Emotion-Based Sequence of Family Photos

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
This paper presents a method for the automatic creation of slideshows from family photo collections based on the emotions of a given group of people. The user specifies the desired person(s) to be included in the slideshow. A natural image sequence is formed based on people's emotions and several other, user-defined image similarity attributes, in order to form meaningful slideshow transitions. This process makes use of a new image dissimilarity function, which can integrate various attribute combinations and preferences, making the system highly user-adaptable and flexible.

Categories and Subject Descriptors
H.2.8 **[Database Management]**: Database applications – *Image databases*; H.1.2 **[Models and Principles]**: User/Machine Systems – *Human factors*; I.5.4 [**Pattern Recognition Applications**]: Computer vision.

Keywords
Emotion Estimation, Automatic Slideshow, People Identification.

1. INTRODUCTION
Humans are social animals. Behavioral studies have shown that recognition of faces and facial expressions plays a vital role in this [1]. Moreover, facial expressions have an emotional impact on viewers by conveying the same emotions to them [2]. All these, in accordance with the fact that classic saliency rules break down when people are present in a scene [3], suggest that the human factor (faces, facial expressions, posture, activity etc.) is the most important characteristic affecting the observer's impression of an image. This means that if there are people in a scene, human observers will immediately focus their attention on them and their faces, largely ignoring other image characteristics.

Consequently, assessing the human factor is a matter of primal importance for an associative image browsing system. However, this turns out to be a very difficult task; the majority of existing techniques that estimate human emotions do not take into account the human factor, but mainly focus on other global or local image features, such as various kinds of histograms or visual words [4]. Although there is a considerable body of research regarding slideshows [5], and even some commercial products about automatic face annotation in personal photo-albums (Apple's iPhoto), there are no existing systems which identify people and their emotions in photolibraries and use this information, along with other similarity features, to form a chain of natural and emotionally smooth image transitions.

2. DESCRIPTION OF THE METHOD
Initially, the system scans the whole photolibrary in order to identify all persons present in the photos. The user is presented with a list of faces of the detected people and selects those who should be included in the slideshow. After the initial retrieval of all relevant images, duplicates and aesthetically inferior images are filtered out.

Our system currently uses four different image similarity attributes to establish meaningful image relations. These attributes are the people's emotions, the date/time when the photo was captured, the colors of the image, and the gist of the scene. Emotions: The system is trained to detect happiness, sadness, anger, surprise, and neutral expression. These are modeled according to the 2-dimensional system of Valence and Arousal [6], in which every emotion is a combination of these two attributes. The emotional similarity between two photos is defined as the Euclidian distance of the moods in the Valence – Arousal plane, normalized to the interval [0,1].

Timeline: The EXIF date/time tags, which are included in the majority of digital photos, are used to define the temporal distance between photos. The timeline similarity between two images is defined as their date/time difference, normalized to the interval [0,1] based on the dates of the oldest and the most recent photos of the retrieval results.

Color: The system utilizes histograms in the Hue-Saturation-Value (HSV) color space. Hue and Saturation histograms are normalized independently and concatenated into one. Value is not used. The similarity metric used is the Bhattacharyya distance, which by its definition has a range of [0,1].

Gist describes the global characteristics of a scene, such as naturalness, openness, roughness, expansion, or ruggedness. It has been used previously for classifying scenes into indoor/outdoor or natural/artificial. The gist similarity between two photos is defined as the Euclidian distance between the gist vectors of the two images, normalized to [0,1].

The proposed system computes the dissimilarity between all the retrieved images of the photolibrary, in which the selected persons are present, as follows:

$$D_{AB} = 1 - \frac{\sum_{k=1}^{N}\left(1 - d_{AB}^{k}\right)W^{k}}{\sum_{k=1}^{N} W^{k}}, \quad D_{AB}, d_{AB}^{k} \in [0,1], \quad W^{k} \in \mathbb{R}^{+} \qquad (1)$$

where D_{AB} is the final dissimilarity between image A and image B. N is the total number of similarity attributes used by the system (N=4), d^{k}_{AB} is the k^{th} similarity attribute and W^{k} its importance weight. Since the formula is normalized, the weights can take any real positive value without affecting the range of D_{AB}. Attributes with higher weights will have greater participation in the dissimilarity metric between images, and vice versa.

Once the dissimilarity metric between all the images have been computed, a graph is formed, in which the nodes are the images, and the edges encode the dissimilarity between them. The main objective of the system is to find the shortest path that passes through all the nodes, without revisiting any of them. The shortest path condition ensures that transitions occur only between similar images. Since the level of image similarity is affected by the assigned importance weights, different weights can result in different shortest paths. The algorithm searches the immediate neighbors of the current image, which have not yet been visited, and selects the one with the smallest dissimilarity. It then moves to the newly selected image and continues the procedure until no other images are available.

3. EXPERIMENTAL RESULTS

In order to test the performance of the proposed system, the dataset of [7] was used, which comprises family photos. For demonstration, we selected 9 images of various emotional states. The results are depicted in Fig. 1. Column A depicts the result based only on emotions. It is evident that similar emotions are clustered together, while there are smooth emotional transitions: surprised is followed by happy, neutral, sad and last, angry. No significant emotional shifts are present, like changing from happy to sad. Column B depicts the results according to timeline. The EXIF date/time metadata shows that the images are correctly arranged from oldest to newest. Column C depicts the result using only color. Clustering of images with similar color is evident. Column D depicts the result based on gist alone. Although there is some grouping of indoor and outdoor scenes, the clustering is not as evident as in the previous cases, probably due to the fact that people occupy a considerable portion of the images. Finally, column E depicts the result for a weighted combination of attributes, still emphasizing emotions. In this case, the algorithm attempts to combine the four different sequences of the previous columns, with more compromises for the attributes with lower weights. Images A1 and A2 are closest in terms of emotions, but the most dissimilar ones in terms of timeline. In this case, the algorithm is forced to make compromises in both attributes. However, the compromise in timeline is greater than the one in emotions: image B9 is moved to the 4th position from last, changing 5 positions. In terms of emotions, the same image is moved only 2 places. More importantly, the images placed between them are all classified as angry, which is the closest emotion to surprised, according to the 2-dimensional model of emotions [6].

4. DISCUSSION AND CONCLUSIONS

We presented a new system for creating slideshows, specifically focused on people and their emotions, along with other similarity attributes. The system's flexible design (new similarity attributes can be easily added) and its adaptability to the user's preferences (different degrees of importance can be defined for the employed similarity attributes), make it a useful tool for associative browsing of photolibraries or slideshows. Additionally, the fact that the proposed system takes into account emotions makes it useful for filtering out expressions that are unwanted or not flattering, while keeping only the desired ones e.g. happy faces.

5. 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.

6. REFERENCES

[1] A.M. Burrows. The facial expression musculature in primates and its evolutionary significance. *Bioessays*, 30(3):212-225, 2008.

[2] B. Wild, M. Erb, M. Bartels. Are emotions contagious? Evoked emotions while viewing emotionally expressive faces: quality, quantity, time course and gender differences. *Psychiatry Research*, 102:109-124, 2001.

[3] E. Birmingham, WF. Bischof, A. Kingstone. Saliency does not account for fixations to eyes within social scenes. *Vision Research*, 49:2992-3000, 2009.

[4] W. Wang, Q. He. A Survey on Emotional Semantic Image Retrieval. In *Proc. ICIP*, pages 117-120. 2008.

[5] J.-H. Su, M.-H. Hsieh, T. Mei, V. S. Tseng. Photosense: Make sense of your photos with enriched harmonic music via emotion association. In *Proc. ICME*, pages 1-6. 2011.

[6] J. A. Russel. A circumplex model of affect. *J. Personality and Social Psychology*, 39(6):1161-1178, 1980.

[7] A. Gallagher, T. Chen. Clothing cosegmentation for recognizing people. In *Proc. CVPR*, pages 1-8. 2008.

Figure 1. Slideshow samples produced by the proposed method using different combinations of weights.