How Do Users Make a People-Centric Slideshow?

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

ABSTRACT

This paper presents a pilot user study that attempts to shed light on the ways users create people-centric slideshows, with the objective of scaling it up to a crowdsourcing experiment. The study focuses on two major directions, namely image selection and image sequencing. Participants were asked to select photos of a specific person from an initial set and arrange them into a slideshow. Results show that there is correlation between specific predictors and selected images, as well as their relative position in the final sequence. This indicates that a crowdsourcing experiment will indeed highlight the characteristics of the average user, which can then be incorporated into an automatic people-centric slideshow creator.

Categories and Subject Descriptors

H.1.2 [**User/Machine Systems**]: Human information processing

General Terms

Algorithms, Measurement, Human Factors

Keywords

Slideshows, Emotions, Crowdsourcing

1. INTRODUCTION

The affordability of digital images, mainly due to the popularity of smart-phones, has resulted in increasingly larger personal photolibraries. The majority of these comprise images of people in family moments, activities with friends, or travels. Browsing these libraries is tedious, especially if someone is interested in specific people.

Previous attempts to tackle this problem have focused mainly on event-based image sequences for photo-album summarization [3, 4]. However, since personal photolibraries are generally populated by images of people, those who are close to the user (e.g. family or friends) will appear in many different events. Consequently, event-based image sequences are not suitable for people-based browsing. A possible solution would be a people-centric automatic slideshow, that takes into consideration the identity of people, their portrayed emotions, and image aesthetics [5].

This paper attempts to answer the question of whether such an automatic system can be developed, using crowdsourcing to learn user preferences and patterns. More specifically, the study focuses on two directions: image selection and image sequencing. We first identify important criteria that users employ when selecting images for a people-centric slideshow, given a set of images of a person. We then identify certain rules that predict how people arrange the selected images into the final slideshow. Our results show that there are identifiable patterns in the way people select and arrange images, paving the way for further crowdsourcing studies using the proposed approach.

ACM MM '13 Barcelona, Spain

http://–enter the whole DOI string from rightsreview form confirmation.

Table 1: Description of the dataset

Category	Attributes	Images
Emotions	Positive	25 (50%)
	Neutral	19 (38%)
	Negative	6 (12%)
Scene	Indoors	18 (36%)
	Outdoors	32 (64%)
Depicted Person	Alone	31 (62%)
	With Others	19 (38%)
Aesthetics	Good	16 (32%)
	Average	20 (40%)
	Bad	14 (28%)

2. EXPERIMENTS & RESULTS

50 family photos were selected from the Gallagher dataset [1], depicting the baby of a family in various settings, expressions, activities, and groups of people. This particular theme (the baby) was selected over an adult, because it usually portrays a greater range of facial expressions and emotions. However, in a crowdsourcing experiment, different themes could also be explored (an adult, or a group of people). Table 1 depicts the characteristics of the selected dataset. Aesthetics were computed using a similar method to [4], and affective tagging was the average of a manual valence/arousal annotation, performed by two researchers.

14 subjects (7 males, 7 females, ages in the range of 22-36 years) participated in the survey. All of them owned at least

Table 2: Partial Pearson correlations ρ and corresponding significance values p relating image selection probability P with the different predictors.

		N	**AS**	**AR**	**VA**
P	ρ	0.0452	0.5803	-0.3346	0.3852
	p	0.7628	1.9×10^{-5}	0.0215	0.0075

Table 3: Regression analysis

Model Variables	R^2	F	p
AS	0.31	21.33	2.9×10^{-5}
AS, VA	0.33	11.78	7.13×10^{-5}
AS, VA, AR	0.41	10.72	1.84×10^{-5}
AS, VA, AR, SC	**0.51**	**11.81**	$\mathbf{1.22}\times10^{-6}$
All	0.51	9.24	4.54×10^{-6}

one digital camera and had photolibraries with more than 3000 images, which qualifies them as 'typical' users. The subjects were presented with the set of 50 images, which they were not familiar with, and were requested to create a slideshow for the main character (i.e. the baby). They were specifically asked to select any type or number of photos that they thought was appropriate for the slideshow, without any time or browsing constraints.

2.1 Image Selection

Following a similar approach to [2] about memorability, low-level and semantic-related factors were analyzed for possible statistical correlations with image selection, in the context of people-centric slideshows. Upon determining the number of times each image was selected by a user (selection probability P), we considered 4 factors that may influence image selection: number of persons N whose faces are visible in the image, aesthetics score (AS), valence (VA) and arousal (AR) of the portrayed facial expressions. Table 2 depicts the estimated correlation between P and the aforementioned factors. The probability P of an image being selected is positively correlated with AS and VA (smiling images were frequently selected) and negatively correlated with AR (images of the baby crying had higher AR values than the smiling ones). Secondly, assuming P as the dependent variable, and N, AS, AR, VA, SC (Scene Category - indoors/outdoors) as predictors, we performed a series of backward linear regression analyses with different sets of variables constituting the model. As Table 3 shows, AS alone explains 31% of the selection behavior. Addition of VA and AR factors improves prediction by 10%, while SC acquires predictive power in combination with the other factors. N is the worst predictor, and the best linear model can explain over 50% of the variance in the image selection patterns. A significant F-statistic is observed for all models. Each of these factors can be automatically estimated (AS using [4], VA/AR using emotion recognition algorithms, and SC using GIST features). Finally, the average slideshow length was 18.2 photos, out of which 11.3, 2.3 and 4.6 had positive, neutral, and negative VA, respectively, highlighting the preference of users for affective images.

2.2 Image Sequencing

The provided slideshows were analyzed in terms of beginning, middle, and end. Table 4 shows the average scores of the predictor factors, along with their standard deviations. At the beginning of the slideshow, users tend to select images in which the main character is depicted alone. This could be considered as establishing the *theme* of the slideshow. AS and AR are also the highest, indicating that the first images should be of very high aesthetic quality and depict more exciting expressions. Finally, there is a clear tendency to prefer indoor images at the beginning of the slideshow. There are no clear results for the middle segment, apart from the fact that they are of medium image quality. This indicates that a greater variability of expressions and scenes could be included here. Finally, the happiest expressions are generally saved for last, indicating a tendency to have a *happy ending* in the slideshow, while maintaining a very high AS.

Table 4: Mean and standard deviation of predictors

Attributes	Start	Middle	End
SC (1=indoors)	**0.71** (0.47)	0.48 (0.5)	0.57 (0.51)
N (1=alone)	**0.93** (0.27)	0.59 (0.5)	0.64 (0.5)
AS (1=best)	**0.75** (0.27)	0.59 (0.28)	0.74 (0.35)
AR $\in [-1, 1]$	**0.06** (0.05)	0.01 (0.12)	0.03 (0.13)
VA $\in [-1, 1]$	0.33 (0.32)	0.24 (0.39)	**0.43** (0.48)

3. CONCLUSIONS

This pilot study revealed interesting patterns in the way users select and arrange photos into people-centric slideshows. We intend to conduct a crowdsourcing experiment with a greater number of workers and themes to further investigate the preferences of the average user or user groups, which can then be incorporated into an automatic people-centric slideshow creator.

4. 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).

5. REFERENCES

[1] A. C. Gallagher and T. Chen. Clothing cosegmentation for recognizing people. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, Anchorage, AK, 2008.

[2] P. Isola, J. Xiao, A. Torralba, and A. Oliva. What makes an image memorable? In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 145–152, Colorado Springs, CO, 2011.

[3] I. Ivanov, P. Vajda, J.-S. Lee, and T. Ebrahimi. Epitome: A social game for photo album summarization. In *Proc. 1st ACM International Workshop on Connected Multimedia (CMM)*, pages 33–38, Firenze, Italy, 2010.

[4] P. Obrador, R. de Oliveira, and N. Oliver. Supporting personal photo storytelling for social albums. In *Proc. ACM International Conference on Multimedia*, pages 561–570, Florence, Italy, 2010.

[5] V. Vonikakis and S. Winkler. Emotion-based sequence of family photos. In *Proc. ACM International Conference on Multimedia*, Nara, Japan, 2012.