Modeling Image Appeal Based on Crowd Preferences for Automated Person-Centric Collage Creation

Vassilios Vonikakis, Ramanathan Subramanian, Jonas Arnfred, Stefan Winkler
Advanced Digital Sciences Center (ADSC)
University of Illinois at Urbana-Champaign, Singapore
bbonik, Subramanian.R, j.arnfred, Stefan.Winkler@adsc.com.sg

ABSTRACT

This paper attempts to model *Image Appeal* of photos portraying people, with the objective of automatically ranking and selecting the most appealing ones for the creation of interesting person-centric collages/collections. To understand the notion of image appeal, we employed *crowdsourcing*, using 350 workers who were asked to select a representative subset of images from five different person-centric album themes (involving a *man*, *woman*, *couple*, *girl* and *baby*). The albums were previously balanced with respect to nine different image attributes using Binary Integer Programming. The crowdsourcing study revealed identifiable patterns in the photo selection process, with more appealing photos securing more hits than less appealing ones. We then employed nine low-level image features and Support Vector Regressors to model photo selection statistics– the best model explained 63% of the selection patterns, and our analyses also confirmed the role of *context* in influencing Image Appeal. Finally, Image Appeal predictions on *unseen* photos are presented to demonstrate the promise of our approach.

Categories and Subject Descriptors

H.1.2 [**User/Machine Systems**]: Human information processing; I.5.2 [**Pattern Recognition Design Methodology**]: Pattern analysis

General Terms

Measurement, Algorithms, Verification, Human Factors

Keywords

Image Appeal, Modeling, Crowdsourcing, Collage Synthesis

1. INTRODUCTION

The affordability of digital images nowadays has led to an ever-increasing size of personal photo collections. It is now fairly common to take 2-3 shots of the same scene, and over a few hundred photos of special events (*e.g.*, vacation trip) in order to capture as many memorable moments as possible. However, users typically adopt a manual and painstaking task of identifying the *best* photos from a collection, in order to synthesize a collage for sharing with relatives and friends. Consequently, there is a definite need for developing automated tools for photo selection from a large collection, also known as *photo triaging*.

The seminal work of Savakis et al. [15] demonstrates that *Image Appeal* (IA) plays a crucial role in the selection of images in personal photo collections. According to the definition in [10], *"IA is the interest that a photograph generates when viewed by human observers, incorporating subjective factors on top of the traditional objective quality measures"*. As such, IA is a broader term than Image Quality (IQ), incorporating additional factors and mainly referring to photographic images. In other words, a photo with low IQ need not necessarily have low IA.

Although many prior works has leveraged on crowdsourcing for IQ or image selection for summarization, very few have directly attempted to model IA. In [16], the authors implement a photo summarization framework optimizing three properties– quality, diversity and coverage, but no user study is performed to guide the summarization process or validate the generated photo summaries. In [3], a crowd-powered camera is presented, where workers quickly filter a short video down to the best single moment for a photo. Summarization of image collections is presented in [14]. Based on crowdsourcing data, the authors propose an automatic image selection approach, which jointly utilizes the analysis of image content, context, popularity, visual aesthetics and sentiment derived from comments posted on social media.

One of the first works to identify the significance of IA in photo-collections is [15]. The authors analyze a series of factors that may affect IA, observing that it is a cumulative function of scene perspective, composition aspects, existence of faces, character pose and action, and basic image quality among many others. Very few works have sought to understand and model IA, specifically from a user-perspective. Another work [10] studies IA through the prism of photographers to automatically model two metrics, one for IA-based image ranking, and another to retrieve appealing images from a collection. Examination of how IA influences selections in photo sequences and correlates with users' visual attention patterns is presented in [6]– however, no attempt is made to automatically replicate observed image selections in this work. A pilot-study employing 14 users to identify low

CrowdMM'14, November 7, 2014, Orlando, FL, USA.
Copyright 2014 ACM 978-1-4503-3128-9/14/11 ...$15.00.
http://dx.doi.org/10.1145/2660114.2660126.

and high-level image attributes influence image selections on a *child*-themed photo album is described in [18].

All the aforementioned studies, with the exception of [18] (which is limited in terms of number of users and album themes), focus on general photo-collections, that may include people, scenery, pets, food, objects etc. However, it has been shown that the majority of personal photo-collections comprise images portraying people and faces, and such images have been found to be most popular in social media [2]. For this reason, we are interested in modeling IA specifically in people-centric photo collections.

In this work, a full-scale crowdsourcing experiment is conducted to understand the factors underlying IA on a large person-centric photo-collection. The 350 workers were asked to select the most appealing images from five different albums containing photos with a *man*, *woman*, *couple*, *girl* and *baby* as central characters. All photo albums were balanced across nine different image attributes (described in Sec. 3) using optimization. The study confirmed that the photo selection process was not random, and some photos (which we term as highly appealing) were consistently selected by workers as compared to other (less appealing) ones. Then, we trained Support Vector Regressors (SVRs) with the image selection probability P_{sel} as the dependent variable, and nine image attributes, which could have influenced the selection process, as predictors. A best-fit model with a coefficient of determination $R^2 = 0.63$ was obtained using the employed predictors, and our analysis also confirmed the influence of context in determining the most appealing photos. We then employed SVRs to predict IA for *unseen* photos, as discussed in Sec. 4.

In summary, we make the following contributions: (1) This paper represents the first work to employ crowdsourcing for understanding IA. (2) This is also the first work to attempt user-centric IA modeling, which enables understanding of how context influences photo selections in person-centric albums. (3) A Binary Integer Programming-based optimization technique was employed to synthesize image sets balanced with respect to multiple features for the crowdsourcing study, which can be utilized by future works as well.

The rest of the paper is organized as follows: Section 2 describes the experimental procedure followed in the crowdsourcing experiment, along with the technique for generating balanced image datasets. Section 3 gives a detailed description of the IA modeling that was used, while Section 4 presents the results and discussion. Finally, concluding remarks are included in Section 5.

2. EXPERIMENTAL PROTOCOL

Five person-centric photo albums with approximately 300 images each were used in the study. The albums' central characters were a *man*, a *woman*, a *couple*, a *girl* and a *baby*. The first three were selected from personal photo-collections, since no publicly available datasets depict the same individual(s) over an extended period of time– the three adult albums included images captured over a 11-year period, enabling us to examine the influence of time on IA. The girl and baby albums were part of the Gallagher dataset [5].

2.1 Creating balanced subset of images

Given a large initial pool of photos, unbalanced with respect to attributes that we are interested in analyzing, a first very important step is to select a small and manageable subset of these images, in which, the attributes we are interested in are equally represented. The main reason for this is that, large image sets are difficult to browse, especially in a crowdsourcing setting. Consequently, the crowdsourcing scenario necessitates narrowing down the number of images the workers have to analyze. This however, is not a straightforward task and may result in an unbalanced subset in which, image attributes which we may be interested in are underrepresented. For example, if a researcher wants to examine the contribution of colorfulness and scene-type on IA, a subset comprising mostly outdoor images would be inadequate since no indoor scenes are included. The larger the number of image attributes we are interested in equalizing, the more complicated the combinatorial problem of arriving at a balanced set becomes, since exclusion (or inclusion) of a particular image may impact multiple other attributes. Fig. 1 depicts a toy example of the dataset balancing concept for 2 simple attributes.

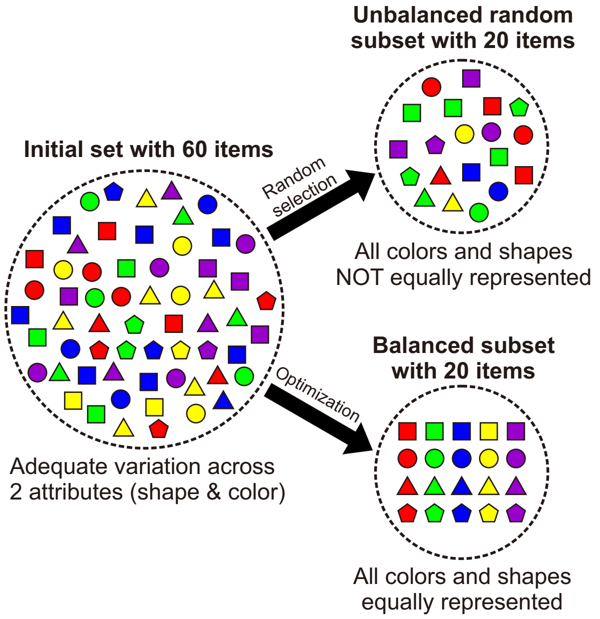

Figure 1: Toy example of the balancing concept for 2 attributes (color and shape).

Let K be the total number of images of the initial set, sufficiently large and involving a large variance in the distributions of the M image attributes which we are interested to balance. The size requirement is important as otherwise, the optimization problem may be infeasible. $\mathbf{A} = [A_{ij}]_{M \times K}$ is the matrix that contains the M attribute values for the K initial images, with $A_{ij} \in \mathbb{R}$ and normalized within the interval $[0, 1]$. Let N be the number of images in the balanced subset, with $N << K$. We formulate the selection process of the N images as a Binary Integer Programming (BIP) problem, where the vector x contains K binary variables representing the selection/elimination of each of the K images in the initial set. Since we are interested in selecting N images, it entails that:

$$\sum_{i=1}^{K} x_i = N \text{ with } K, N \in \mathbb{Z} \quad (1)$$

In order to ensure that all attributes are equally represented, we would like to minimize the distance between each attribute's mean in the final N images and $1/2$, thus forcing a mean value close to $1/2$ for all attributes. This minimization can be expressed as:

$$\sum_{i=1}^{M} \left| mean(A_i) - \frac{1}{2} \right| = \sum_{i=1}^{M} \left| \frac{\sum_{j=1}^{K} A_{ij} x_j}{\sum_{j=1}^{K} x_j} - \frac{1}{2} \right| =$$
$$= \sum_{i=1}^{M} \left| \sum_{j=1}^{K} (A_{ij} x_j) - \frac{N}{2} \right|$$

In order to handle the M absolute values, we introduce M slack variables $(C_1, \cdots, C_M)$, one for each attribute. Consequently, the final BIP becomes:

$$\text{Minimize} \sum_{i=1}^{M} C_i \quad (2)$$

$$\text{s.t.} \sum_{j=1}^{K} x_j = N,$$
$$C_i \geq \sum_{j=1}^{K} (A_{ij} x_j) - \frac{N}{2},$$
$$C_i \geq \frac{N}{2} - \sum_{j=1}^{K} (A_{ij} x_j)$$

In order to end up with a relatively manageable balanced subset of images, and ensure that the workers have a manageable task on hand, we solved the BIP by selecting ($N = 60$) images, balanced with respect to ($M = 9$) attributes, from the initial ($K = 300$) photos for each album. These attributes are described in Section 3.

2.2 Crowdsourcing

The crowdsourcing experiment was set up in the Microworkers platform [1]. Workers were presented with the balanced subsets for each of the five albums, and were asked to select any number of images from each subset in order to create a photo collage for the particular person(s) depicted. Prior to the task, workers had to answer some personal screening questions in order to assess their demographic background and their level of experience with digital photography.

2.2.1 Pilots

Three pilot trials were set up prior to the final experiment in order to identify a good combination of *incentives*, *interface*, and *instructions* that would yield higher quality of results in the final experiment. Apart from these 3 characteristics, all pilot trials were otherwise identical to the final experiment.

During the first trial, 50 workers were awarded with 0.5$ for participating in the experiment. The average time spent during the image selection process was approximately 147 seconds. However, the quality of the results, was found to be lower than expected.

In order to increase the engagement of workers in the selection process, an additional textbox was added during the second trial, in which, the workers had to justify the reasons for selecting the images they did. At the same time, in order to investigate the impact of compensation in the quality of the results, the awarded amount was reduced to 0.4$. 50 workers participated in the second trial. It was found that, although the compensation was reduced, the justification that the workers had to give, resulted in an increased average time spent during the selection process (approximately 171sec). Higher quality of results were also observed, in terms of rejection rate of workers who did not meet our minimum set of specifications (see section 2.2.2).

During the third trial, the basic compensation of the workers was reduced in half (0.2$). At the same time, it was explicitly mentioned that an equally big bonus of 0.2$ would be given to those workers who gave a good justification of the reasons for selecting the images they did. This approach, which has been also used in other crowdsourcing tasks [8], was selected in order to provide incentives for increasing the workers' engagement during the image selection task. 20 workers were hired during the third trial and their results were manually evaluated. It was found that this particular combination of incentives resulted in the best quality of crowdsourcing results, compared to any of the previous trials. For this reason, this particular setup was selected to be used in the final experiment.

2.2.2 Evaluation of workers

One of the most challenging issues in crowdsourcing is to identify which workers provide valid results. In our study, each worker went through 3 different screening levels that had to do with his/her relevance to the nature of the study, possible inconsistencies in the provided answers, and quality of submitted results. The first two filtering levels had to do with the initial screening session that included the following seven questions:

1. *How old are you?* (<10/10–19/20–29/30–39/>40)
2. *What is your gender?* (M/F/Other)
3. *How many cameras have you owned?* (0/1/>1)
4. *Do you own a smart-phone?* (Y/N)
5. *How many years approximately is it since you had your first digital camera?* (<5/5–10/>10/0)
6. *How many images approximately do you take on average every month?* (<20/20–100/>100)
7. *How many images does your photo-library contain?* (<1000/1000–5000/>5000)

First, since we are interested in photo-collections and images, we wanted to hire workers that had at least some minimum experience with photos, and eliminate those who do not use at all digital images in their daily lives. Consequently, the results of any worker who declared that was less than 10 years old, or never had a digital camera or a smart-phone, were not used in the experiment.

In the second level we attempted to identify inconsistencies in the workers' answers for questions 3–7. Based on the given answers we tried to estimate the minimum and maximum number of photos that someone could have in his/her library, given the declared years of possessing a camera (or a smartphone) and the average number of photos taken every month. Particularly, if one of the following impossible conditions is satisfied, there is a strong indication that the worker has provided false answers to questions 3–7, and thus,

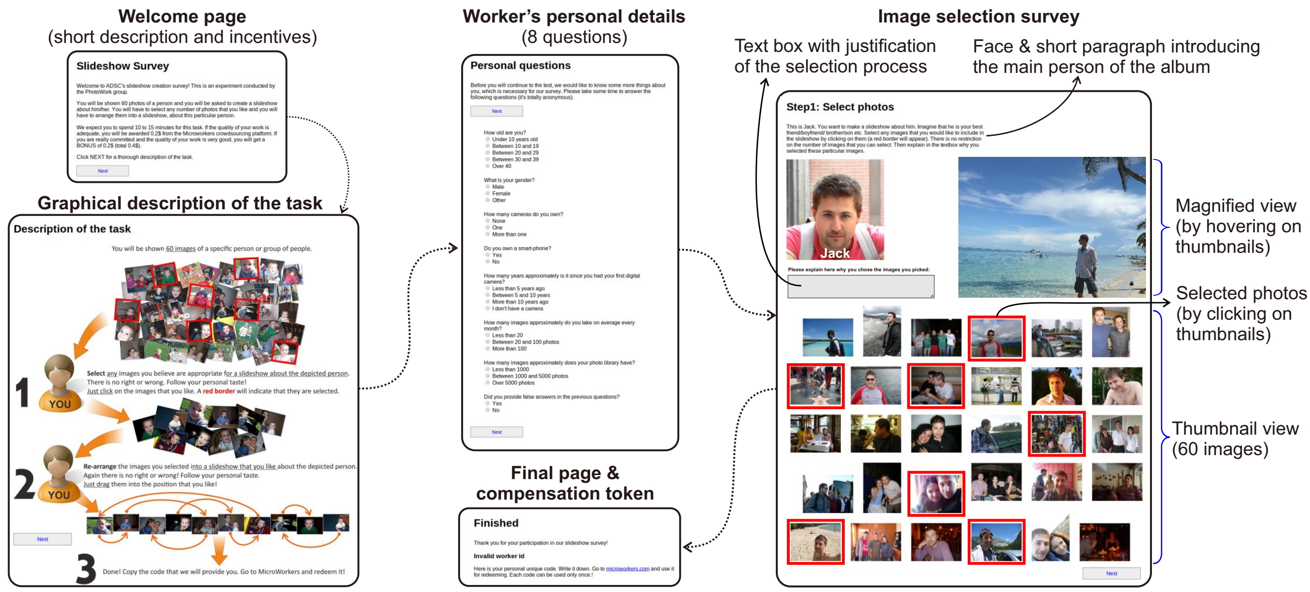

Figure 2: Flowchart and description of the crowdsourcing survey.

his/her results are not used in the study.

$$Y_{min} \times M_{min} \times 12 > L_{max} \quad (3)$$
$$Y_{max} \times M_{max} \times 12 < L_{min} \quad (4)$$

where Y_{min} and Y_{max} are the minimum and maximum years, respectively, of a person possessing an imaging device (as indicated by questions 3–5), M_{min} and M_{max} are the minimum and maximum number of photos that a person captures in a monthly basis (as indicated by question 6) and L_{min} and L_{max} the minimum and maximum number of photos that a person's photolibrary contains (as indicated by question 7). If a worker declared having a smartphone but never possessed a camera, then 2007 was used as as a starting year of his/her photolibrary, which is when smartphones came into vogue.

If a worker passed through the first two screening stages, we analyzed the results that he/she has provided. Particularly, we wanted to eliminate workers who spent small amount of time during the selection process, selected very few images, or did not provide adequate justification for their choices. Previous works have utilized statistical criteria in order to identify outlier results and eliminate them from the processing [13]. Our main objective was to eliminate very obvious cases of inattention. As such, we attempted to identify the baseline conditions according to which a submitted result could be considered valid. Failing to meet these criteria would essentially mean that it would be impossible for a submitted work to be valid.

In this direction, we performed trial tests, with subjects interacting for the first time with the interface, and recorded the time they needed in order to browse all 60 photos. It was found that the amount of time someone has to devote in order to even briefly browse all 60 photos, cannot be less than 45 seconds (0.75 sec/photo). In other words, it is unrealistic for anyone to spend less than 45 sec and pay equal attention to all 60 images. Consequently, the results of any worker who did not satisfy this simple timing condition were not used in the study. This approach has also been used in many other crowdsourcing tasks [8]. Similarly, statistics for the English language suggest that the the median length of a sentence is approximately 75 characters ($\approx$ 15 words/sentence $\times$ 5 characters/word) [9]. Using this number as a guideline and analyzing the received justifications during trials, we concluded that the minimum size of a small sentence that can actually convey useful information regarding a worker's selections, is 30. Consequently, any provided justification below 30 characters, was highly unlikely to include a meaningful comment, and thus it was eliminated. Finally, we set a lower threshold of 5 images for the total number of photos that were selected.

At the end of this elimination process, 71.4% of the workers were qualified as reliable (250 valid workers out of the 350 initial ones) and were selected to participate in the study.

2.2.3 *Main Experiment*

The flowchart of the final experiment is depicted in Fig. 2. Workers were randomly assigned to one of the 5 different albums (resulting approximately to 50 workers/album) and were presented with an interface which featured three different regions. The upper part included a brief description of the task. In order to make the workers more engaged, we tried to associate them with the person depicted in the dataset, by providing a name for each of the album central characters and including messages like (depicted here for the *man* dataset):

"*You need to make a photo collage of Jack. Imagine that he is your best friend/brother/son. Select any images that you would like to include in the collage by clicking on them (a red border will appear). There is no restriction on the number of images that you can select. Then explain in the textbox why you selected the particular images*".

The lower part of the interface included thumbnails of the 60 images (in random positions), along with a larger picture viewer, which displayed a magnification of any thumbnail on which the mouse pointer hovered. This way, workers could effortlessly browse all images, both by looking at them as a whole (thumbnails) and in a more detailed view. Clicking on a thumbnail would activate a red border around the thumbnail, indicating that this particular image was selected. Finally, workers had to type specific comments as to why they selected the particular images they did. This was included in order to force the workers to be more engaged in the task. For each worker, 4 different pieces of information were recorded: the images that he/she picked, the total number of selected images, the total time spent during the selection process and the personal comments describing why he/she selected these particular images.

2.2.4 Crowdsourcing results

Analysis of the crowdsourcing results showed that the average and median number of selected images over all albums were 17.05 and 13 respectively, which is in accordance with the previous findings of [18]. Another interesting result is that all 60 images of all 5 albums were selected at least once, *i.e.*, there were no images that were never selected. We computed the selection probability for each image P_{sel} as the number of times the image appeared in a valid selection list, normalized by the total number of reliable workers. For all albums, there was significant variance in the P_{sel} values, with $max(P_{sel}) \approx 0.6$, and $min(P_{sel}) \approx 0.1$. A notable exception was the *baby* album for which $max(P_{sel}) = 0.44$. Considerable variance in the P_{sel} values confirmed that the image selection process was not random. Rather, some underlying factors should have influenced the frequent selection of some images, or in other words, contributed to their enhanced appeal.

3. MODELING IMAGE APPEAL

We modeled 9 low-level factors to determine their influence on photo selections. The following attributes were modeled:

1. **Face count:** The Viola-Jones [17] face detector (frontal and profile) was used, in conjunction with skin detection, in order to eliminate false-positive detections. This attribute can indicate whether an image contains an individual or a group.
2. **Face size:** The ratio of the size of facial bounding boxes over the image size was used. This attribute can indicate the type of the shot, *i.e.*, close-up or full-body.
3. **Scene type:** Indoors/outdoors. This attribute is not a binary one, since many intermediate cases can exist (*e.g.*an image of a room with an open window depicting natural scenery). For this reason we employed the *Relative Attributes* approach proposed in [12], which ranks images according to their relative strength for each attribute. We employed gist features and color histograms and trained the system to compute a real-valued rank specifying the indoor/outdoor-ness for each image.
4. **Age of depicted person:** Facial appearance changes over time, and the adult photo albums were compiled over a year period. Using EXIF timestamps, we computed the time durations between captured photos in order to examine if primacy/recency effects influenced the image selection process.
5. **Sharpness:** Overall perceived sharpness of an image, computed similar to [11].
6. **Contrast:** Overall perceived contrast of an image, as computed in [11].
7. **Colorfulness:** Perceived colorfulness of an image (vividness of colors), as computed in [7].
8. **Exposure:** Overall perceived exposure levels of an image, computed similar to [11].
9. **Combined Aesthetics:** Combination of attributes 5-8 into a single metric, computed similar to [11].

All the above attributes were normalized to the interval $[0, 1]$ using *min-max* normalization. Some of the above attributes (sharpness, colorfulness, contrast and exposure) were selected based on the basis of previous studies ([15], [10]). Others, such as *face size* and *face count* were selected to describe *closeup* and *group* photos, which were noted to influence IA in [15]. Finally, attributes like *scene type* and *age of depicted person* have not been considered by previous studies while studying IA, and were considered as intuitively, they could be related to IA.

Assuming the above factors to be predictors influencing the characterization of a photo as highly appealing (associated with high P_{sel} values) or less appealing (low P_{sel}), we trained linear, polynomial and radial-basis function (RBF) support vector regressors (SVRs) available as part of the *libsvm* [4] package. To examine if context plays an important role in determining IA (*e.g.*, if factors underlying IA for the *baby* and *couple* albums are similar or different), we trained an SVR for (crowdsourced or training) images from each photo-album, and a generic SVR with training images from all albums. The best C, γ values for all models were obtained via grid-search based cross-validation.

4. RESULTS AND DISCUSSION

Regression results are presented in Table 1– coefficients of determination obtained using the generic (R_g^2) and specific (R_s^2) SVR models are shown for the five albums. As expected, RBF-SVR generates the best album-specific models, with a maximum 63% variance explained for *male* photo selections. Analyzing linear-SVR coefficients (Table 2), we noted that outdoor photos were generally preferred in *male*, *female* and *couple* photos, while time-stamps, face counts and sizes only marginally influenced photo selections. Sharpness and overall aesthetics also played a key role in influencing image appeal.

The generic SVR model explained P_{sel} characteristics poorly than album-specific models, implying that context is crucial towards determining the most interesting/appealing images. The model fits are particularly worse for the *girl* and *baby* albums, suggesting that these images are perceived differently compared to photo collections involving adults.

In order to automatically rank images in novel albums based on IA, we applied the album-specific RBF-SVR models to estimate P_{sel} for those photos that were not part of the crowdsourcing study. Fig. 3 presents the higher and lower exemplar results with their P_{sel} estimates for *unseen* images (not part of the crowdsourcing study). Fig. 4 depicts interesting results on crowdsourced (training) images where there is a significant discrepancy between model estimates and ground-truth P_{sel} values. Evidently, photos acquired under sufficient illumination, and with colorful backgrounds (as with outdoor images) are determined as having high ap-

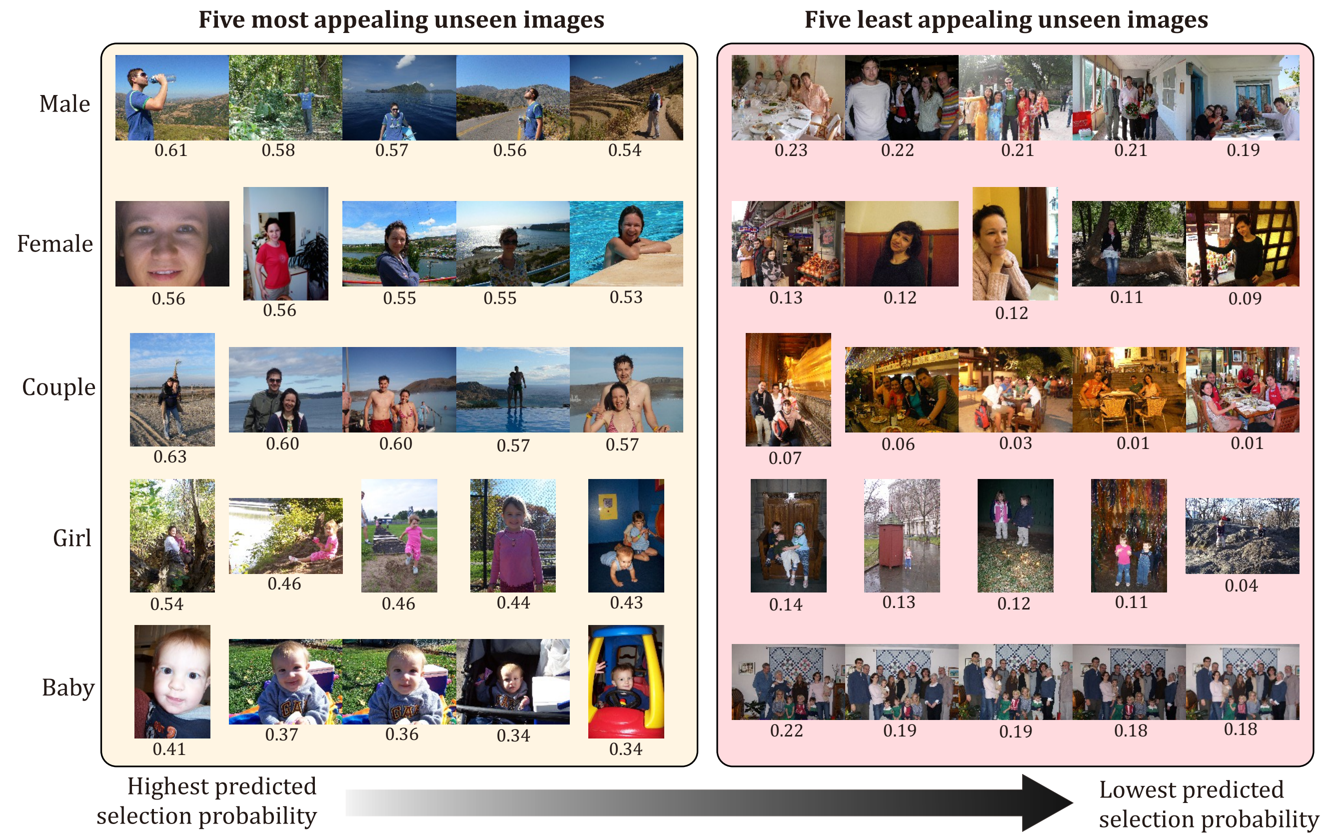

Figure 3: Top/lower 5 most/less appealing predicted unseen images for each of the 5 albums, along with their estimated selection probability P_{sel}.

peal, while those acquired under low lighting conditions and relatively plain backgrounds, typical of indoor settings, are deemed as less appealing.

Comparisons between estimated and actual P_{sel} measures presented in the third and fourth columns indicate some limitations of our approach. This is mainly due to the fact that image appeal is not only determined by low-level factors but also high-level semantics, such as photos taken in unique settings (*e.g.*, underwater *couple* image) or containing interesting facial expressions (especially for *girl* and *baby* pictures), which we do not consider in our IA-modeling currently. Incorporating these aspects is part of our future work.

5. CONCLUSION

This paper attempts to model IA in personal photo collections through a user-centric perspective. To understand how users deemed an image as being more/less appealing, an extensive crowdsourcing experiment was conducted with 350 workers and five different albums. The significant variance in selection probabilities for the *most* and *least* appealing images indicated that images were not selected randomly, and there were underlying factors that influenced some images to be selected more often than others. We then employed nine low level image attributes to model the image selection process, and trained SVRs which could adequately predict image selections for the album-specific conditions. However, a generic SVR failed to model the selection patterns as adequately as the album-specific SVRs suggesting that context

Table 1: Best SVM regression models for training images used in the crowdsourcing study. For the five photo-albums, coefficients of determination of P_{sel} with generic (R_g^2) and specific (R_s^2) models are shown (highest values per column are in bold font).

	Man		*Woman*		*Couple*		*Girl*		*Baby*	
Kernel	R_g^2	R_s^2	R_g^2	R_s^2	R_g^2	R_s^2	R_g^2	R_s^2	R_g^2	R_s^2
Linear	0.29	0.33	0.23	0.36	0.36	0.42	0.03	0.28	**0.26**	0.59
Poly	**0.36**	0.62	**0.37**	0.49	0.37	0.47	0.09	0.47	0.04	0.56
RBF	0.22	**0.63**	0.24	**0.50**	**0.41**	**0.50**	**0.23**	**0.47**	0.00	**0.60**

Table 2: Descending order of attribute significance for linear SVR model. Attribute indices are 1: Time, 2: Num.of faces, 3: Face size, 4: Indoors/outdoors, 5: Sharpness, 6: Contrast, 7: Colorfulness, 8: Exposure, 9: Combined Aesthetics

Album	Attribute significance								
Man	4	9	7	2	5	8	3	6	1
Woman	5	4	9	3	1	7	2	8	6
Couple	2	5	4	9	1	7	8	6	3
Girl	9	5	7	3	2	1	4	8	6
Baby	5	9	2	3	7	6	8	1	4

Figure 4: *Training* images for which regression estimates are significantly lower or higher than actual P_{sel}'s.

greatly influences the categorization of what is more and less appealing. Experimental results demonstrate that our approach is promising. However, more attributes (related to image semantics) are needed to accurately model image selection characteristics.

6. ACKNOWLEDGMENTS

This study is supported by the research grant for ADSC's Human Sixth Sense Programme from Singapore's Agency for Science, Technology and Research (A*STAR).

7. REFERENCES

[1] Microworkers. https://microworkers.com/, 2014.

[2] S. Bakhshi, D. A. Shamma, and E. Gilbert. Faces engage us: Photos with faces attract more likes and comments on instagram. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 965–974, New York, NY, USA, 2014. ACM.

[3] M. S. Bernstein, J. Brandt, R. C. Miller, and D. R. Karger. Crowds in two seconds: Enabling realtime crowd-powered interfaces. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, UIST '11, pages 33–42, New York, NY, USA, 2011. ACM.

[4] C.-C. Chang and C.-J. Lin. Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 27(2):1–27, 2004.

[5] A. C. Gallagher and T. Chen. Clothing cosegmentation for recognizing people. In *Computer Vision and Pattern Recognition*, pages 1–8, 2008.

[6] S. Gilani, R. Subramanian, H. Hua, S. Winkler, and S.-C. Yen. Impact of image appeal on visual attention during photo triaging. In *IEEE Int'l Conference Image Processing*, pages 231–235, 2013.

[7] D. Hasler and S. E. Susstrunk. Measuring colorfulness in natural images. In *Proc. SPIE*, pages 87–95, 2003.

[8] T. Hossfeld, C. Keimel, M. Hirth, B. Gardlo, J. Habigt, K. Diepold, and P. Tran-Gia. Best practices for qoe crowdtesting: Qoe assessment with crowdsourcing. *Multimedia, IEEE Transactions on*, 16(2):541–558, Feb 2014.

[9] A. Kornai. *Mathematical Linguistics*. Advanced Information and Knowledge Processing. Springer, 2008.

[10] P. Obrador. Region based image appeal metric for consumer photos. In *IEEE Workshop on Multimedia Signal Processing*, pages 696–701, 2008.

[11] P. Obrador and N. Moroney. Low level features for image appeal measurement. pages 1–12, 2009.

[12] D. Parikh and K. Grauman. Relative attributes. In *Int'l Conference on Computer Vision*, pages 503–510, 2011.

[13] J. A. Redi, T. Hossfeld, P. Korshunov, F. Mazza, I. Povoa, and C. Keimel. Crowdsourcing-based multimedia subjective evaluations: A case study on image recognizability and aesthetic appeal. In *Proceedings of the 2Nd ACM International Workshop on Crowdsourcing for Multimedia*, CrowdMM '13, pages 29–34, New York, NY, USA, 2013. ACM.

[14] S. Rudinac, M. Larson, and A. Hanjalic. Learning crowdsourced user preferences for visual summarization of image collections. *Multimedia, IEEE Transactions on*, 15(6):1231–1243, Oct 2013.

[15] A. E. Savakis, S. P. Etz, and A. C. P. Loui. Evaluation of image appeal in consumer photography. In *Proc. SPIE*, volume 3959, pages 111–120, 2000.

[16] P. Sinha, S. Mehrotra, and R. Jain. Summarization of personal photologs using multidimensional content and context. In *ACM International Conference on Multimedia Retrieval*, pages 4:1–4:8, 2011.

[17] P. Viola and M. J. Jones. Robust real-time face detection. *Int'l Journal of Computer Vision*, 57(2):137–154, 2004.

[18] V. Vonikakis, R. Subramanian, and S. Winkler. How do users make a people-centric slideshow? In *ACM Int'l Workshop on Crowdsourcing for Multimedia*, pages 13–14, 2013.