Photo Retrieval from Personal Memories using Generic Concepts

Rui M. Jesus[1,2], Arnaldo J. Abrantes[1], Nuno Correia[2]

[1]Multimedia and Machine Learning Group, Instituto Superior de Engenharia de Lisboa
Rua Conselheiro Emídio Navarro, nº1, 1940-014 Lisboa, Portugal
[2] Interactive Multimedia Group, DI/FCT, New University of Lisbon
Quinta da Torre, 2825 Monte da Caparica, Portugal
+351 212948536
rjesus@deetc.isel.ipl.pt, nmc@di.fct.unl.pt

Abstract. This paper presents techniques for retrieving photos from personal memories collections using generic concepts that the users specify. It is part of a larger project for capturing, storing, and retrieving personal memories in different contexts of use. Semantic concepts are obtained by training binary classifiers using the Regularized Least Squares Classifier (RLSC) and can be combined to express more complex concepts. The results that were obtained so far are quite good and by adding more low level features, better results are possible. The paper describes the proposed approach, the classifier and features, and the results that were obtained.

Keywords: multimedia retrieval, personal memories, classification based on kernel.

1. Introduction

Humans like to keep information about their lives in order to later remember important moments or to create personal histories. One way of doing this is by collecting photos or videos. With the technological advances, multiple devices including phones, PDA's, digital stills cameras or digital video cameras are being increasingly used. The storage capacity of these devices has also increased a lot in the last few years, thus providing a very convenient way to store materials in digital form. The success of the WWW as a global platform for sharing information also promoted the media exchange process. Digital pictures are easier to capture and collect than the traditional pictures and for this reason many people have large collections of personal photos and videos.

The traditional way of organizing pictures in directories with suitable names for their content will not be enough to search and browse for personal pictures in an easy and fruitful way. One possible solution is the manual annotation with keywords of all the images, but this is not an easy task for large databases. Moreover, home users lack the expertise, the tools and, most of all, the time to perform this task. Automatic

annotation is generically done using low-level features [1] or context metadata obtained at the capture time [2]. The combination of these two sources of information [3] improves the performance of the retrieval but the systems based on low-level features (CBIR) still have some limitations [4]. The performance of these systems depends essentially of the low-level features, and for this reason, sometimes, these systems present a low performance because it is hard to capture semantic concepts (semantic gap [5]). These difficulties can be overcome by the previous training of semantic models and then, automatically associate textual descriptions to the images [6]. However, the image perception of the user must be the same of the annotator. Other solution is the use of relevance feedback [7, 8], including the user in the search process. The user interacts with the system by providing additional information in the retrieval task. The main difficulty of this solution is related with the initial results presented to the user. If they are not relevant and many different results are provided the relevance feedback will not work properly. Querying the system with just a few images samples will never be good for relevance feedback. However, relevance feedback is the best way to annotate images.

The tools and techniques that we are presenting here are part of a larger project to capture, store, and retrieve personal memories in different contexts and with different devices and user interfaces.

This paper presents an image retrieval system that uses previously trained generic concepts that are suitable to search in a personal picture collection. These generic concepts have the ability to provide relevant and distinctive images that could be used in the relevance feedback process. The models of the generic concepts are obtained by training binary classifiers using the Regularized Least Squares Classifier (RLSC) proposed in [9] and used in our previous work in relevance feedback [10, 11]. To combine several generic concepts the sigmoid function is applied to the output of RLSC.

The paper is structured as follows. Section 2 presents the related work, section 3 describes the features used and how the RLSC and the sigmoid function are used to rank the database. Section 4 presents the results obtained and the last section presents some conclusions and directions for future work.

2. Related Work

To manage personal memories with pictures several commercial applications (e.g., Adobe Photoshop Album, Paint Shop Pro, Picasa, Photofinder) and online sites like www.flickr.com or www.phlog.net are available. All of them use directories to organize pictures and some of them allow visualizing the directories chronologically. Most of these applications use the manual annotation for search photos. In fact, annotation is very important in order to explore personal collections. The manual annotation is the most effective way but it is time consuming. The MyAlbum [12] system use a semi-automatic strategy to annotate picture based on low-level features and in relevance feedback. In [13] camera phone users can annotate their photos instants after the capture, some annotations are done automatically (data and location) and then users can interact with the system to do some corrections. Automatic

systems rely on context metadata [14] or in visual content [2, 3, 15, 16]. Most of these systems use CBIR techniques that were used in other contexts different from the personal memories and combine them with context metadata. Other important aspect in multimedia retrieval is the user, in [17] several factors are discussed.
Concerning the problem of associating semantic concepts with low-level features, one of the first approaches proposed is described in [6]. They divided the images into rectangular regions and applied a co-occurrence model to words and regions. Following this work several proposals were made, some of them also associating words to images [16] and others, words to image regions [18, 19]. Naphade and Huang [19] proposed a probabilistic framework based on Bayesian Belief Networks. They segment the images in blob regions to create *multijetcs* (probabilistic multimedia objects) and to build a network of concepts. Recently, it was proposed a method in the domain of personal memories [16] that explores context metadata and visual content. The visual part of this work is similar to ours, but they use SVMs.

3. Query by Generic Concepts

This section describes the method proposed to query the database for pictures that belong to generic concepts. A set of generic concepts that were previously trained is available and the user can defines the query by combining them. These concepts are trained using the Regularized Least Squares Classifier for binary classification and the sigmoid function to convert the output of the classifier in a pseudo probability.

3.1 Low-Level Features

The low-level features that were used are the Marginal HSV color Moments [10] and Gabor Filter [20] to represent texture.
To extract the color feature, first each image is divided in 9 tiles (3x3) and for each tile individual histograms for the three color channels are computed. Then, the mean and the second central moment of each histogram are calculated. The color feature of each image is represented by a vector of 54 values.
The texture feature is extracted by applying to each picture a bank of 6 orientations and 4 scales sensitive filters that map each image pixel to a point in the frequency domain. The feature consists of the mean and standard deviation of the modulus of the filtering results. Each image is represented by a vector of 48 values.

3.2 Training the Generic Concepts

Useful concepts (e.g., people, indoor, outdoor, beach and snow) to search things in a personal collection are trained using the Regularized Least Squares Classifier. 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 $S_m = \left\{ \left(x_i, y_i \right)_{i=1}^{m} \right\}$ where labels $y_i \in \{-1,1\}$, the decision boundary between the two classes (e.g., indoor, outdoor) is obtained by the discriminant function,

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

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 [9],

$$(m\gamma I + K)c = y \tag{2}$$

where I is the identity matrix, K is a square positive definite matrix with the elements $K_{i,j} = K(x_i, x_j)$ and y is a vector with coordinates y_i. To choose the optimal σ for the Gaussian kernel the cross-validation method is used. Training the classifier is equivalent to solve a linear system with m equations.
The points $\{x_i\}$ with $f(x_i) \leq 0$, are classified in non relevant class $(y_i = -1)$, and the points with $f(x_i) > 0$ are classified in relevant class $(y_i = 1)$.

3.3 Ranking the Database

The output of the classifier is used to rank the database, however when several concepts are combined we need to convert the output in a pseudo probability p. Assuming w is a class (concept), given the output of the RLSC, the probability $p(w/x)$ is obtained by the sigmoid function in a similar manner of [21],

$$p(w/x) = \frac{1}{1 + e^{-f(x)}} \tag{3}$$

Given a query formed with k concepts $Q = \{w_1, w_2, ..., w_k\}$ and using f features the rank of each image is obtained by the probability,

$$p(w_1, w_2, ..., w_k / x) = \sum_{j=1}^{f} a_j \prod_{i=1}^{k} p(w_i / x) \tag{4}$$

where a_j is the weight of each features and $\sum_{j=1}^{f} a_j = 1$.

4. Experimental Results

The proposed method to query the database was tested 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. These pictures were manually annotated in order to evaluate the results obtained.
Personal memories are essentially composed by pictures of people, nature or urban scenes, holidays and parties. Five binary classifiers for 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.
Figure 1 shows a simple interface that was implemented to evaluate the method. The left panel shows toggle buttons labeled with the trained concepts. By selecting some of these buttons the user defines the query. Then, all the pictures are ranked and the top 30 images are presented in the central panel from top left corner to bottom right corner. For the query, “outdoor+beach”, only 3 pictures are incorrect.

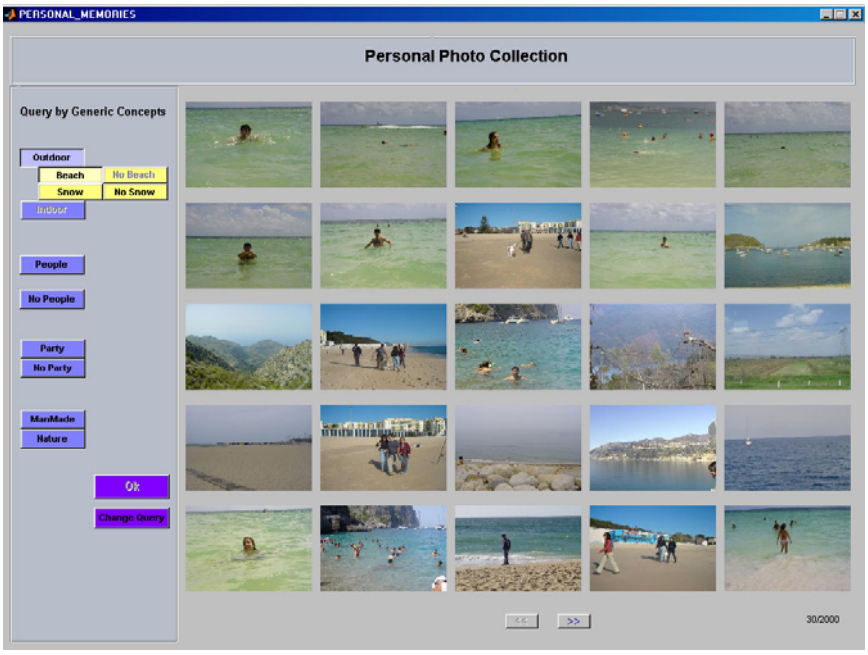

Fig. 1. Interface with the results obtained for the query Outdoor + Beach combining the 2 features.

To evaluate the system seven queries (see tables below) were performed in Rui Jesus collection and in the entire database. The features were evaluated individually, joined together in a vector obtained by concatenation of the color and texture feature, and by combining the ranks obtained by each feature individually. Mean average precision (MAP) was used to measure the performance of the system.

Tables 1 and 2 show the results obtained by the proposed method. As expected the results obtained by the Rui Jesus personal collection were slightly better than in the entire database, but only because the number of pictures. The combination of the rank of the two features presents the best results in both tables.

Table 1. Mean Average Precision obtained using the personal collection of one person (Rui Jesus).

Query	Color Moments	Gabor Filter	Color and Texture Feature	Combined Feature
Indoor	0,69	0,76	0,76	0,82
Outdoor	0,92	0,92	0,94	0,96
People	0,78	0,77	0,73	0,80
Party	0,05	0,06	0,03	0,05
Outdoor+ Beach	0,51	0,34	0,44	0,48
Outdoor + Snow	0,08	0,04	0,03	0,04
Indoor + People + Party	0,19	0,21	0,29	0,34
MAP	0,46	0,44	0,46	0,49

Table 2. Mean Average Precision obtained using all the images in the database (Rui Jesus and his friends).

Query	Color Moments	Gabor Filter	Color and Texture Feature	Combined Feature
Indoor	0,61	0,64	0,65	0,70
Outdoor	0,82	0,79	0,84	0,86
People	0,77	0,77	0,76	0,78
Party	0,11	0,13	0,05	0,12
Outdoor+ Beach	0,41	0,17	0,34	0,36
Outdoor + Snow	0,11	0,03	0,03	0,04
Indoor + People + Party	0,15	0,18	0,18	0,20
MAP	0,42	0,39	0,40	0,48

The concepts indoor, outdoor and people presents the best results and, the texture feature is better to retrieve indoor pictures which is explained by the struture of the manmade objects.
The query “Indoor + People+Party“ presents better results than the query “Party“ and this illustrates the main idea of the paper.

5. Conclusions and Future Work

This paper presents a method to retrieve images from personal memories based on generic concepts trained using the Regularized Least Squares Classifier. To combine

several concepts in a query the sigmoid function was applied to the output of the classifier. The method proposed was tested with 2582 pictures of a personal collection with a performance measured by a mean average precision of 0,48 when combining the rank of the two features. In the future, more generic concepts for querying personal memories will be trained and specific features for some concepts (party, face, snow) will be developed and evaluated.

We are currently building a complete environment to capture images and videos and additional contextual data, annotate, browse and search the collections. A previous project provided some of the scientific guidelines for this effort. We have implemented a mobile storytelling project (InStory) in a cultural and historical site [22] that uses PDAs to navigate narratives in a physical space. Users can navigate the story threads that are provided but they can also contribute with their own materials. The environment that we are building includes a PDA client that allows to capture the images along with GPS data and user annotations. We are also researching the way people use digital memories. Clients for browsing are also envisaged for different target platforms, including desktop/laptop PC, PDA, and augmented reality devices. The developed system will target to main types of users: (1) people engaged in tourism activities; (2) and also people that have difficulties to remember past events or are away from their natural environment, e.g., when in a hospital. In this case special care has to taken when designing the interfaces and we are planning to repurpose the stored materials, using techniques such as the ones reported in this paper and access them in other devices including mobile phones, TV sets and even paper or other custom made physical objects.

References

1. Veltkamp, R., and Tanase, M., *Content-Based Image Retrieval Systems: A Survey.* Technical Report UU-CS-2000-34, October, 2000.
2. Hori, T., and Aizawa, K., *Context-based video retrieval system for the life-log applications.* In Proceedings of the Fifth ACM SIGMM International Workshop on Multimedia Information Retrieval (Berkeley, CA, Nov. 7,). ACM Press, New York, 2003: p. 31-38.
3. O'Hare, N., 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, London, 2005.
4. Lew, M., 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, 2006. **2**(1).
5. Smeulders, A., 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, 2000. **22**(12): p. 1349 -1380.
6. 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.
7. Yong, R., Huang, T., and Mehrotra, S., *Relevance feedback techniques in interactive content-based image retrieval.* In Storage and Retrieval for Image and Video Databases (SPIE), 1998: p. 25-36.
8. Zhou, X., and Huang, T., *Relevance feedback in image retrieval: A comprehensive review.* Multimedia Systems, 2003. **8**(6): p. 536-544.

9. Poggio, T., and Smale, S., *The mathematics of learning: Dealing with data.* Notice of American Mathematical Society, 2003. **50**(5): p. 537-544.
10. Jesus, R., Magalhães, J., Yavlinsky, A., and Rüger, S., *Imperial College at TRECVID.* TREC Video Retrieval Evaluation (TRECVID), Gaithersburg, MD, Nov, 2005.
11. Jesus, R., Abrantes, A. , and Marques, J., *Relevance feedback in CBIR using the RLS classifier.* In 5th EURASIP Conference focused on Speech and Image Processing, Multimedia communications and Services, Bratislava, Junho, 2005.
12. Wenyin, L., Sun, Y., and Zhang H., *MiAlbum-A System for Home Photo Management Using the Semi-Automatic Image Annotation Approach.* ACM Multimedia, 2000.
13. Wilhelm, A., Takhteyev, Y., Sarvas, R., Van House, N., and Davis, Marc, *Photo Annotation on a Camera Phone.* Proc. ACM CHI 2004: p. 1403-1406.
14. *World-Wide Media eXchange.* http://wwmx.org, 2005.
15. Cooper, M., Foote, J., and Girgensohn, A., *Automatically organizing digital photographs using time and content.* In Proc. of the IEEE Intl. Conf. on Image Processing (ICIP 2003), 2003.
16. Jiebo, L., 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, 2006. **23**(2): p. 101-114.
17. Jaimes, A., *Human Factors in Automatic Image Retrieval System Design and Evaluation.* Proceedings of SPIE, Internet Imaging VII, San Jose, CA, USA, 2006. **6061**.
18. Cusano, C., Ciocca, G., and Schettini, R., *Image annotation using SVM.* Proceedings of the SPIE, Internet Imaging V 2003. **5304**: p. 330-338
19. Naphade, M.R., and Huang, T.S., *A probabilistic framework for semantic video indexing, filtering, and retrieval.* IEEE Transactions on Multimedia, 2001. **3**(1): p. 141-151.
20. Manjunath, B.S., and Ma, W. Y., *Texture features for browsing and retrieval of image data.* IEEE Trans. Pattern Anal. Machine Intell., 1996. **18**: p. 837-842.
21. Platt, J.C., *Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods,* in *Advances in Large Margin Classifiers,* P.B. A. Smola, B. Schölkopf, D. Schuurmans, Editor. 1999, MIT Press. p. 61-74.
22. Correia, N., Alves, L., Correia, H., Morgado, C., Soares, L., Cunha, J., Romão, T., Dias, A. E., and Jorge, J., *InStory: A System for Mobile Information Access, Storytelling and Gaming Activities in Physical Spaces.* ACE2005 - ACM SIGCHI International Conference on Advances in Computer Entertainment Technology, Universidade Politècnica de Valencia (UPV), Spain, 15 - 17 June, 2005.