Automatic Rating and Selection of Digital Photographs

Daniel Kormann, Peter Dunker, and Ronny Paduschek

Fraunhofer Institute for Digital Media Technology, Ehrenbergstrasse 31,
98693 Ilmenau, Germany
kormdl@idmt.fraunhofer.de, peter.dunker@ieee.org, pdk@idmt.fraunhofer.de

Abstract. This paper presents methods for automatically rating and selecting digital photographs. The importance of each photograph is estimated by analyzing its content as well as its time-metadata. The presence of people is estimated by combining face and skin detection. Finally, the appeal of each photograph is calculated using a trained SVM classifier. The results of a conducted user study show that the automatically obtained rating coincides well with the perception of the test persons.

1 Introduction

The digitalization has led to an ever increasing size of personal photo collections. Today, consumers often take hundreds or even thousands of pictures of an event. For many intended purposes like the creation of slideshows or image galleries, a rating of the images would be desirable, so that the top-scoring images could be automatically selected. Existing approaches solve this task by screening for low quality images or by considering certain image content aspects like colorfulness [1] . Other works deal with the estimation of aesthetics in digital images [2]. Compared to the above mentioned work, our method focuses on the selection of the "best" images by considering various criteria.

2 Overview

We want to single out the aspects upon which a human observer would rate digital photographs. Based on a study of Savakis et al. [7] as well as our own conducted online survey, the following three criteria can be stated, which are all considered within our system, resulting in separate scores:

Image appeal - Is the image appealing, is it a successful photography?

Image importance - Does the image show an important subject or event?

Presence of people - Does the image show people, friends and family?

2.1 Image Importance

In order to calculate an *importance score* for each image i, we consider three aspects: First, if something interesting is happening during an event, the photographic rate rises. Hence, we calculate the local photographic rate f_i for each

T.-S. Chua et al. (Eds.): SAMT 2009, LNCS 5887, pp. 192–195, 2009.

image i using a weighted window. Second, if a photographer takes many pictures of one subject, this implies that the subject is of a certain importance to him. To detect whether two images are showing the same subject, the SIFT algorithm is used [6]. If there is a reasonable number of matching SIFT keypoint descriptors between two images, they are regarded as showing the same subject. s_i denotes the number of images showing the same subject as image i. Third, if a photographer takes two or more near identical images, it implies that this particular shot is of special importance to him. Such duplicate images are detected using the MPEG-7 color layout descriptor [4], where an empirically determined threshold defines whether two images are regarded as duplicates. Similar to above, d_i denotes the number of duplicate images of image i.

A weighted combination results in a final importance score $s^i_{importance} = u_1 * f_i/f_{max} + u_2 * d_i/d_{max} + u_3 * s_i/s_{max}$, where f_{max}, s_{max} and d_{max} are the maximum values for the entire photo collection, and u_i are weighting factors.

2.2 Image Appeal

Machine learning techniques are used to calculate an *appeal score* for each image. Therefore, 15 low- and mid-level image features were designed, covering the aspects *technical quality*, *composition*, *simplicity* and *colorfulness* (see Tab. 1). Fourteen test persons of different age groups were asked to provide personal photographs of particularly high and low image appeal. Thus, a training database consisting of 840 images was built. Using this database, an SVM classifier was trained. The resulting SVM model is used to estimate the image appeal. The probability estimation for the "high appeal" class membership of image i is used as the final appeal score s^i_{appeal}.

Table 1. Overview of selected appeal classification features

technical quality	**simplicity**
highlight/shadow clipping	number of salient regions
sharpness	number of distinct hues
colorfulness	**composition**
color distribution	rule of thirds
standard deviation hue-channel	centrality

2.3 Presence of People

In order to rate the presence of people in each photograph i, we combine face and skin detection. The number of faces n_i in each image as well as their relative size a_i are obtained using a face detector based on Haar features and AdaBoost training. The relative amount of skin s_i in each image is calculated using a skin detector trained on skin colors. All three aspects are combined to a final people score $s^i_{people} = v_1 * log(n_i) + v_2 * a_i + v_3 * s_i$ where v_i are weighting factors.

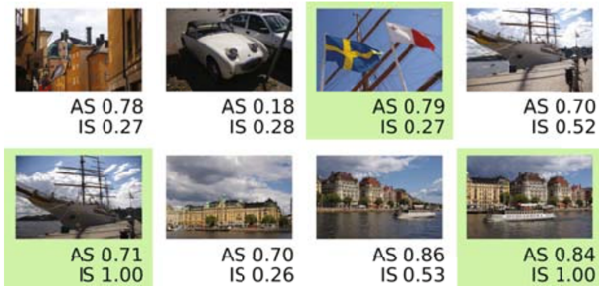

Fig. 1. Example of an event with calculated appeal score (AS) and importance score (IS) and the resulting 3 selected images (green) after duplicate removal

2.4 Ranking Framework

The three calculated scores are combined to obtain a final combined score for ranking the images: $s^i_{combined} = w_1 * s^i_{importance} + w_2 * s^i_{appeal} + w_3 * s^i_{people}$. The weighting w_i of each score can be determined by the user or specified by certain presets, according to the application scenario. Due to the calculated importance score, the top-scoring images usually contain many pictures showing the same subject. We again use SIFT keypoints for detecting similar subjects. With this information, only the highest rated images which are showing different subjects can be selected (as depicted in Fig. 1).

The representativity of the selected images can be increased by performing event detection prior to the ranking process, and selecting the highest-rated images from each event separately. To detect events, we combine and modify two existing approaches [3] [5]. The time gaps between images are first clustered using the k-means algorithm ($k = 2$). These initial events are then further partitioned by looking for outliers.

3 Evaluation

As the user's satisfaction with the rating of the images is the main measure for the quality of our method, a user study was conducted. Fourteen participants provided personal photo collections. Images were rated and selected using different presets. Each test person rated the selected images of his/her own photo collection by filling in a questionnaire, which was designed according to guidelines of qualitative research, e.g. using 5-point Likert scales and cross-check questions. The boxplot in Fig. 2 exemplarily shows one result of the user study. It represents the overall satisfaction of the test persons with the selected images. As can be clearly seen, the images selected by our method obtained much better overall ratings than randomly selected images.

The test results in general are promising. The calculated image appeal was reflected well by the user ratings. Throughout, selections generated by our method obtained considerably better ratings compared to random selections.

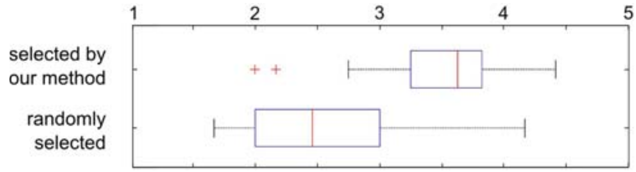

Fig. 2. Overall user-satisfaction with the selected images (1-worst, 5-best)

4 Conclusions

We briefly presented our method for rating and selecting personal digital photographs, which rates images based on various criteria, decreases redundancy and increases representativity amongst the selected images. A conducted user study delivered promising results, and verifies our concept of the automatic selection of images.

References

1. Boll, S., Sandhaus, P., Scherp, A., Thieme, S.: MetaXa - Context-and Content-Driven Metadata Enhancement for Personal Photo Books. In: Cham, T.-J., Cai, J., Dorai, C., Rajan, D., Chua, T.-S., Chia, L.-T. (eds.) MMM 2007. LNCS, vol. 4351, pp. 332–343. Springer, Heidelberg (2007)
2. Datta, R., Joshi, D., Li, J., Wang, J.: Studying Aesthetics in Photographic Images Using a Computational Approach. In: Nguyen, N.T., Kowalczyk, R., Chen, S.-M. (eds.) IWSM 2000. LNCS (LNAI), vol. 5796, pp. 152–162. Springer, Heidelberg (2009)
3. Graham, A., Garcia-Molina, H., Paepcke, A., Winograd, T.: Time as essence for photo browsing through personal digital libraries. In: Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries, pp. 326–335. ACM Press, New York (2002)
4. Kasutani, E., Yamada, A.: The MPEG-7 color layout descriptor: a compact image feature description for high-speed image/video segment retrieval. In: Proceedings of the 2001 International Conference on Image Processing, vol. 1 (2001)
5. Loui, A., Savakis, A., Co, E., Rochester, N.: Automated event clustering and quality screening of consumer pictures for digital albuming. IEEE Transactions on Multimedia 5(3), 390–402 (2003)
6. Lowe, D.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
7. Savakis, A., Etz, S., Loui, A., et al.: Evaluation of image appeal in consumer photography. In: Proceedings-SPIE the International Society for Optical Engineering, pp. 111–121 (2000)