Expo: An Expectation–oriented System for Selecting Important Photos from Personal Collections

Andrea Ceroni
L3S Research Center
ceroni@L3S.de

Vassilios Solachidis
ITI / CERTH
vsol@iti.gr

Claudia Niederée
L3S Research Center
niederee@L3S.de

Olga Papadopoulou
ITI / CERTH
olgapapa@iti.gr

Vasileios Mezaris
ITI / CERTH
bmezaris@iti.gr

ABSTRACT

The diffusion of digital photography lets people take hundreds of photos during personal events, such as trips and ceremonies. Many methods have been developed for summarizing such large personal photo collections. However, they usually emphasize the coverage of the original collection, without considering which photos users would select, i.e. their expectations. In this paper we present *Expo*, a system that aims at selecting which photos users perceive as most important and would have selected, thus meeting their expectations. It does not rely on any manually provided annotation, thus keeping the effort of users low. Photos are processed by applying a wide set of image processing techniques and a subset of a required size is selected. Users can review and modify the automatic selection. The system can also be used to gather training data by letting users select their preferred photos from the imported collections.

CCS CONCEPTS

• **Information systems** → **Multimedia information systems**;

KEYWORDS

Photo Selection; User Expectations; Image Importance; Coverage

1 INTRODUCTION

With digital photography, taking pictures is effortless and makes people easily ending up with hundreds of photos of personal events like holidays, trips, ceremonies. Different methods and applications have been developed to summarize such large collections and make their revisiting less onerous. They are usually centered around the concept of coverage, generating summaries that resemble the original collection. This is achieved either by clustering and picking representatives from each cluster [5, 8, 12] or by explicitly modeling and optimizing coverage [13, 14]. However, these strategies ignore user expectations, i.e. which photos users would select as most important for revisiting or preservation purposes. We argue that selecting photos that are important to users from personal collections is different than generating comprehensive summaries, as the set of important images might not fully cover the original collection [4]. For instance, users might not like photos of boring moments of a trip.

We present *Expo*, an expectation–oriented system for selecting photos from personal collections. Differently from previous systems (e.g. [2, 11]), it aims at emulating human selections by identifying those photos that users perceive as most important for revisiting or preservation, meeting their expectations. A demonstrative video is available[1]. To make selections, *Expo* relies on the method proposed by Ceroni et al. [4], which has been proved to be more effective in emulating user selections than methods based on coverage (e.g. [12–14]). It considers information at both image- and collection-level (by applying image quality assessment, concept detection, face detection, near-duplicate detection, event-based clustering) and estimates the selection probability of images via supervised Machine Learning.

Photos are first processed to extract the information required by the back-end selection method, which identifies the requested number of most important images. Users can browse the images, seeing also the information extracted from them and used in the selection process, as well as revise the automatic selection according to their preferences. The photos are organized in clusters (sub-events), to make browsing and reviewing easier for the users. *Expo* keeps the level of user investment low by not requiring any manual annotation. The system can also be used to acquire training data by letting users import new collections and manually select their preferred photos from them. Each imported collection and the associated selection data are sent to a common repository and used to re-train the model.

2 RELATED WORK

Different photo selection and summarization works are driven by the aspect of coverage, identifying clusters of images based on time and visual content and then picking a number of representative from each cluster [5, 8, 12]. Our selection method [3, 4] differs from them as it does not impose such a strict

[1] www.l3s.de/~ceroni/IPV/ICMR_2017_demo.mp4

ICMR '17, June 6–9, 2017, Bucharest, Romania
 ACM ISBN 978-1-4503-4701-3/17/06…$15.00
DOI: http://dx.doi.org/10.1145/3078971.3079011

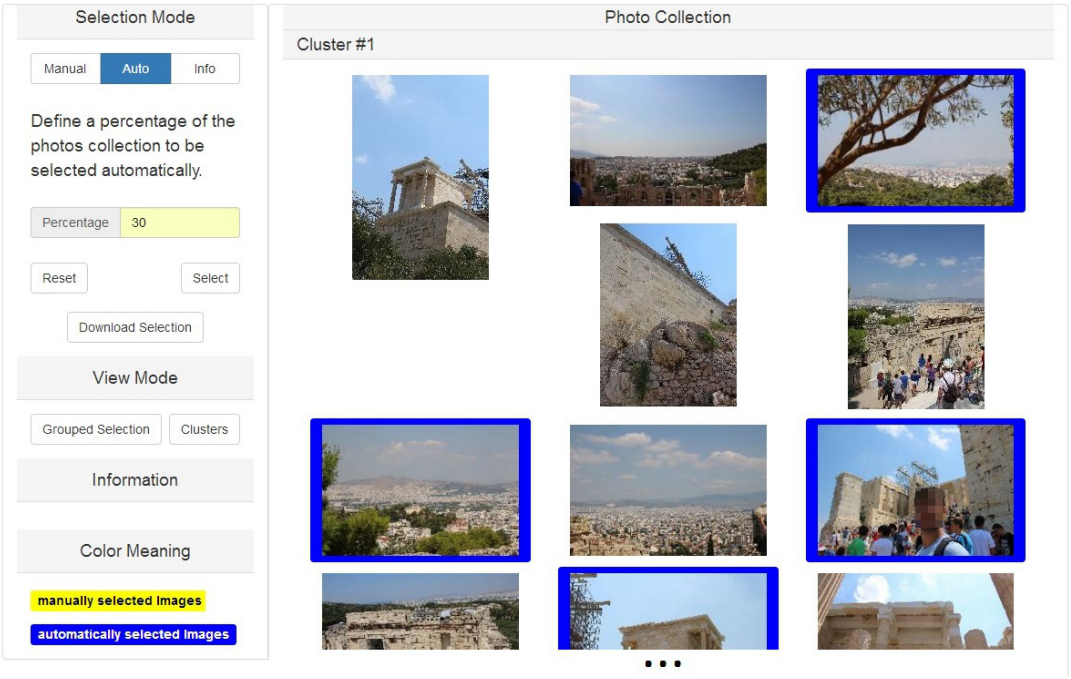


Figure 1: Performing automatic selections with *Expo*.

notion of coverage, but rather considers clusters and other global information together with image-level information, learning their different impact in a single model. The works in [13, 14] are closer to ours, as they consider coverage in a relaxed way as one dimension of a multi-goal optimization, but coverage still has the highest importance in the optimization and the summaries are not evaluated based on user assessments. Moreover, they make partial use of manually created text to associate semantic descriptors to images, while our system does not require any manual input, once the selection model has been learned.

Also available systems for photo selection and summarization [2, 11] are strongly based on coverage and diversity aspects and rely on a consistent amount of human supervision. These are the main differences from our system, which rather aims at identifying those photos that users perceive as most important for revisiting or preservation without any online input, after having learned the importance of coverage, diversity, and other aspects from human selections. Batko et al. [2] create textual and visual summaries of personal photo collections by combining clustering methods and image annotation tools. This requires a considerable effort to manually annotate pictures in the clusters. In [11], collections are first organized in clusters and images in the clusters are ranked based on visual quality, aesthetics, and the presence of faces. However, the weights of the different aspects are defined in presets and not learned from user selections, like our method does.

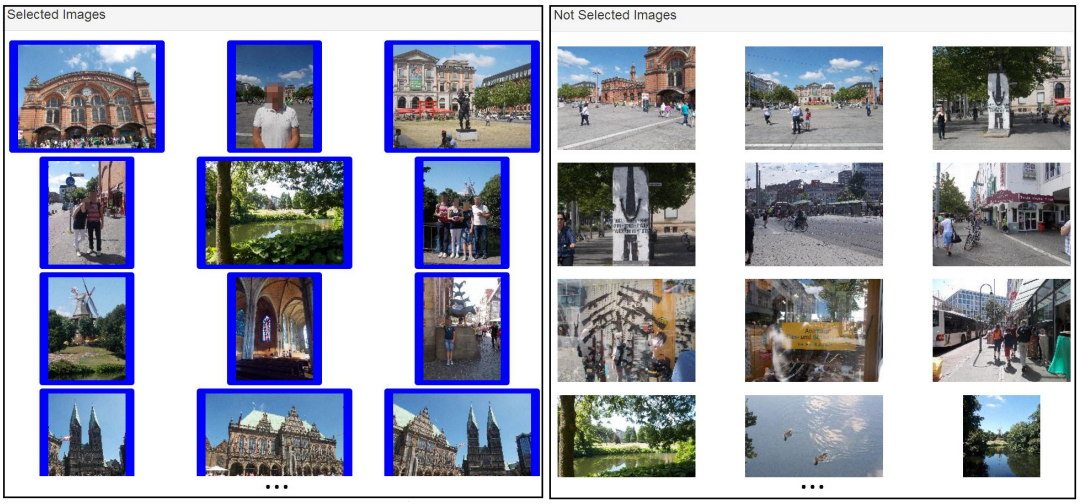
3 EXPECTATION-ORIENTED SELECTION

In this Section, we provide an overview of the main steps of the expectation-oriented selection method presented in [4], which is used in our system. Please refer to the original paper for further details.

Different image processing techniques are applied on input photo collections to extract features from them. These are: concept detection [9], to capture the semantic content of images beyond aesthetic and quality indicators; face detection [4], reflecting the importance of the presence of people in photos; near-duplicate detection [1], to take the redundancy of many pictures of the same scene as a signal of importance, and to eliminate very similar images; quality assessment [10], as good quality photos might be preferred in case of comparable photos; event clustering [6] and collection-level information for sake of browsing and to reflect the role of coverage in photo selection. This information is made visible to the user when browsing collections within the system.

Each photo is represented as a feature vector $\boldsymbol{f}_p$ by using the information listed above and a SVM trained on the data considered in [4] is used to predict the probability $I = SVM\left(\boldsymbol{f}_p\right)$ of each photo p to be selected, i.e. its importance. Finally, after predicting the importance of all the photos in the collection, the photos are ranked based on their importance and the top-n are selected. The parameter n, specified by the user, represents the desired size of the selection.

(a) Images grouped by clusters (sub-events).



(b) Images grouped by selection decisions.

Figure 2: Different ways of showing photos: grouping either by clusters (2a) or by selection decisions (2b).

4 SYSTEM DEMONSTRATION

In this Section we provide information regarding the implementation of the system (4.1) as well as the use cases that it supports (4.2).

4.1 Implementation and Functionalities

The system has been implemented as a Web application using the Java Play Framework[2], which simplifies the development of web applications with Java and Scala. The image processing techniques listed in Section 3 are exposed as RESTful Web Services, which are called once for each input collection. The classifier used for importance prediction has been built using the Support Vector Machine implementation of LibSVM[3] (with Gaussian Kernels) and has been trained on the dataset of [4].

The user interface is shown in Figure 1. Images are organized in clusters (sub-events) to simplify the browsing. The panel on the left-hand side contains all the functionalities available in the system. In the *Auto* mode, the one shown in the picture, users specify the number of images to select (as a percentage of the whole collection) and run the automatic selection. The selected images get a blue border and users

[2]https://www.playframework.com/

[3]http://www.csie.ntu.edu.tw/~cjlin/libsvm/

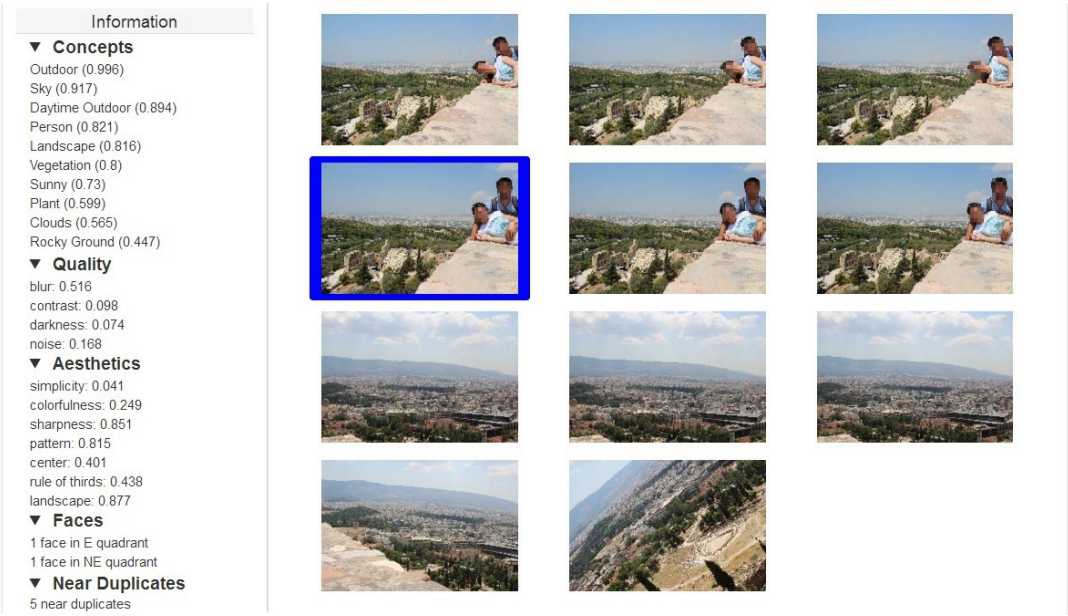


Figure 3: Looking at the features extracted from images.

can revise the selection according to their preferences. Besides organizing photos in clusters (Figure 2a), the "Grouped Selection" view (Figure 2b) shows all the selected images together to give an overall view of the selection.

Clicking on *Info*, the user can look at the information extracted from images. For instance, Figure 3 shows the information extracted from the image with blue border. Finally, clicking on *Manual*, users can produce their own selections manually and submit them (along with the whole collection) as training data for future re-training of the selection model.

4.2 Use Cases

The *Expo* system supports two use cases, one for the automatic identification of important images and one for providing new training data by performing manual selections.

Making Automatic Selections. The main functionality offered by the system to end users is automatically identifying those photos that are likely to be perceived as important by the owner of the collection, thanks to the underlying selection method aiming at meeting user selections and expectations. Starting from such automatic selection, the user can revise it according to his/her preferences. The availability of a fairly good initial selection alleviates the manual revision, which would become cumbersome when starting from scratch. At the end of the process, the selected photos can be downloaded and used in place of the whole collection, for instance, when looking back at the associated event in the future or sharing photos with relatives and friends. Such tasks would be more demanding and, to some extent, frustrating, if no selection was available and the original, potentially large collection had to be used.

Gathering Training Data. The system can be used to acquire new training data for the selection model. Users upload new photo collections, select their preferred photos from scratch, and submit the selection information and the whole collection to a central repository. Such collections annotated with user selections are used to re-train the model and to better match user expectations and selection patterns. This can be the starting point for user personalization, as previously investigated in [7].

5 CONCLUSION

In this paper we presented *Expo*, an expectation–oriented system for selecting pictures from personal collections. It offers a solution for the management of large personal photo collections by generating sets of highly important photos, which are more enjoyable for revising and sharing purposes. Differently from previous summarization systems based on the concept of coverage, it aims at emulating human selections by exploiting a selection method [4] that selects those pictures that users perceive as most important for revisiting or preservation, thus meeting their expectations. The notion of image importance has been learned from user selection patterns. *Expo* keeps the level of user investment low by not requiring any manual annotation. Users can review and modify the automatic selection according to their preferences. The system can also be used to gather training data by letting users select preferred photos from the imported collections.

Acknowledgments This work was funded by the German Research Foundation (DFG) in the context of the SPP 1921 on *Intentional Forgetting in Organisationen*, project *Managed Forgetting*.

REFERENCES

[1] K. Apostolidis, C. Papagiannopoulou, and Vasileios Mezaris. 2014. CERTH at MediaEval 2014 Synchronization of Multi-User Event Media Task. In *Proc. of MediaEval 2014 Workshop*.

[2] Michal Batko, Petra Budikova, Petr Elias, and Pavel Zezula. 2014. CLAN Photo Presenter: Multi-modal Summarization Tool for Image Collections. In *Proc. of ICMR '14*.

[3] Andrea Ceroni, Vassilios Solachidis, Mingxin Fu, Nattiya Kanhabua, Olga Papadopoulou, Claudia Niederée, and Vasileios Mezaris. 2015. Investigating human behaviors in selecting personal photos to preserve memories. In *International Conference on Multimedia & Expo Workshops (ICMEW '15)*.

[4] Andrea Ceroni, Vassilis Solachidis, Claudia Niederée, Olga Papadopoulou, Nattiya Kanhabua, and Vasileios Mezaris. 2015. To Keep or not to Keep: An Expectation–oriented Photo Selection Method for Personal Photo Collections. In *Proc. of ICMR '15*.

[5] Wei-Ta Chu and Chia-Hung Lin. 2008. Automatic Selection of Representative Photo and Smart Thumbnailing Using Near-duplicate Detection. In *Proc. of MM '08*.

[6] Matthew Cooper, Jonathan Foote, Andreas Girgensohn, and Lynn Wilcox. 2005. Temporal event clustering for digital photo collections. *In ACM TOMCCAP* (2005).

[7] Mingxin Fu, Andrea Ceroni, Vassilios Solachidis, Claudia Niederée, Olga Papadopoulou, Nattiya Kanhabua, and Vasileios Mezaris. 2015. Learning personalized expectation-oriented photo selection models for personal photo collections. In *International Conference on Multimedia & Expo Workshops (ICMEW '15)*.

[8] Jun Li, Joo Hwee Lim, and Qi Tian. 2003. Automatic summarization for personal digital photos. In *Proc. of ICICS-PCM '03*. DOI:https://doi.org/10.1109/ICICS.2003.1292724

[9] Foteini Markatopoulou, Nikiforos Pittaras, Olga Papadopoulou, Vasileios Mezaris, and Ioannis Patras. 2015. A Study on the Use of a Binary Local Descriptor and Color Extensions of Local Descriptors for Video Concept Detection. In *MultiMedia Modeling*.

[10] E. Mavridaki and Vasileios Mezaris. 2014. No-Reference blur assessment in natural images using Fourier transform and spatial pyramids. In *Proc. of ICIP '14*.

[11] Stefanie Nowak, Ronny Paduschek, and Uwe Kühhirt. 2011. Photo Summary: Automated Selection of Representative Photos from a Digital Collection. In *Proc. of ICMR '11*.

[12] Mohamad Rabbath, Philipp Sandhaus, and Susanne Boll. 2011. Automatic Creation of Photo Books from Stories in Social Media. *ACM TOMM* (2011).

[13] Boon-Siew Seah, Sourav S. Bhowmick, and Aixin Sun. 2014. PRISM: Concept-preserving Social Image Search Results Summarization. In *Proc. of SIGIR '14*.

[14] Pinaki Sinha, Sharad Mehrotra, and Ramesh Jain. 2011. Summarization of Personal Photologs Using Multidimensional Content and Context. In *Proc. of ICMR '11*.