application (see figure 2) to take pictures and the multimedia retrieval system to share them with other visitors or other members of community. The results returned by the retrieval system can help in choosing the path for a given visit.

Users can submit three types of queries to the retrieval system: query by image, query by semantic concepts and query using GPS location data. Additionally the user can combine the last two types of queries to retrieve nearby images given a context (e.g., outdoor images, with nature elements and with people). When the visitor request information to the server (submitting a query) she is sharing her information with the community (query information sent). The returned results (images shared by the community) by the multimedia retrieval can increase the personal collection and guide the visit.

The Memoria mobile application (see figure 2) has three main functions: picture capture, browsing and retrieval of photos from a database. The interface contains a title, status and navigation bars; list menus; toolboxes and a search box. The generic "Menu" lists all functions and options whereas "Filter" aggregates every item available to query the database (e.g., concepts, directions). The "Search Box" is used to define queries by dragging and dropping filters into the container space.

4. MULTIMEDIA RETRIEVAL SYSTEM

The multimedia retrieval system used in the Memoria project uses multimodal information to represent each image of the database: visual content, GPS data and audio information annotated at capture time. Users can submit three types of queries to the retrieval system: query by image, query by semantic concepts and query using GPS location data. Additionally the user can combine the last two types of queries to retrieve nearby images given a context (e.g., outdoor images, with nature elements and with people). The following subsections explain the system behavior for each type of information.

4.1 Image Query

To take a picture and using it to retrieve more information about the local, is a common scenario when visiting a museum [14] in the context of the augmented reality. The proposed system follows this idea but the returned information are images shared by other visitors. Each image is represented by a bag of features which recently have presented a good performance in the context of image classification [15]. Then, the Latent Semantic Analysis (LSA) [16] is applied to obtain a term-document matrix with lower rank like in text retrieval.

4.1.1 Bag of Features

Given a database composed by a set of N images $D_B = \{I_1 ... I_N\}$ and a vocabulary $V_w = \{w_1, ..., w_M\}$ of M "words" extracted from the database D_B, the bag of features for an image k is represented by a histogram that counts the number of occurrences, $N_{lk} = n(w_l, I_k)$ of each w_l word in the I_k picture,

$$h_k(l) = \sum_{i=0}^{M-1} N_{ki} \delta(l - i)$$

where $\delta(l)$ is the impulse unitary function.

The vocabulary V_w can be composed by words, visual features, context features (e.g., GPS or time) or any other useful information to better understand the image content. At this time the system only uses visual features to build the vocabulary:

- Texture vocabulary consists of a set of clusters centers obtained applying the k-means method to a large set of SIFT (Scale Invariant Feature Transform) descriptors [5] extracted from all images of the database. From each image is extracted 1000 descriptors approximately and each texture word is represented by a vector of 128 elements.
- Color vocabulary is composed by a set of vectors that describes the most relevant color regions that appears in the database. The k-means is applied to find these regions. In each image, segmentation is performed using the Mean Shift (MS) algorithm [17]. Each region is described by the mean and variance of the color pixels in LUV color space, the percentage of pixels in the region and the coordinates of the middle pixel.

Given the vocabulary V_w the bag of features that represents each image is obtained by applying the Euclidean distance between the features extracted and the *w* words (clusters). If the distance fall below a threshold the occurrence of the w_l in the image I_k is incremented. After that, the database is expressed by an occurrence matrix, X of MxN elements, known as term-document matrix in text analysis and the term frequency – inverse document frequency (td-idf) weight is applied to each element of the matrix.

4.1.2 Image Retrieval with Latent Semantic Analysis

To discover the latent relationships between correlated visual features (words of the vocabulary) and images the Latent Semantic Analysis is applied. This method finds a low rank of the X matrix and removes the instance of terms less relevant by using the Singular Value Decomposition (SVD). The number of singular values used was obtained empirically.

To retrieve the relevant images given a query, the system represents the bag of features of the query in the same space of the database and ranks the database using the cosine distance.

4.2 Semantic Concepts

The most intuitive way to define queries to search images is by using concepts (keywords). Semantic concepts used in the proposed system are trained using the Regularized Least Squares Classifier [18]. The images used to train were obtained from some CD's of the Corel Stock Photo, from the TRECVID2005 database and from www.flickr.com, in order to build a more generic training set.

Given the training set $S_m = \{(x_i, y_i)_{i=1}^m\}$ where labels $y_i \in \{-1,1\}$ and x_i is a vector of image features, the decision boundary between the two classes (e.g., indoor, outdoor) is obtained by the discriminant function,

$$f(x) = \sum_{i=1}^{m} c_i K(x_i, x)$$

where $K(x, x')$ is the Gaussian Kernel $K(x, x') = e^{-\frac{\|x - x'\|^2}{2\sigma^2}}$, m is the number of training points and $c = [c_1, ..., c_m]^T$, is a vector of coefficients estimated by Least Squares [18],

$$(m\gamma I + K)c = y$$

where I is the identity matrix, K is a square positive definite matrix with the elements $K_{i,j} = K(x_i, x_j)$, y is a vector with coordinates y_i and γ is a regularization parameter. To choose the optimal values for σ and γ the cross-validation method is used.

A point x with $f(x) \le 0$, is classified in non relevant class $(y = -1)$, and a point with $f(x) > 0$ is classified in relevant class $(y = 1)$.

The output of the classifier can be used to rank the database, however when several concepts are combined we need to convert the output to a pseudo probability p. Assuming w is a binary random variable where the outcome can be one of two concepts (e.g., indoor/outdoor). The probability $p(w/x)$ can be obtained using the output of the function $f(x)$ and the sigmoid function in a similar manner of [19],

$$p(w/x) = \frac{1}{1 + e^{-Af(x) + B}}$$

In [19] are discussed several methods to estimate A and B parameters. We set them manually but in the future they will be estimated.

Considering a query formed by k independent concepts $Q = \{w_1, w_2, ..., w_k\}$ that describes the background and some objects presented in the desired pictures (e.g., indoor, people and computers). Each w_i is a binary random variable and the probabilities $p(w_i/x)$ are obtained using the previous equation with different classifiers $f_i(x)$. Each image in the database can be represented by several low-level features (e.g., color and texture) and also for each feature a different classifier is obtained.

Given a query formed with k concepts and using r low-level features the rank of each image is obtained by the probability,

$$p(w_1, w_2, ..., w_k/x) = \sum_{j=1}^{r} a_j \prod_{i=1}^{k} p(w_i/x^j)$$

where x^j is a vector for the j^{th} feature, $0 < a_j < 1$ is the weight of each features and $\sum_{j=1}^{r} a_j = 1$.

This method was evaluated for several concepts suitable for personal memories in our previous work [20] with good results.

4.2.1 *Audio Information*

The proposed system automatically annotates GPS data and audio information at capture time. Instants after the picture capture, the system opens the microphone to save some possible user comments. The idea is to record some comments that the people usually do when they are visiting a point of interest. This audio information is converted to text using ASR (Automatic Speech Recognition) tools. Then, each image is represented by a bag of words and the Latent Semantic Analysis is applied as explained in subsection 4.1, but using words that were recognized. The database is ranked given a query defined by a set of keywords. The ranking obtained by the semantic concepts (visual content only) is combined with the ranking given using audio information. The final ranking is obtained by summing the ranking obtained by the semantic concepts and the audio ranking.

4.3 Query with GPS Information

Queries with GPS information can guide the visit. Using a location provided by GPS data, the visitor can see a set of nearby images or a set of images in a given direction. When concepts are added to the query, the visitor can see a set of filtered images (nearby images or in a direction) according to the context defined by the user (e.g., outdoor, people, nature). Using this query the visitor can improve his knowledge about the site being visited. The distance between locations 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.

5. EXPERIMENTAL RESULTS

This aplication was tested with a database of about 1500 pictures of Quinta da Regaleira, a cultural heritage site in Sintra, Portugal. These pictures were taken by several members of our research group during visits to Quinta da Regaleira in the scope of a mobile storytelling project. Some images are annotated with GPS, a subset of all is annotated with keywords extracted from the audio information and all of them are represented by visual features. This heterogeneity in the way data is represented is common in a collective collections and the system should be able to deal with that constrain. Several tests were made to evaluate the retrieval system using the mobile device (PDA), Fujitsu Siemens Pocket Loox 720.

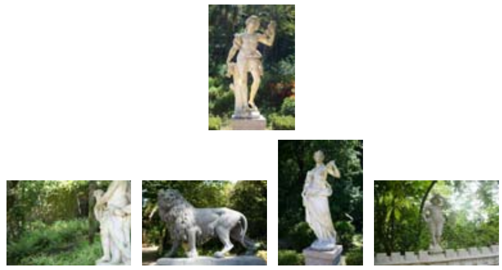

Figure 3. Results obtained using the bag of feature for the item sculpture using the query at top.

The queries that use only GPS information, to retrieve nearby images or images in a given a direction, presents good results unless when there are errors on the GPS data.

To evaluate the query by image (retrieve similar pictures), four images for each of the following items were selected to query the database: people (226 images with people in the database), palace (102), nature (869), tower (57), sculpture (265) and tiles (31). For each item the mean of the precision obtained with the four images of each item was calculated. Image retrieval systems for mobile devices should, above all, have high precision because the screen is small and only a few images can be shown. Additionally, in a mobile usage scenario, there is usually little tolerance for managing non-relevant images. For these reasons, the query by image was evaluated by calculating the precision in the first 10 returned images. Figure 3 shows an example of the results obtained for the item sculpture using the bag of SIFT's.

Images (Queries)	Bag of Gabor	Bag of SIFT	LSA SIFT
People	0,50	0,55	0,55
Palace	0,70	0,78	0,80
Nature	0,95	0,93	0,85
Tower	0,25	0,35	0,40
Tiles	0,10	0,17	0,30
Sculpture	0,37	0,45	0,60
Mean	**0,48**	**0,54**	**0,58**

Table 1. Texture results - precision using images queries and considering 10 returned images.

To evaluate the bag of features (subsection 4.1.1), we compared the results obtained by the bag of SIFT's with the bag of Gabor filter features [21] obtained in the locations given by the SIFT keypoints. We also compared the bag of MS color regions with the Marginal HSV color moments [22]. The image is divided in 9 tiles and the color moments of each tile and his position represents each visual "words".

Images (Queries)	Bag of color Moments	Bag MS color regions	LSA MS color regions
People	0,45	0,45	0,50
Palace	0,25	0,53	0,50
Nature	0,95	1,0	1,0
Tower	0,13	0,18	0,15
Tiles	0,60	0,74	0,74
Sculpture	0,73	0,60	0,70
Mean	**0,52**	**0,58**	**0,60**

Table 2. Color results - precision using images queries and considering 10 returned images.

Table 1 compares the texture features and Table 2 the color features. In general, the results are very similar although the features explained in subsection 4.1.1 present slightly better results. We also evaluate the Gabor filter features as proposed in [21] and the Marginal HSV color moments as proposed in [22] to compare with their bags but the results obtained were similar.

Images (Queries)	Bag of SIFT + MS color regions	LSA SIFT + MS color regions
People	0,50	0,60
Palace	0,60	0,60
Nature	1,0	1,0
Tower	0,27	0,58
Tiles	0,60	0,87
Sculpture	0,73	0,78
Mean	**0,63**	**0,74**

Table 3. Color and Texture results - precision using images queries and considering 10 returned images.

When we used the bag of SIFT and the bag of MS color regions individually the advantage in using the LSA was small but when we combine them (by summing the individual ranks) the results presented by LSA are better and justifies its use (see table 3).

Table 4 presents the results obtained when using the query by semantic concepts. We evaluate five concepts: outdoor (1110 images in the database), indoor (314), people (226), manmade (846), nature (869). To compare the audio with the visual information, the precision was calculated considering 30 retrieved images. The visual features used were the Marginal HSV color moments and the Gabor filter features. When using the audio combined with the visual information the results are better than using only visual information and are similar with the results obtained using only audio. This shows the relevance of the audio in the system. As expected, the concepts with more images in the database present better results. The "Outdoor" and "Outdoor+Nature" concepts present the best results using only visual content. This demonstrates that the visual features used are suitable for these concepts.

Concepts (Queries)	Visual	Audio	Visual + Audio
Outdoor	1,0	0,90	0,90
Indoor	0,66	0,67	0,80
Nature	0,77	0,83	0,80
Manmade	0,80	0,93	0,80
People	0,24	0,46	0,57
Indoor + Manmade	0,27	0,56	0,56
Outdoor + Nature	1,0	0,80	0,90
Mean	**0,68**	**0,74**	**0,76**

Table 4. Precision for several concepts considering 30 returned images.

The results obtained by the query that uses GPS data, a direction, and a set of semantic concepts are presented in table 5. In this test, only the images in the desired direction are ranked by the concepts. Each concept was evaluated in three different locations. Results presented in table 5 are the mean of the precision obtained in these three locations considering 10 returned images. Again, the audio information improves the performance of the system. The results are similar with the results obtained without using GPS because the number of images to rank decreases but also the number of relevant images decreases.

Concepts (Queries)	GPS + Visual	GPS + Visual + Audio
Outdoor	0,8	0,87
Indoor	0,8	0,94
Nature	0,67	0,67
Manmade	0,74	0,84
People	0,34	0,5
Indoor + Manmade	0,47	0,5
Outdoor + Nature	0,67	0,77
Mean	**0,64**	**0,73**

Table 5. Images in a direction - precision for several concepts using GPS, audio and visual information and considering 10 returned images.

Table 6 shows the results obtained using a GPS location to define a set of nearby images and the semantic concepts. The results are inferior to the previous results with and without GPS because in some locations, for some concepts, there are few or none nearby images. For example, indoor places can only be found in a small part of this cultural heritage site.

Concepts (Queries)	GPS 60m+ Visual	GPS 60m + Visual + Audio
Outdoor	0,84	0,84
Indoor	0,40	0,40
Nature	0,64	0,74
Manmade	0,67	0,77
People	0,1	0,24
Indoor + Manmade	0,27	0,3
Outdoor + Nature	0,64	0,67
Mean	**0,51**	**0,57**

Table 6. Nearby images - precision for several concepts using GPS, audio and visual information and considering 10 returned images.

6. CONCLUSIONS AND FUTURE WORK

This paper presents the multimedia retrieval system of the Memoria project. This system uses multimodal information to retrieve, in different contexts and physical locations such as historical sites or museums, digital media items (mainly images) from personal memories shared by a community. Results that demonstrate the relevance of the audio information in the performance of the proposed system were presented. We also present results that show the performance of the Latent Semantic Analysis with images.

An important feature of these collective databases is their continuous growth which probably will decrease the performance of the multimedia retrieval system. However, with the increase of the database size more information obtained by previous searches is available as well as more annotated metadata. In the future, it would be interesting to evaluate the performance of the multimedia retrieval system using all this information. Another point is the image annotation that plays an important role in the performance of a multimedia retrieval system. Therefore, collective semi-automatic annotation should be provided since there are some errors with the automatic annotation that these collaborative environments can correct. Finally, video is also being considered and will be explored in future work.

7. REFERENCES

[1] M. Lew, Sebe, N., Djeraba, C., and Jain R., "Content-based Multimedia Information Retrieval: State-of-the-art and Challenges," ACM Transactions on Multimedia Computing, Communication, and Applications, vol. 2, pp. 1-19, 2006.

[2] A. Smeulders, Worring, M., Santini, S., Gupta, A., and Jain, A., "Content-based image retrieval at the end of the early years," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, pp. 1349 -1380, 2000.

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

[4] K. Mikolajczyk, and Schmid, C., "Scale and affine invariant interest point detectors," International Journal of Computer Vision, vol. 60, pp. 63--86, 2004.

[5] D. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal Computer Vision, vol. 60, pp. 91-110, 2004.

[6] N. Sebe, and Lew, M., "Comparing salient point detectors," Pattern Recognition Letters vol. 24, pp. 89-96, 2003.

[7] Y. Mori, 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.

[8] N. O'Hare, Jones, G., Gurrin, C., and Smeaton, A., "Combination of content analysis and context features for digital photograph retrieval," IEE European Workshop on the Integration of Knowledge, Semantic and Digital Media Technologies, pp. 323- 328, 2005.

[9] L. Jiebo, Boutell, M., and Brown, C., "Pictures are not taken in a vacuum - an overview of exploiting context for semantic scene content understanding," Signal Processing Magazine, IEEE, vol. 23, pp. 101-114, 2006.

[10] M. Cooper, Foote, J., and Girgensohn, A., "Automatically organizing digital photographs using time and content," Proc. of the International Conference on Image Processing pp. 749-752, 2003.

[11] A. Khella, and Bederson, B., "Pocket PhotoMesa: a Zoomable image browser for PDAs," Proc. of the

international conference on Mobile and ubiquitous multimedia pp. 19-24 2004.

[12] C. Gurrin, Jones, G. J., Lee, H., O'Hare, N., Smeaton, A. F., and Murphy, N. , "Mobile access to personal digital photograph archives," Proc. of MobileHCI '05, vol. 111, pp. 311-314, 2005.

[13] M. Davis, Smith, M., Stentiford, F., Bambidele, A., Canny, F., Good, N., King, S., Janakiraman, R., "Using Context and Similarity for Face and Location Identification," Proc. of the IS&T/SPIE Symposium on Electronic Imaging Science and Technology Internet Imaging VII, 2006.

[14] D. Raptis, Tselios, N., and Avouris, N. , "Context-based design of mobile applications for museums: A Survey of existing practices," Proceedings of the 7th international Conference on Human Computer interaction with Mobile Devices &Amp; Services, vol. 111, pp. 153-160, 2005.

[15] E. Nowak, Jurie, F., and Triggs, B., "Sampling strategies for bag-of-features image classification," European Conference on Computer Vision, 2006.

[16] S. Deerwester, Dumais, S., Furnas, G., Landauer, T., and Harshman, R., "Indexing by Latent Semantic Analysis," Journal of the Society for Information Science vol. 41, pp. 391-407, 1990.

[17] D. Comaniciu, and Meer, P., "Mean Shift: A Robust Approach Toward Feature Space Analysis," IEEE Transactions on Pattern Analysis ans Machine Intelligence, vol. 24, 2002.

[18] T. Poggio, and Smale, S., "The mathematics of learning: Dealing with data," presented at Notice of American Mathematical Society, 2003.

[19] J. C. Platt, "Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods," presented at Advances in Large Margin Classifiers, 1999.

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

[21] B. Manjunath, and Ma, W., "Texture features for browsing and retrieval of image data," IEEE Trans. Pattern Anal. Machine Intell., vol. 18, pp. 837-842, 1996.

[22] R. Jesus, Magalhães, J., Yavlinsky, A., and Rüger, S., "Imperial College at TRECVID," TREC Video Retrieval Evaluation (TRECVID), 2005.