Proc. Int. Workshop on Human Memory-Inspired Multimedia Organization and Preservation (HMMP) at the IEEE Int. Conf. on Multimedia and Expo (ICME 2015), June-July 2015, Torino, Italy

INVESTIGATING HUMAN BEHAVIORS IN SELECTING PERSONAL PHOTOS TO PRESERVE MEMORIES

Andrea Ceroni[1], *Vassilis Solachidis*[2], *Mingxin Fu*[1], *Nattiya Kanhabua*[1], *Olga Papadopoulou*[2], *Claudia Niederée*[1], *Vasileios Mezaris*[2]

[1]L3S Research Center, Leibniz Universität Hannover, Germany
{ceroni, kanhabua, niederee}@L3S.de
[2]Information Technologies Institute / CERTH, Greece
{vsol,olgapapa,bmezaris}@iti.gr

ABSTRACT

Photos are excellent means for keeping and refreshing memories. Digital photography, however, imposes new challenges for keeping photos accessible on the long run due to threats such as hard disk crashes, format changes, or storage medium decay. Safe long-term preservation, ensuring the longevity of photos, comes at a cost, suggesting a restriction of this investment to the most valuable photos. Therefore, understanding how people behave when selecting their most valuable photos from large collections is an important first step towards the development of automatic preservation approaches. In this paper, we conduct a user study of 35 participants in selecting personal photos for long-term preservation. The results of the study provide insights for the photo selection process. In addition, we propose a photo selection method based on machine learning and compared human selections with leading edge clustering techniques, highlighting significant issues in emulating human decision patterns.

***Index Terms*—** Photo Selection; Digital Preservation; Human Expectations; Long-term Memories

1. INTRODUCTION

Photos are taken for many different purposes [1]. Although immediate sharing has become very popular, photos are also taken for capturing memorable moments targeting the long-term perspective of reminiscence. The diffusion of high-resolution cameras allows people to take hundreds or thousands of images during relatively short events, and cheap storage indeed allows them to store all these images in some devices. Although stored, digital photos are subject to a new and less obvious type of fragility, which leads to a form of random "digital forgetting": over decades storage devices break down and formats and storage media become obsolete, making random parts of the photo collections inaccessible. One example is how difficult it would be today to access photos stored years ago in .mos format in a floppy disk. This can be alleviated by preservation techniques, which require either continuous effort of the content owners (e.g., for periodically moving their photos to modern storage devices and transcoding them to a modern format), or paying for preservation services (e.g. a specialized cloud service storing the photos in redundant drives and doing the necessary transcodings). Due to the growing volume of content, it is expected that preservation effort and cost (which goes far beyond mere storage cost) will not allow preserving every bit of content that is created, but just the most valuable content. Thus, the problem arises, how to support the user in selecting the subset of photos to preserve. In this paper, we present a user study to understand how people identify the most "valuable" photos from personal collections for subjecting them to special preservation activities. The insights gained provide a starting point for developing automatic preservation procedures.

Automated photo selection, in more general, has already been studied in various other contexts, such as photo summarization [2, 3], selection of representative photos [4, 5], and the creation of photo books from social media content [6]. Going beyond those approaches, photo selection for preservation also has to take into account social practices surrounding photos [1] such as why the photos are taken. For example, today people tend to take more photos of more mundane motives; and the cultural probe study presented in [7] shows that photos of everyday life are also considered as valuable mementos, e.g. because they are exemplary of the person's character or because of their social value. Such subjective aspects make automated photo selection for preservation a very challenging task. Our aim is to better understand the human selection process for long-term photo preservation, identifying preferences and behaviors that can drive automatic selection approaches dedicated to this task. For this purpose, we perform a photo selection study of 35 participants using real-world personal photo collections, each one containing some hundreds photos. In total, we obtain more than 8,000 photos that are manually labeled and exploited in our experiment.

Complementing this activity, we propose a machine learning method for automatically selecting photos and perform semantic event clustering on the collections, discussing the correlation between the identified clusters and the selections of the users. The use of event clustering for this purpose is motivated by (a) the successful clustering and coverage-based approaches in other photo selection tasks and (b) the importance given to event-coverage by the users of our study. This gives an indication of the issues that clustering–based approaches to photo selection and summarization might face in the scenario of long-term preservation.

The contributions of this paper are: (1) a user study and insights on the photo selection process for long-term preservation, (2) usage of personal and large data in both the study and experiments, (3) the development of an automatic approach to photo selection for preservation, and (4) a comparison between event-based clustering and human selections.

2. RELATED WORK

In the context of preservation of public images, a qualitative study assessing the value of images for representing social history is reported in [8]. The evaluators were asked to rate five images, selected from Flickr, considering their worthiness of long-term preservation. This study is mostly limited in (i) not considering personal photos and (ii) the small number of photos considered. Wolters et al. [9] investigated which photos from an event people tend to delete over time. The participants took pictures during a common event, and then they were asked for deletion decisions at different points in time. Although preservation ("keep") and "delete" decisions are related, we explicitly asked our evaluators to make selection decisions for preservation of images, rather than for deletion. Moreover, in our study the users were asked to make joint selection decisions (i.e. select a sub-collection) instead of atomic decisions. This is potentially a key difference, since selecting one photo might affect the decisions for other similar photos. Finally, instead of taking photos of a common event explicitly for the study, we work with personal real-world collections belonging to diversified events.

The task of automatic photo selection, more generally, has been tackled from different perspectives: identifying clusters of images based on time and visual content [4, 6], producing summaries that cover the content and concepts in the original collection [2, 3], and making selections based on quality criteria [10, 11]. The evaluation criteria and data considered in these works do not entirely match the requirements for long-term preservation: the images are either not judged by humans, or are rated according only to aesthetic criteria. The work in [2] generates summaries from personal photo collections by considering coverage and diversity within a multi-goal optimization problem. The method is evaluated without considering human judgments for the summary, thus there is no guarantee that the generated summaries match the ones expected by the user. In [3], images are clustered via an optimization method that jointly considers visual coherence between images, concept preservation, and coverage of the summary with respect to the original collection. The images are not personal and human judgments are not considered as an evaluation criterion. Yeh et al. [10] perform ranking of good photos based on aesthetic criteria, without considering user preferences and judgments as evaluation criteria.

3. USER STUDY

We performed a user study for a photo selection task, with the goal of gathering insights on behaviors exhibited by humans when selecting personal photos for long-term preservation. Participants were asked to provide their personal photo collections and to select a subset of photos that they would want to preserve, i.e. to ensure that the selected photos stay accessible for a long period of time. The user study was complemented by a survey about the task, which we asked the participants to fill-in after completing the photo selection task.

Participants. The experiment involved 35 users (28.6% females and 71.4% males) with 15 different nationalities: 25.7% of the participants came from Greece, 17.1% from Germany, 11.4% from Italy, 11.4% from China, 5.7% from Vietnam, and the rest from Ethiopia, Turkey, Kosovo, Iran, UK, Thailand, Sweden, Brazil, Albania, and Georgia. Regarding their ages, 60.0% of the participants are between 20 and 30 years, 25.7% between 30 and 40, 11.4% between 40 and 50, 2.9% between 50 and 60.

Task Definition. Since our task of selecting photos for preservation is not an everyday task for the users, it was important to find a good metaphor for supporting the task. After discussing a number of options with cognitive experts, we decided to use the metaphor of a "magic digital vault", which incorporates the ideas of protection, durability, and a sort of advanced technologies to keep things accessible in the long-term. Therefore, the task consisted in selecting a subset of valuable photos to be put in the magic digital vault, which would protect the images against loss and would ensure that they remain readable and accessible over the next decades.

Photo Collections. Previous works mostly consider either *public photo* collections (e.g., available on social media like Facebook and Flickr), e.g. [6], or photos from a *shared event* in which all the evaluators took part [12]. One difficulty we see with using public collections of photos from different people, even if they attended the same event, is that according to the different experiences of the individuals in the event they might also have a different level of appreciation for the same photo, thus influencing their decisions. In contrast, we use personal photo collections. For instance, these can be photos from business trips, vacations, ceremonies, or other personal events the evaluator participated in. This means that each collection is not just a bunch of photos, which might exhibit different degrees of quality and aesthetics, but there are experiences, sub-events, and memories that might influence the selection behavior. We decided to focus on such personal collections because we wanted to observe the personal photo

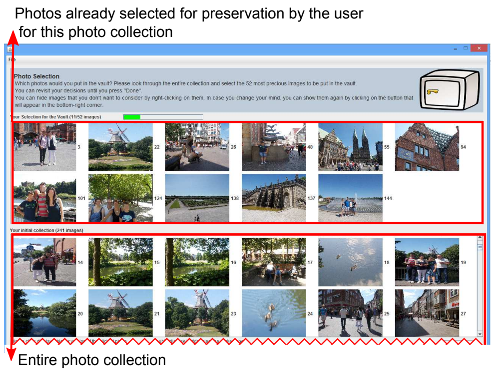

Fig. 1. GUI used by participants to select photos to preserve.

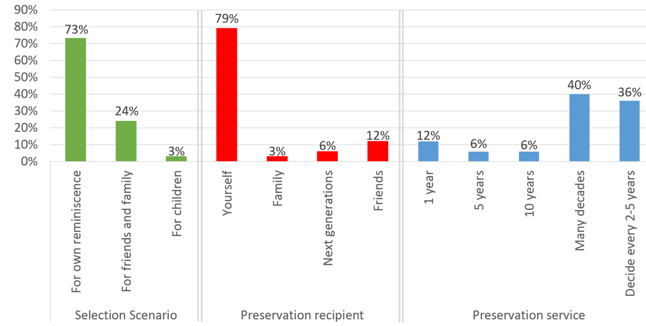

Fig. 2. Survey results with respect to preservation scenario, preservation target group, and preservation as a service.

selection decisions in a setting that is as realistic as possible. In total, 39 collections were used in the experiment (four users evaluated two collections), resulting in 8,528 images. The size of collections ranges between 100 and 625 images, with an average size of 219 and a standard deviation of 128.7. These collection sizes also emphasize the need for automated selection support, since manually browsing for photo selection becomes time-consuming. Overall, 51% of the collections represent vacations, 30% business trips, and 19% other events like music festivals and graduation ceremonies.

User Interface. We developed a desktop application, depicted in Figure 1, that was used by the evaluators to import collections and to select the photos they wanted to preserve. The images contained in the imported collection are displayed in the bottom panel, while the ones selected are shown in the top panel. The images in the collection were shown in the same order in which they were taken, since this makes the browsing, remembering, and selection easier and more realistic for the users. We verified that keeping the original order did not introduce any significant bias in the selection towards the early photos in the collection. This might have been a risk, since users might lose attention or even complete the selection without going through the entire collection.

User Evaluation Methodology. Before starting the evaluation, the users were personally introduced to the photo selection task as well as to the user interface (application) that they were asked to use. No guidelines were given about the criteria to use for selection, in order not to influence the selection process. The application asked them to select 20% of photos from the collection for preservation. This selection percentage (20%) has been empirically identified as a reasonable amount of representative photos, after a discussion with a subset of users before the study.

4. USER SURVEY AND DISCUSSION

After the photo selection step, the users were asked to fill a survey that can be conceptually split into two parts. The first group of questions refers to the scenario of photo selection process for personal preservation, while the second one looks into the criteria that were considered during the selection.

Regarding the first group of questions, the users were asked to provide information about (1) which scenario they had in mind when selecting the images; (2) for whom they are preserving the images; (3) whether they would be ready to pay, and for how many years, if preservation was a paid service. The answers to each question were posed as multiple choices and are reported in Figure 2. Questions (1) and (2) reveal that the process of long-term preservation is centered around the owner of the photos: more than 70% of the evaluators said that they thought about own future reminiscence when they selected the photos, and almost 80% indicated themselves as a main consumer of the preservation outcome. Looking at the preservation as a valuable service to be paid (question (3)), the evaluators were mostly split into two groups: either being ready to pay for many decades (39%) or needing flexibility to make new preservation decisions every 2-5 years (36%). In both cases, these answers highlight a clear need for preservation of personal photo collections.

In the second group of questions, we suggested different photo selection criteria and asked the users to rate how much each criterion was considered during the selection. The suggested criteria, which are in line with the insights on *keep* and *delete* decisions in [9], were rated via star ratings on a scale between 1 and 5 (5 stars mean very important, 1 means not important at all). The criteria along with statistics about their ratings are reported as box plots in Figure 3. Note, that medians are represented as horizontal bold bars, while sample mean is indicated with a bold cross. For sake of clarity, we grouped the criteria into three classes: (1) *content-based criteria* refer to objective and subjective measures for individual images such as image quality, image typicality (i.e. how suitable it is for serving as an iconic summary of the event), the presence of important people in images, whether images are generally important, and the evocation of memories, (2) *collection-based criteria* - here represented by coverage of events - consider an image in the context of its collection, and (3) *purpose-based criteria*, indicating the importance of different selection goals (in our case, sharing and preservation).

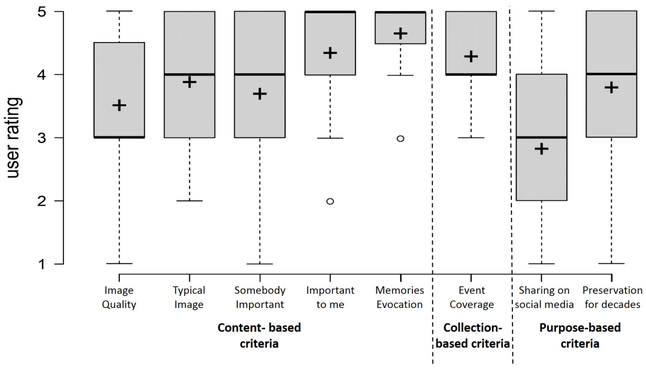

Fig. 3. Boxplots of the different selection criteria.

An important finding of this evaluation is that the objective quality of photos is rated as the second least important selection criterion, after the sharing intent. This shows that quality and aesthetics, although being important and used for "general purpose" photo selection [11], are not considered very important in case of selecting photos for preservation.

In contrast, criteria more related to reminiscence, such as event coverage, typical image, and "the picture evokes (positive) memories" are all rated high, with highest ratings for memory evocation. The remaining two criteria "picture is important to me" and picture "shows somebody important" refer to the personal relationship to the picture and are also both rated high. These results anticipate that the task of predicting images to be selected for long-term preservation is likely to be difficult, since many of the criteria that are rated high, e.g. memory evocation, personal importance and "typical image", are difficult to assess for a machine, because they contain a high level of subjectivity. Another complicating fact is that there is no single dominant selection criterion, but a combination of highly rated criteria. In the criteria ratings we can see clear differences to the ratings of the partially overlapping set of criteria reported in [12], where photos on shared events were used and the selection was not directly related to preservation and reminiscence. In that work, much higher ratings are given to criteria such as quality, whereas event coverage and importance of depicted persons are rated relatively low (although with high variance). Interestingly, photos that capture a memory are also rated high in this case.

5. AUTOMATIC PHOTO SELECTION METHOD FOR PRESERVATION

After having studied the impact of different criteria to photo selection for preservation, we experiment with automatic procedures to represent them and automate the selection process. In this work, we automatically assess the photos by applying three pre-processing steps: quality assessment, face detection, concept detection. By leveraging this information, we train different models to predict the importance of images and to make automatic selections. Figure 4 gives an overview of our proposed system framework. The image processing module extracts a set of features in order to (1) learn models to predict the importance of a photo, i.e., the probability of being selected, or (2) exploit in a photo clustering process. The importance prediction is described in the rest of this section, while a comparison between the event-based clustering and selections done by humans will be discussed in Section 6.

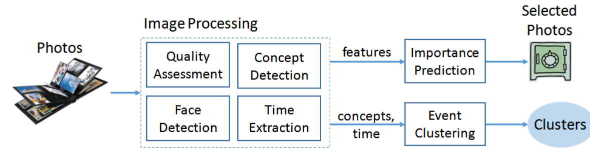

Fig. 4. High-level overview of our approach to automatically selecting photos for preservation.

5.1. Image Processing

Image Quality Assessment. We employ four image quality measures, namely blur, contrast, darkness, and noise, along with their aggregated value (Minkowski sum), following the procedure presented in [13].

Face Detection. One of the most successfully face detectors is the Haar-like features based introduced in [14], which we have applied using several detection classifiers. Each detected region is accepted as facial one by taking into account the number of classifiers that detected it, its color histogram (skin like color) and whether other facial features (eyes, mouth, nose) have been detected in it.

Concept Detection. This step analyzes the visual content of an image and automatically assigns concept labels to it. Recent approaches extract local features from images (e.g. SIFT and SURF), build a global image representation from the local features (e.g. with BoW, VLAD, Fisher vectors), and train concept detectors that rely on machine learning techniques [15]. We extended the 2–layer stacking architecture proposed in [16] by using first–layer classifiers that exploit VLAD encoding [15] for the local features (SIFT). Using this method, 346 concept detectors were trained for the 346 concepts defined as part of the TRECVID 2013 benchmarking activity [17]. As training corpus, the TRECVID 2013 dataset comprising 800 hours of video was used.

5.2. Importance Prediction

Three groups of features resulting from the above analysis techniques have been selected to describe each image. The *quality features* consist of the aforementioned quality measures: blur, contrast, darkness, noise, and their fused value. For *face features*, we divide each image in nine quadrants, and computed the number of faces and their size in each quadrant. This results in 19 features: the number and size of faces in

Proc. Int. Workshop on Human Memory-Inspired Multimedia Organization and Preservation (HMMP) at the IEEE Int. Conf. on Multimedia and Expo (ICME 2015), June-July 2015, Torino, Italy

	Random	Quality	Faces	Concepts	All
P@top-20%	0.2	0.3311	0.3476	0.3862	0.3953

Table 1. Precision at top-20% of different features.

each quadrant, plus an aggregated one representing the total number of faces in the image. The *concept features* consist in a vector of 346 elements, where the i-th value represents the probability for the i-concept to appear in the image.

Once images in our collections have been described in terms of these features, a prediction model represented by a Support Vector Machine (SVM) is learned to predict the selection probabilities of new unseen images. Given a training set made of photos p_i, their feature vectors $\boldsymbol{f}_p$, and their selection labels l_p (i.e. *selected* or *not selected*), an SVM is trained and the learned model M can be used to predict the probability $P = M\left(\boldsymbol{f}_q\right)$ for a new unseen image q to be selected by the user. We use the selection information obtained in the user study as data set. To avoid overfitting, the model was trained via 10-fold cross validation over the collections and the generated output probabilities were considered in our evaluation. Once the importance of each image is predicted, images in the same collection are ranked based on this value and the top-20% is finally selected.

We use SVM with Gaussian Kernel (LibSVM[1] implementation) with parameters $C = 1.0, \gamma = 1.0$.

5.3. Results

Since the users were asked to select the 20% of their collection, we let the models select the 20% of each collection as well. We then measured the precision of the selections as the fraction of the number of suggested photos that were selected by the users with respect to the total number of photos in the selection. The results presented in the rest of this section are averaged over the 39 collections considered in our work.

Different prediction models have been trained by using the subsets of the features described before, and the results are reported in Table 1. *Quality* and *faces* features are the ones that perform worst. For the quality features, this is expected from the results of the survey and has also already been observed for other photo selection tasks [12]. In addition, *faces* features alone also do not seem a very good indicator, although the presence of importance people was rated as highly important in the survey. As a matter of fact, the mere presence of people cannot give any indication about their importance for the user. The performance achieved when only using *concepts* features is clearly better than the ones of *quality* and *faces*: they are able to capture the semantic content of the photos. Examples of concepts with high importance in the model are *person*, *joy*, *cheering*, *entertainment*, and *crowd*. The model *all*, trained with all the available features, slightly improves the performance of using concept features alone.

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

6. IMAGE CLUSTERING AND PRESERVATION

Finally, we analyze the applicability of current selection and summarization approaches to the scenario of preservation, highlighting possible issues that they might face in this scenario. The main uncertainties in applying state of the art methods to our task are (a) that they are developed with other photo selection scenarios in mind and (b) that they often do not compare the performances of their output with selections done by users. They, for example, identify sub-sets of photos that provide comprehensive summaries of the initial collections [2, 6], without checking if the summary meets the user expectations, or they consider judgments based on more objective criteria such as aesthetics [10, 11].

Since a wide part of the state of the art methods for photo selection and summarization [2, 3, 4, 6] considers clustering and/or coverage for generating selections and summaries, we clustered photos and compared clustering results with human selections. This analysis is corroborated by the fact that the *event coverage* criterion, representable through clustering, has been identified as important during our study (Section 4).

Clustering. Several general-purpose clustering methods (e.g. k–means or hierarchical clustering) can be used in combination with visual features or vectors of responses of concept detectors, as in [18]. Based on the outcomes of [18], where image clustering has been tested using a variety of clustering methods and several image representations, we adopted in our experiments the K–means - model vectors (i.e. vectors of concept detection responses) combination that achieved the best performance. For capturing sub-events, we merge the results of the semantic clustering based on concept features with the results of a temporal clustering of the images, as in [19].

Discussion. In our opinion, one of the main risks of applying clustering to emulate human selections for long-term preservation is that not all the clusters might be important for the users. There might be photos from a sub-event that the user either simply does not like or considers less important than others. We corroborated this hypothesis by counting the number of human-selected images in each cluster formed in our collections. As to be expected, only for a few clusters (7.3%) all images of the cluster were selected. However, for a considerable part of the clusters (43%) no images were selected at all. Given these statistics, the selection done by any pure *coverage–based* method that picks an equal number of images from each cluster will contain at least 43% of images that would not have been selected by the user.

Finally, making the assumption that bigger clusters might be more important for the users (as indicated by the users' choice to take more photos that capture that part of the event), we consider the size of the clusters with respect to the number of user-selected images that they contain. Figure 5 shows the correlation between relative size of clusters (x axis) and the percentage of selected images in them (y axis). It is possible to observe that the selections done by the users result in many clusters with few selected images in each, which is coherent

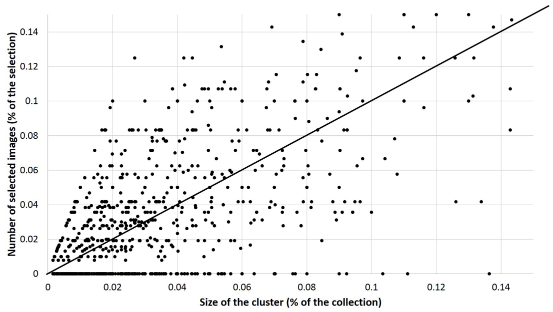

Fig. 5. Amount of selected images in clusters (with respect to the size of selection) versus relative size of clusters.

with the notion of coverage. However, what is more interesting is that the size of the cluster seems to be only marginally correlated with the importance of the cluster (i.e. the number of selected images it contains). This is potentially another limitation for all those methods that select an amount of images from each cluster proportionally to its size.

7. CONCLUSION

In this paper, we presented a user study to better understand the human selection process in the scenario of long-term photo preservation. The results of the study gave insights that can be the starting point for the development of automatic selection approaches dedicated to the scenario of preservation. In addition, we experimented the modeling of the insights within automatic selection methods, and analyzed the correlation between selections done by humans and automatically identified photo clusters. The analysis highlighted issues that clustering–based approaches might face when applied to the scenario of preservation. We plan to exploit these results to advance automatic approaches to photo selection for preservation by investigating how clustering and coverage-based approaches can be improved to meet user selections.

Acknowledgments This work was partially funded by the European Commission in the context of the FP7 ICT project ForgetIT (under grant no: 600826).

8. REFERENCES

[1] S. E. Lindley, A. Durrant, D. Kirk, and A. S. Taylor, "Collocated social practices surrounding photos," *International Journal of Human-Computer Studies*, 2009.

[2] P. Sinha, S. Mehrotra, and R. Jain, "Summarization of personal photologs using multidimensional content and context," in *Proc. of ICMR '11*, 2011.

[3] B.-S. Seah, S. S. Bhowmick, and A. Sun, "Prism: Concept-preserving social image search results summarization," in *Proc. of SIGIR '14*, 2014.

[4] W.-T. Chu and C.-H. Lin, "Automatic selection of representative photo and smart thumbnailing using near-duplicate detection," in *Proc. of MM '08*, 2008.

[5] E. Guldogan, J. Kangas, and M. Gabbouj, "Personalized representative image selection for shared photo albums," in *Proc. of ICCAT*, 2013.

[6] M. Rabbath, P. Sandhaus, and S. Boll, "Automatic creation of photo books from stories in social media," *ACM Trans. Multimedia Comput. Commun. Appl.*, 2011.

[7] I. Mols, E. v. d. Hoven, and B. Eggen, "Making memories: A cultural probe study into the remembering of everyday life," in *Proc. of NordiCHI '14*, 2014.

[8] A. Copeland, "The use of personal value estimations to select images for preservation in public library digital community collections," *Future Internet*, vol. 6, 2014.

[9] M. K. Wolters, E. Niven, and R. H. Logie, "The art of deleting snapshots," in *Proc. of CHI EA'14*, 2014.

[10] C.-H. Yeh, Y.-C. Ho, B. A. Barsky, and M. Ouhyoung, "Personalized photograph ranking and selection system," in *Proc. of MM '10*, 2010.

[11] C. Li, A. C. Loui, and T. Chen, "Towards aesthetics: A photo quality assessment and photo selection system," in *Proc. of MM '10*, 2010.

[12] T. Walber, A. Scherp, and S. Staab, "Smart photo selection: Interpret gaze as personal interest," in *Proc. of CHI '14*.

[13] E. Mavridaki and V. Mezaris, "No-reference blur assessment in natural images using fourier transform and spatial pyramids," in *Proc. of ICIP'14*, 2014.

[14] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision*, 2004.

[15] R. Arandjelovic and A. Zisserman, "All about vlad," in *Proc. of CVPR '13*, 2013.

[16] F. Markatopoulou, V. Mezaris, and I. Kompatsiaris, "A comparative study on the use of multi-label classification techniques for concept-based video indexing and annotation," in *Proc. of MultiMedia Modeling*, 2014.

[17] P. Over, G. Awad, M. Michel, J. Fiscus, G. Sanders, W. Kraaij, A. F. Smeaton, and G. Quenot, "Trecvid 2013 – an overview of the goals, tasks, data, evaluation mechanisms and metrics," in *Proc of TRECVID 2013*, 2013.

[18] C. Papagiannopoulou and V. Mezaris, "Concept-based image clustering and summarization of event-related image collections," in *Proc. of HuEvent'14*, 2014.

[19] K. Apostolidis, C. Papagiannopoulou, and V. Mezaris, "CERTH at MediaEval 2014 synchronization of multi-user event media task," in *Proc. of MediaEval*, 2014.