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

# LEARNING PERSONALIZED EXPECTATION–ORIENTED PHOTO SELECTION MODELS FOR PERSONAL PHOTO COLLECTIONS

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

Human expectations and practices are key aspects to consider when developing semi-automatic methods to select important photos from personal collections, e.g. for creating an enjoyable sub-collection for revisiting or preservation. The photo selection process (especially for personal data) can be highly subjective and the factors that drive the selection can vary from individual to individual. Thus, generic selection models might have limitations in meeting the different expectations and preferences of each user. In this paper, we propose a personalized photo selection model to assist users in photo selection, which adapts to their selection behaviors and preferences. Given an initial selection model trained on the available data, selection decisions done for new collections are acquired and the selection model is re-trained accordingly. Our experiments, based on real-world personal photo collections with overall more than 18,000 images, show promising adaptation capabilities of our personalized selection models.

***Index Terms*—** Photo Selection; Human Expectations; Personalization

## 1. INTRODUCTION

Nowadays, with the rich availability of cameras and smart phones, people continuously take photos from vacations and every day life. While the decreased storage price allows to easily store photos in cheap storage devices, the huge size of the collections (hundreds or thousands of photos) makes going through them as well as manually categorizing and sorting tedious tasks. This leads to the risk of having personal photos and memories "forgotten" by users, who rarely access (and enjoy) them again. Moreover, as time goes on, the stored photos have to face the risk of inaccessibility and unreadability in the future, either because of the obsolescence of software and image format, or because of hard disk crashes. These threats suggest to select, supported by automated methods, the most valuable photo subsets from personal collections, and keeping them accessible and enjoyable in the long run.

Human expectations and practices are key aspects to consider when developing semi-automatic photo selection methods for personal collections, especially for preservation purposes [1]. An expectation-oriented photo selection method has been proposed by Ceroni et al. [2] to identify photos from personal collections that users perceive as important and would have selected. Although it has been proved to be more effective in meeting user expectations than the common process of first clustering the photo collection and subsequently picking the most representative photos from the clusters [3, 4, 5], this method generates one single selection model to be used for any user and input collection. As a matter of facts, the photo selection process (especially for personal data) can be highly subjective and the factors that drive the selection can vary from individual to individual [6]. Some users might be particularly interested on photos depicting many people, while others might prefer pictures with landscapes or buildings. It is then important to provide personalized selection models that adapt to the preferences of the user.

In this paper, we propose a personalized photo selection model to assist users in photo selection, which adapts to the photo selection behaviors and preferences of the user. Starting from the general model presented in [2], selection decisions done by a given user on new collections are acquired and the selection model is updated according to them. Feeding the revisions of the user for automatically generated selection back into the selection model can, on the long run, bridge the gap between the general selection model and the user preferences. Moreover, in order to tackle the problem of having limited initial data to train the model (cold-start scenario), we experiment whether the exploitation of data from other users can boost the adaptation of the model to a given user when a limited amount of personal training data is available. This is based on the assumption that, despite the subjectivity of the task, common selection patterns exist and could be captured through a sample of selections done by other users.

In this paper we make the following contributions: (i) we present a personalized expectation-oriented photo selection method, which adapts to user preferences by updating the model based on new selection feedback; (ii) we investigate the benefit of exploiting data from other users to train the initial selection model when limited personal training data is available (cold-start scenarios); (iii) our evaluation with real-world personal collections shows that incorporating user feedback can benefit the selections on new unseen collections of the user, and exploiting annotated collections from other users can boost the system performances in cold-start scenarios.

## 2. RELATED WORK

The problem of photo selection has already been studied in different contexts, such as, photo summarization [4, 7, 8], selection of representative photos [3, 9], the creation of photo books from social media content [5], and the identification of appealing photos based on aesthetic criteria [10, 11]. We consider the task of selecting photos from personal collections that are the most important to the owners and meet their expectations (e.g. for revisiting or preservation). This task has been already tackled in [2], where the proposed method aims at predicting the photos that users would have selected (i.e. their expectations) by considering information at both image- and collection-level and learning their different impact through a single model. In this paper, we use such a selection model as a starting point for personalization.

Huang et al. [12] proposed an image retrieval system for personalized portraits ranking based on four kinds of features and two user interfaces to acquire personalized feature weights from the user for ranking. A similar approach has been presented in [11], where the ranking is not restricted to portraits. In both works, the user preferences are explicitly expressed through the interfaces instead of implicitly learning them from the data, which is indeed what we aim at. The work in [13] is closer to ours, as it learns to rank photos from a dataset of public rankings and realizes personalization by exploiting examples of personal rankings for re-ranking (i.e. personal and public rankings are considered together when re-ranking). Besides the different learning algorithm employed, this work differs from ours because the ranking is done only based on aesthetic features. Relevance feedback has been used to iteratively refine search results based on human feedback [14, 15]. These approaches are not directly applicable to our scenario, where there is not any explicit query to be considered as reference when refining the result set. Finally, collaborative filtering techniques have been used (e.g. in [16]) to accumulate records of user feedback and exploiting such relations among images to help future users.

## 3. EXPECTATION-ORIENTED SELECTION

In this Section, we provide an overview of the expectation-oriented selection method presented in [2], since we exploit it as starting point for personalization.

A wide set of features, both at image- and collection-level, is extracted from photos. These consist in (a) advanced concept detection (to capture the semantic content of images beyond aesthetic and quality indicators), (b) face detection (reflecting the importance of the presence of people in photos), (c) near-duplicate detection (to take the redundancy of many pictures of the same scene as a signal of importance, and to eliminate very similar images), (d) quality assessment (good quality photos might be preferred in case of comparable photos), and (e) event clustering and collection-level information to reflect the role of coverage in photo selection.

These features are combined via machine learning, providing a model that predicts the probability of a photo to be selected, i.e. its importance. In more detail, a prediction model represented by a Support Vector Machine (SVM) [17] is learned to predict the selection probabilities of new unseen photos. Given a training set made of photos $p_i$, their corresponding feature vectors $\boldsymbol{f}_{p_i}$, and their selection labels $l_{p_i}$ (i.e. *selected* or *not selected*), an SVM is trained and the learned model $M$ is used to predict the importance $I = M\left(\boldsymbol{f}_{p_{new}}\right)$ of a new unseen photo $p_{new}$, i.e. its probability to be selected by the user.

Once the importance of each photo is predicted, photos in the same collection are ranked based on this value and the top-$n$ is finally selected. The parameter $n$ represents the requested size of the selection (specified by the user).

## 4. PERSONALIZED SELECTION MODELS

Previous works on photo selection [6, 18, 11] have revealed that the photo selection task is, to some extent, subject to the preferences of each user. General selection models, although capable of representing common selection patterns (e.g., photos depicting people might be usually appreciated), might be improved by considering the preferences of each single user separately and derive personalized models for them. In this section, we show how personalized models have been derived from the photo selection approach described in Section 3, denoted *general model* hereafter.

### 4.1. Overview

We adopt an incremental learning strategy to achieve personalization, re-training the model each time new data (i.e. selection decisions) is provided by the user. The personalization workflow is summarized in Figure 1, which emulates the application of the personalized model in a real-world settings. As described in [2], the annotated photo collections available to train the general model are first pre-processed through image processing techniques and features are extracted from them. For each new collection provided by the user, a first selection is made by the trained general model and the selected photos are displayed to the user, who gives feedback revising the automatically generated selection. The training dataset is

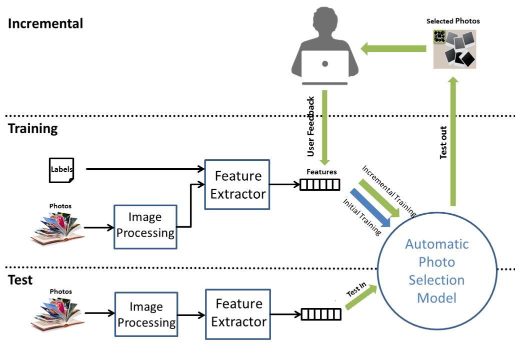


**Fig. 1**. Overview of the personalized photo selection model.

**Algorithm 1:** Model Update

**Input** : test collections $C = \{C_i\}$, annotated collections $C^*$
**Output**: collections with selection labels $C^{'} = \{C_i^{'}\}$
$T = C^*$
**for** $i := 1$ **to** $|C|$ **do**
  $M = \text{Training}(T)$
  **for** each $p \in C_i$ **do**
    $i_p = M(p)$
    $I_{list} = I_{list} \cup i_p$
  **end**
  $C_i^{'} = \text{RankingAndSelection}(I_{list}, C_i)$
  $C_i^* = \text{UserFeedback}(C_i^{'})$
  $T = T \cup \{C_i^*\}$
**end**

then expanded by adding the feedback data and the general model is retrained with the updated training dataset. Iterating this process, it is expected that the gap between user expectations and model's selections gets lower, due to the adaptation of the model towards the selection preferences of the user.

This workflow represents the envisioned behavior once the whole system has been finalized and released to the end user. However, in order to easily repeat evaluations when designing and implementing the model, we collected the data from each user once for all, i.e. users evaluated all the collections from scratch without revising any automatically generated selection. Although we are aware that the selections done by the user starting from an automatically generated selection might differ from those done when selecting photos from scratch, repeating the evaluation multiple times when designing the system would have been unfeasible for the users. Moreover, acquiring evaluations done from scratch is unbiased towards the initial selection proposed automatically.

### 4.2. Incremental Learning

A recurrent problem in machine learning is continuously managing new data, so that the existing model can be updated to accommodate new information and to adapt to it. Two common approaches for updating the model to new incoming data are *online learning* [19], where the model is updated only considering the new data, and *incremental learning* [20], where the model update considers the old training data along with the incoming data. We consider the latter strategy because, in our scenario, the updated model has to be aware of the entire data available, not just of the most recent one.

Although efficient and effective incremental versions of off-line learning algorithms exist (e.g., [20]), we perform the model update by including the new data in the training set and re-train the model from scratch. We implemented such more straightforward but functionally equivalent approach because our scenario does not impose strict time constraints for the model update, thus making the efficiency benefit of incremental versions of secondary importance. The time taken by a user to produce a new collection (e.g. after a trip or vacation) can be considered sufficient to re-train the model with the whole available data. Should the temporal constraints of the envisioned scenario become stricter, the incremental version of the employed algorithm could be plugged in without changing the functionalities of the whole application.

### 4.3. Model Update

Our personalized photo selection models, one for each given user, are built by re-training the model every time that a new collection is imported and the automatic selection done by the current selection model is revised by the user.

The pseudo-code of the model update is shown in Algorithm 1. The input includes a set of new unseen collections $C = \{C_1, \ldots, C_n\}$ from the user as well as a set of collections $C^*$ with selection labels available, which represents the initially available training data. The output is the set of the test collections with prediction labels (selected or not selected) which is denoted as $C^{'} = \{C_1^{'}, \ldots, C_n^{'}\}$. At the beginning, the training dataset $T$ is composed by the initial data $C^*$ and an initial prediction model $M$ is trained from it applying the method described in [2]. For each photo $p$ in the user collection $C_i$, the selection probability (i.e. importance) $i_p$ is predicted by the general model $M$ and added in the importance list which records the importance of photo in the entire collection. Following, according to [2], the photos are ranked based on their importance value and top-$n$ of them are selected which results in the selections $C_i^{'}$. In order to know which photos the user would really have selected or not selected, we ask the user to give feedback by revising the generated selections. This is finally included within the available training dataset. The prediction model $M$ will be retrained by using such new training data and applied to make predictions for the next coming collection $C_{i+1}$ of the user.

### 4.4. Cold–Start Problem

Usually, the adaptation of a system within the initial rounds of user interactions is affected by the so called *cold-start prob-*

*lem*: there is not enough (or even not at all) training data to let the model adapt to the user. This holds in our scenario as well, where the selection model might not perform proper predictions because of the lack of annotated collections in the initial training set $T$. We consider two ways of building the initial training set. One consists in using one annotated collection of the given user as initial training set. The other is based on using annotated collections from other users to train the initial selection model, based on the assumption that some common selection patterns could be captured through a sample of selections done by other users. We will experiment and compare these two strategies in our experiments.

## 5. EXPERIMENTS AND RESULTS

### 5.1. Experimental Setup

**Dataset.** We use personal photo collections with user annotations as dataset, which gives us a ground truth for assessing the adaptation capabilities of our method. We performed a user study where participants were asked to provide their personal photo collections and to select the 20% that they perceive as the most important for revisiting or preservation purposes. The 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.

Overall 91 annotated collections were collected from 42 users, resulting in 18,147 images. Near-duplicates have been unified considering the centroid of the set as representative photo, as done in [3]. Similarly to [6], each representative is marked as selected if at least one photo in its set has been marked as selected, and marked as not selected otherwise.

In order to assess personalization performances, we consider users who contributed at least 5 collections as test users. Among the overall 91 photo collections, there are 11 users who provided more than 5 collections (10 users contributed 5 collections, 1 user contributed 6 collections) which result in 56 collections totally. According to this, our dataset is split into two parts: one part contains 35 collections from 31 users, whereby each user provided at most 2 collections, which is called *general dataset*; another part contains 56 collections from 11 users, whereby each user provided at least 5 collections, which is called *personalized dataset*.

**Evaluation Metrics.** The selection method employed in this paper can generate a selection $S$ of size $n$ from the original collection, where $n$ can assume different values. We evaluate our method considering the precision P@$n$ of the selection $S$ of size $n$ that they produce, computed as the ratio between number of photos in $S$ that were originally selected by the user and the size of $S$. Since the collections in our dataset have high size variability, we decided to express $n$ as a percentage of the collection size, instead of an absolute value. In particular, we compute the precision for $n = 20\%$, which is indicated as P@20%, coherently with our user study where participants were asked to select the 20% most important photos from their collections. In order to assess the adaptation of our personalized model to users, we apply the personalization process described in Section 4 to the collections of each user separately and average the P@20% among the test collections available at each iteration $k$, where $k$ denotes the number of collections that are used for training the personalized model.

**Parameter Settings.** The classifier employed in this paper for importance prediction, built using the Support Vector Machine implementation of LibSVM[1], has Gaussian Kernels and has been trained via 10–fold cross validation on the training set. Note that the training set is expanded at each iteration (i.e. each time a new annotated collection of the user is provided), and the training via 10–fold cross validation is repeated each time. The open parameters were tuned via grid search and updated at each iteration. The ones identified for the *general dataset* where $C = 1.5$, $\gamma = 0.25$.

### 5.2. Training and Test Sets

We evaluate the performances of the model update (Section 4.3) over different rounds of adaptation. The *personalized dataset* is split by users where each user owns 5 collections (one user owns 6). At each iteration $k$, for each user with $N$ collections, $k$ collections are added to the initial training set to learn the personalized model of the user, and $N-k$ collections are used for testing. The ways in which the original training set is built are described in Section 5.3.

We experiment all the values of $k$ ($k = 0, 1, 2, 3, 4$), and for each of them we repeat the split and evaluation 5 times so that all the collections could be selected the same times as training collections. Note that the iteration $k = 0$ corresponds to the situation when the selection model is trained only on the initial training set. The selection strategy to select training collections is the following. When $k = 1$, we ensure that each collection of the user that we are considering is selected once as initial training data and the remaining four collections are treated as test data, then we average the performances. When $k = 2$, we pick two collections at each time from 5 collections, with the constraint that each collection could only be selected twice in all 5 repetitions (to be fair to all collections). We then average the performance achieved at each time. The cases when $k = 3$ and $k = 4$ can be done in the same manner. Finally, we average the performances over users for the same value of $k$.

### 5.3. Different Training Sets

The three considered ways of building training sets are described hereafter. The model update and the split in train and test set previously described are the same in each case.

**Stand-alone.** The initial model is trained with one random collection of the user, and the model update is incre-

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

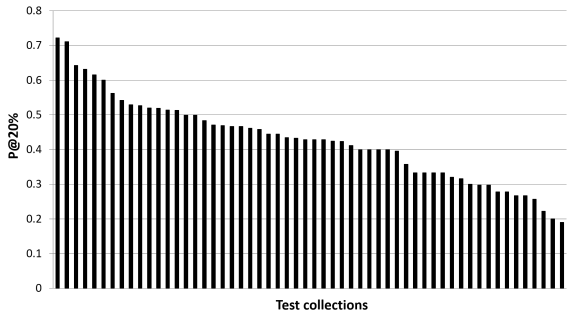


**Fig. 2**. P@20% for test collections (training on general data).

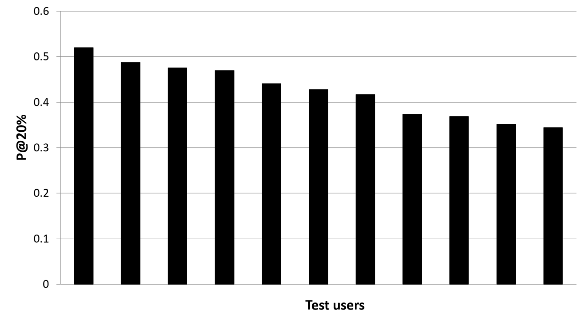


**Fig. 3**. P@20% for test users (training on general data).

mentally done considering the remaining collections (starting from iteration k = 1). The iteration $k = 0$ is not considered since the training set would be empty at this stage.

**Collaborative.** The initial training set at $k = 0$ is formed by all the collections within the *general dataset*. This case represents the situation where, in absence of large amount of annotated personal data for training, annotated collections of other users are used to alleviate the cold-start problem.

**User-agnostic.** Similarly to the *collaborative* case, the *general dataset* is used as initial training set. However, at each iteration $k$, instead of including $k$ collections of the user that we are considering, we add $k$ randomly selected collections from the other test users. This case is motivated by the assumption that, if one collection, which is not from the user that we are considering, is included in the training set at each iteration, then the adaptation performances should be smaller than including collections that are from the user that we are considering. This would highlight the importance of incorporating selection information of the user in the training set when making selections for new collections of the same user.

### 5.4. Results

As a motivation to the need of personalization in photo selection, we trained a not personalized selection model on the *general dataset* and we tested its performances (P@20%) on the *personalized dataset*. The results are showed in Figure 2 (split by collections) and Figure 3 (split by users). We can observe a large amount of variability in performances over the different collections (Figure 2), with precision values ranging between 0.190 and 0.722. The same behavior can be noted when grouping collections by test users (Figure 3), although the differences in performances are less prominent. This shows that a single selection model has limitations in meeting the expectations and preferences of different users, and the overall performances of the system could be improved by learning selection models personalized to each single user.

The results of our personalization procedure, considering the three different ways of constructing the training set described before, are shown in Table 1. Along with the precision when selecting the 20% of the original collection (P@20%) and its standard deviation over the test users, we also explicitly report the relative gain ($\Delta$) obtained between two consecutive iterations. For instance, the $\Delta$ for $k = 3$ represents the relative gain in P@20% with respect to the one achieved for $k = 2$. It is possible to observe that the precision of both *stand-alone* and *collaborative* increases at each iteration, i.e. with the increase of the number of user's collections considered for training the model. This shows that having a selection model partially aware of the user preferences (by exploiting a certain amount of the selection behavior in the training phase) can improve the precision of new unseen collections of the same user. The precision of *collaborative* is higher than the one of *stand-alone*, especially at the first iterations, showing that the selection data from other users can alleviate the cold-start problem. The gain $\Delta$ of *stand-alone* at each iteration is higher than the one of *collaborative*, because the initial model is weaker (due to the limited training set) and the inclusion of new training collections has a higher impact on the learning. It is important to clarify that the standard deviation observed in these experiments is relatively high. This can be due to a mixture of aspects, such as (i) a limited size of test set (both in terms of users and iterations), (ii) intrinsic changes of difficulty among collections of the same user. For this reason, although a promising adaptation to the user emerges from our results, the inclusion of a wider data set would be required to show it more significantly.

Comparing *user-agnostic* and *collaborative*, the former exhibits an almost null gain in performances over iterations (it is even negative for $k = 4$), while the latter leads to a higher and increasing performance gain iteration after iteration. This shows that the increase of performance at each iteration is due to the inclusion of a new collection of the same user in the training set and not simply caused by the fact that the training set is expanded at each iteration, since in this case the gain of *user-agnostic* should have been higher as well. Given the relatively high values of standard deviation, this promising result would require an extended number of test collections and iterations to be more evident and statistically significant.

| | k = 0 | | k = 1 | | k = 2 | | k = 3 | | k = 4 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **P@20%** | **Δ** | **P@20%** | **Δ** | **P@20%** | **Δ** | **P@20%** | **Δ** | **P@20%** | **Δ** |
| **Stand-alone** | - | - | 0.353 ± 0.060 | - | 0.374 ± 0.068 | +5.9% | 0.383 ± 0.067 | +2.4% | 0.402 ± 0.069 | +5.0% |
| **Collaborative** | 0.427 ± 0.057 | - | 0.430 ± 0.054 | +0.7% | 0.432 ± 0.055 | +0.5% | 0.437 ± 0.050 | +1.2% | 0.444 ± 0.061 | +1.6% |
| **User-agnostic** | 0.427 ± 0.057 | - | 0.427 ± 0.052 | +0.0% | 0.428 ± 0.055 | +0.2% | 0.429 ± 0.053 | +0.2% | 0.426 ± 0.055 | -0.7% |

**Table 1**. P@20%, standard deviation, and performance gain of the personalized models at each iteration.

## 6. CONCLUSION

In this paper, we presented an approach to personalized photo selection that adapts to user preferences by exploiting user feedback. It is based on re-training the selection model every time a new annotated collection of the user is available. Our evaluation led to promising results, showing that (i) including new annotated collections for the same user when training the model can benefit the selections on new unseen collections of the same user, and (ii) exploiting annotated collections from other users as initial training data can boost the system performances in cold-start scenarios. Given the high standard deviation of performances observed in our analysis, an evaluation on a larger number of users and personal collections would be required to make our results more evident and statistically significant. We also plan to experiment different strategies of incremental and on-line learning, which would make our approach less computationally demanding.

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

## 7. REFERENCES

[1] A. Ceroni, V. Solachidis, M. Fu, N. Kanhabua, O. Papadopoulou, C. Niederée, and V. Mezaris, "Investigating human behaviors in selecting personal photos to preserve memories," in *Proc. of ICME Workshops (ICMEW)*, 2015.

[2] A. Ceroni, V. Solachidis, C. Niederée, O. Papadopoulou, N. Kanhabua, and V. Mezaris, "To keep or not to keep: An expectation–oriented photo selection method for personal photo collections," in *Proc. of ICMR*, 2015.

[3] 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.

[4] J. Li, J. H. Lim, and Q. Tian, "Automatic summarization for personal digital photos," in *Proc. of ICICS-PCM*, 2003.

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

[6] A. E. Savakis, S. P. Etz, and A. C. P. Loui, "Evaluation of image appeal in consumer photography," in *Proc. of SPIE*, 2000.

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

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

[9] T. Walber, A. Scherp, and S. Staab, "Smart photo selection: Interpret gaze as personal interest," in *Proc. of CHI '14*, 2014.

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

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

[12] L. Huang, T. Xia, J. Wan, Y. Zhang, and S. Lin, "Personalized portraits ranking," 2011, Proc. of MM '11.

[13] C.-H. Yeh, B. A. Barsky, and M. Ouhyoung, "Personalized photograph ranking and selection system considering positive and negative user feedback," *ACM Trans. Multimedia Comput. Commun. Appl.*, 2014.

[14] Y. Rui and T. Huang, "Optimizing learning in image retrieval," in *Proc. of CVPR*, 2000.

[15] J. Zhang, L. Zhuo, L. Shen, and L. He, "A personalized image retrieval based on user interest model," *IJPRAI*, 2010.

[16] S. Uchihashi and T. Kanade, "Content-free image retrieval based on relations exploited from user feedbacks," in *Proc. of ICME*, 2005.

[17] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[18] P. Obrador, R. de Oliveira, and N. Oliver, "Supporting personal photo storytelling for social albums," 2010, Proc. of MM.

[19] A. Bordes, S. Ertekin, J. Weston, and L. Bottou, "Fast kernel classifiers with online and active learning," *Journal of Machine Learning Research*, 2005.

[20] G. Cauwenberghs and T. Poggio, "Incremental and decremental support vector machine learning," in *Proc. of NIPS*, 2001.