

System for Creating Slideshows Based On People and Their Emotions

Vassilios Vonikakis and Stefan Winkler

Advanced Digital Sciences Center (ADSC), University of Illinois at Urbana-Champaign

1 Fusionopolis Way, Singapore 138632

{bbonik, stefan.winkler}@adsc.com.sg

ABSTRACT

We demonstrate a system for the automatic creation of slideshows from photo collections, based on a user-specified group of people, their emotions and other image similarity criteria such as color, timeline and scene characteristics. The main objective of the system is to form meaningful image sequences for slideshows, with minimal user input. The intuitive interface allows the user to provide a personalized definition of image similarity, resulting in different image sequences, and making the system highly user-adaptable and flexible.

Categories and Subject Descriptors

H.5.2 [Information Interfaces and Presentation]: User Interfaces

Keywords

Automatic Slideshow, Emotion Estimation, People Identification.

1. INTRODUCTION

Most pictures in personal photo collections contain people and faces. Psychophysical studies have demonstrated that classic saliency rules breakdown in social scenes, meaning that people are more interested in watching other people, their actions and their emotions, than any other object in a scene [1].

Although there is a considerable body of research in slideshow creation [2] and even some commercial products supporting automatic face annotation in personal photo-albums (Apple's iPhoto), there are no existing systems which make use of this

information, along with other similarity features, to form a chain of smooth image transitions. More specifically, none of the existing systems employing emotions [3] consider people, while systems taking into account people (iPhoto) do not consider emotions.

We demonstrate a system that takes into account people *and* their emotions for creating people-centric, meaningful image sequences. These sequences can be displayed to the user in the form of a slideshow, allowing a smoother and more personalized browsing experience.

2. SYSTEM OVERVIEW

2.1 Algorithm

Fig. 1 depicts the block diagram of the proposed system. The system has two operating modes; offline and runtime. In offline mode, the system scans the photolibrary and performs face detection. This is done in order to identify the people who are present in the photolibrary and depict them later as icons, allowing the user to tag them. Additionally, face recognition and facial emotion estimation is performed on the detected faces, along with extraction of the timestamps from the EXIF data.

In runtime mode, the user interacts with the system as described in the following section. The user specifies the retrieval conditions (the persons who will be included in the slideshow, their desired emotions and the timeline range), along with a personalized definition of image similarity, which will be used to create the image sequence. Subsequently, the photolibrary is scanned for images which satisfy the retrieval conditions.

The retrieved images are then passed to the main algorithm, which arranges them into a meaningful sequence, according to the personalized image similarity definition provided by the user. It employs four criteria to define the dissimilarity between a pair of images: emotions of people, timeline, color (hue-saturation histograms) and global scene characteristics (gist descriptor). This dissimilarity is expressed as a weighted average of the four individual criteria, with user-specified weights. The dissimilarity formula is computed for all possible pairs among the retrieved images. A graph is formed in which nodes represent the retrieved photos and edges encode the dissimilarity between them. The system uses a greedy algorithm to compute the shortest path, which passes through all the nodes, without revisiting any of them. This approach results in an image sequence with small dissimilarity values for the image transitions. Different importance weights for the similarity criteria will result in different graph edges, different shortest paths and thus, different image sequences. For a more detailed description of this algorithm, please, refer to [4].

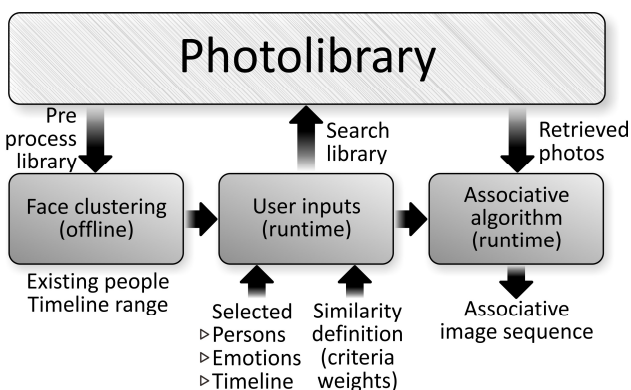


Figure 1. Block diagram of the proposed system.

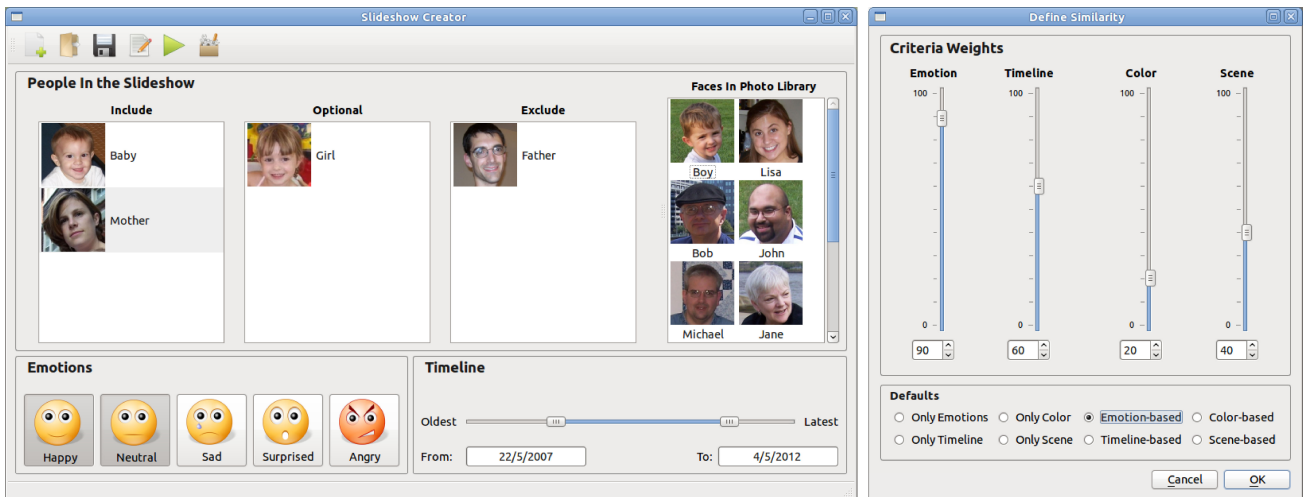


Figure 2a. The Slideshow Creator interface. b. The personalized similarity definition interface.

2.2 User Interface

The main interface of the proposed system, the “slideshow creator”, is depicted in Fig. 2a. It allows the user to create, save, edit, and display slideshows. The larger part of the window is occupied by three groupboxes “People in the Slideshow”, “Emotions” and “Timeline”. These are the three retrieval criteria used for searching the photolibrary.

The right part of the first groupbox presents a list with all the people who have been identified in the photolibrary. Every person’s face is depicted, together with his/her name, which can be edited by the user. The middle of the “People in the Slideshow” groupbox comprises three lists, where the user can drag-and-drop any of the face icons. The three lists are “Include”, “Optional” and “Exclude”, representing the operators AND, OR and NOT, respectively, allowing the user to form Boolean expressions for the image search. In the example depicted in Fig. 2a, the user defines a slideshow, in which images must include “Mother” and “Baby”, may also include “Girl”, but should exclude “Father”.

The “Emotions” groupbox contains five buttons, representing the basic emotions that the system can identify: happy, neutral, sad, surprised, and angry. The user can select any combination, requiring everybody in the retrieved images to have one of the selected emotions. This feature can also serve as a filter for discarding images, in which people are captured with unwanted facial expressions. In the example of Fig. 2a, the user is requesting that all people in the slideshow should either be happy or neutral. Any photos which do not satisfy this condition will not be used.

The “Timeline” groupbox features a double-slider control, which allows the user to define the date interval of the images that will be included in the sequence. This double slider is normalized in a way that its leftmost side represents the date of the oldest photo of the collection, and its rightmost side the date of the latest photo.

The “settings” button in the upper toolbar opens the window depicted in Fig. 2b. This interface allows the user to personalize the definition of similarity, which will be used during the creation of the image sequence. There are four sliders, which control the weights assigned to the four similarity criteria used by the

algorithm. A higher slider value indicates that the transition of images will be smoother in regards to that similarity criterion compared to the others with lower values. Setting a weight to zero effectively deactivates the contribution of this criterion. As a result, the user can control the image sequence by assigning his/her preferences regarding the weights of the four similarity criteria. Fig. 3 depicts the slideshow viewer of the system.

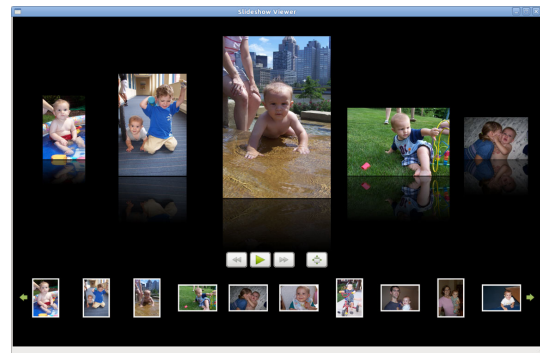


Figure 3. The Slideshow Viewer.

3. 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). We also thank Hongwei Ng for his help with developing and testing the system.

4. REFERENCES

- [1] E. Birmingham, W. F. Bischof, A. Kingstone. Saliency does not account for fixations to eyes within social scenes. *Vision Research*, 49:2992-3000, 2009.
- [2] V. Singh, J. Luo, D. Joshi, M. Das, P. Lei, P. Stubler. Dynamic media show drivable by semantics. In *Proc. ACM Multimedia*, pages 815-816. 2011.
- [3] P. Dunker, C. Dittmar, A. Begau, S. Nowak, M. Gruhne. Semantic high-level features for automated cross-modal slideshow generation. In *Proc. CBMI*, pages 144-149. 2009.
- [4] V. Vonikakis, S. Winkler. Emotion-based sequence of family photos. In *Proc. ACM Multimedia*. 2012.