

Personal Photo Album Summarization

Pinaki Sinha
Department of Computer
Science
University of California
Irvine, CA
psinha@ics.uci.edu

Hamed Pirsiavash
Department of Computer
Science
University of California
Irvine, CA
hpirsiav@ics.uci.edu

Ramesh Jain
Department of Computer
Science
University of California
Irvine, CA
jain@ics.uci.edu

ABSTRACT

Photo album summarization is the process of selecting a subset of photos from a larger collection which best preserves the information in the entire set and is semantically coherent. In this paper we propose a system which uses heterogeneous information sources associated with digital photos and generates a summary. Our algorithm adapts itself based on the type of event it is summarizing (Yearbook, Week or Single Day Event) We model the summarization problem as a retrieval problem based on different types of queries. We propose some evaluation metrics for the summary. We use an intuitive web based interface to present the results so that users can further explore the summary in an interactive way. This system is our submission to the CeWe Challenge for the Next Generation of Tangible Multimedia Products.

Categories and Subject Descriptors

H.3.3 [Information Storage Retrieval]: Information Search and Retrieval; I.4.9 [Image Processing and Computer Vision]: Applications

General Terms

Algorithms, Performance, Design

Keywords

Multimedia Grand Challenge, CeWe, Photo Album Summarization, Photo Story

1. INTRODUCTION

Increasing popularity of digital cameras, camera-phones and cheap memory cards has resulted in gigabytes of personal photo collection sitting on our storage devices or the web. These photos are testimonials to the events taking place in our lives. Most consumers use this personal photo collection for two major purposes. The first is to skim through the data to refresh our memories about the life events. The second is to share our photos among friends. The surging popularity of online photo-sharing system including those in social networks (Facebook, Twitter etc) testify to the importance of this use case. However, given the sheer size of the photo collections, it is time and labor

intensive to select photos for personal viewing or for sharing. (Semi)-Automatic summarization can help users narrow down to a more manageable subset of photos and help them explore their collection more efficiently. This process can also help photo printing companies to choose a subset of photos which the user might want to print. Previous research in this area [2], selected key photos based on time stamps and face content. Similar problems have been addressed in the video summarization community too [6].

2. THE SUMMARIZATION SYSTEM

Our summarization system shown in Figure 1 has two major stages: offline and online. Each of these stages has various modules which we discuss here. The feature computation module uses pixel data to find GIST features [3] (state of the art for scene recognition) as well as color and edge histograms. It also detects faces in the photos and filters them using human skin color segmentation (thus removing a lot of outliers). Optical context data (e.g., exposure time, focal length, flash etc) has been used to predict the situation in which the photo was shot [5]. The system uses this information for indoor-outdoor-night shot classification. It retrieves the location data for each photo either from the text-tags, gps-tags or from calendars of the users. Our summarization system considers three types of events: single day events (e.g., birthday celebrations), week events (e.g., trips) and year events (e.g., Year Book) (as required by the grand challenge) and adapts itself based on that. The temporal and spatial distribution of photos for a day event is very different from that of a yearbook. In the next module, based on the type of event and associated data present we do a macro partitioning. E.g., if the location information is present in the photos as discrete text labels, we partition the set based on location. If the temporal distribution is sparse (very typical in yearbook) we partition based on time-thresholds. Day events typically do not need such macro partition. In the next module we use exemplar selection algorithm [1] to find representative/ key photos in temporal and pixel domains. This algorithm is useful because we do not need to input the number of clusters, we can modify the importance of a candidate exemplar based on other knowledge sources (tags/comments by people) and we can vary the number of exemplars by modulating the prior belief of each photo being an exemplar. For each exemplar, we also compute the similarity with other photos in the same cluster. The next module is the ranking module. We rank the photos in each cluster based on human photographic behavior. Presence of faces is a useful cue for the importance

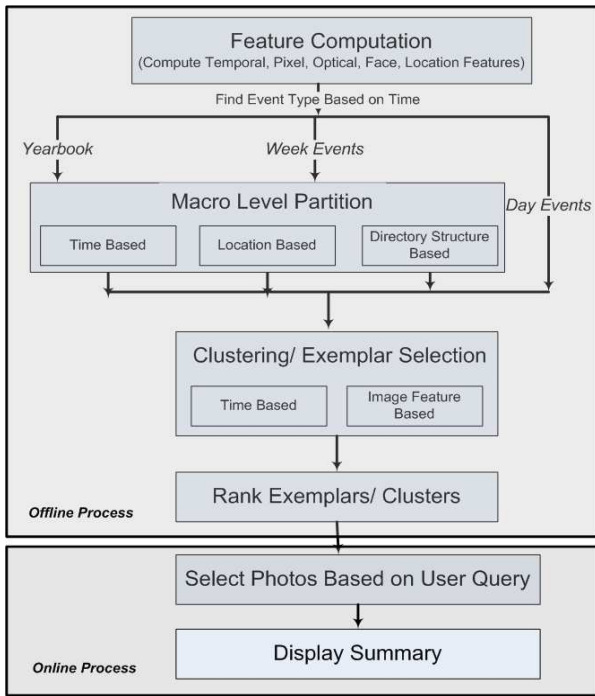


Figure 1: Photo Summary Architecture

of a photo. Other factors include uniqueness of the photo (not too many photos shot in the same time) the time and image feature domains. The online stage consists of the processes that happen in real time (when a user is generating a summary). We model the photo selection task a query-retrieval procedure based on some parameters chosen by the user. The set of queries could be When, What and Where. Each of these queries give more importance to time, image content and location respectively, and reranks the photos to generate the final summary. User can choose one or more of these queries in sequence. A typical summary could be “summarize my birthday event based on when events happened with more importance to people” (a When query augmented by presence of faces). Another query “summarize my yearbook based on places where I have been to and select unique photos “ (a Where query followed by What query). After the reranking, we show the photos in ranked order. The final process in the online stage is the summary display module. To make the process of summary viewing more exploratory, we build an interactive presentation interface by fusing timeline, maps, lists and other mashups.

3. EVALUATION OF THE SUMMARY

The evaluation of the summary can be both quantitative and qualitative. For quantitative evaluation, we compare the summary with ground truth. The ground truth set is either from user selected sets from online photo albums/social networks where people usually upload a subset/summary of photos. We also do qualitative evaluation based on the user feedback of our summary. The users were satisfied with a birthday-event summarization and had the following comments: the summary had the cake cutting event, the group photo of all participants and it outlined all the sub-events (dinner, games etc) that happened through that event. Fig-

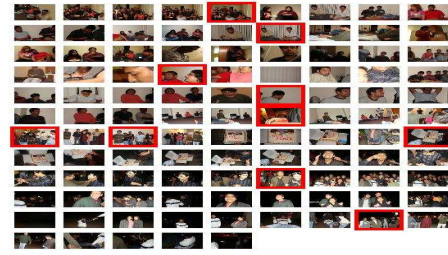


Figure 2: Entire Set of 104 Photos from a Birthday Event



Figure 3: Summary of 10 Photos based on Time and People

ure 2 shows photos from a birthday. The photos in red border are the ones chosen for summary. Figure 3 shows the summary in detail. Please zoom in to get a better view.

4. CONCLUSIONS

Our initial results [4] look promising. We successfully combined multiple information sources and the user was satisfied in our approach of modeling the summarization as a retrieval problem. However we plan to use more information like face clustering/ recognition and social network data to generate personalized photo-stories.

5. REFERENCES

- [1] B. Frey and D. Dueck. Clustering by passing messages between data points. *Science*, 315(5814):972, 2007.
- [2] J. Li, J. Lim, and Q. Tian. Automatic summarization for personal digital photos. In *ICSP, Pacific Rim Conf in Multimedia*, volume 3, 2003.
- [3] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *IJCV*, 42(3):145–175, 2001.
- [4] P. Sinha. <http://esl.ics.uci.edu/photosummary.html>.
- [5] P. Sinha and R. Jain. Classification and annotation of digital photos using optical context data. In *Proc of CIVR*, pages 309–318. ACM New York, NY, USA, 2008.
- [6] S. Uchihashi, J. Foote, A. Girgensohn, and J. Boreczky. Video Manga: generating semantically meaningful video summaries. In *Proc of ACM Multimedia (Part 1)*, pages 383–392. ACM New York, NY, USA, 1999.