Audiovisual Slideshow: Present Your Journey by Photos

Jun-Cheng Chen[1], Wei-Ta Chu[1], Jin-Hau Kuo[1], Chung-Yi Weng[1], Ja-Ling Wu[1,2]

1. Dept. of Computer Science and Information Engineering, National Taiwan University, Taiwan

2. Graduate Institute of Networking and Multimedia, National Taiwan University, Taiwan

{pullpull, wtchu, david, chunye, wjl}@cmlab.csie.ntu.edu.tw

ABSTRACT

This demonstration presents a novel way to systematically display photos and enhance the viewing experience of photo browsing. In contrast to conventional photo slideshow, multiple photos that have similar characteristics are well arranged and displayed at the same layout. Moreover, the displaying pace is coordinated with the beat of the user-selected incidental music. To automatically generate the audiovisual slideshow, we develop a system that consists of three main components: photo analysis, music analysis, and audiovisual composition. Audiovisual content analysis and cross-media synchronization issues are addressed in this work. This novel demonstration is especially suitable to present photos taken in a journey. It vigorously presents the delights of traveling and helps us recall or experience the trip.

Categories and Subject Descriptors

H.5.1 [**Multimedia Information Systems**]: animations. H.3.1 [**Content Analysis and Indexing**]: abstracting methods, indexing methods.

General Terms

Algorithms, design, experimentation.

Keywords

Slideshow, photo clustering, music analysis, and image content analysis.

1. INTRODUCTION

Digital camera has become an indispensable commodity for each family or individual in recent years. With the advance of digital capturing/storage, people can take photos at will and have been more accustomed to record everything by photographs rather than text. Nevertheless, large amounts of photos without appropriate organization draw a potential problem in information access. Due to the difficulties of accessing or organizing such a huge amount of photos, we have urgent needs in advanced content analysis and presentation techniques.

The easiest way to access these disordered photos is through a photo slideshow. In conventional photo slideshows, photos are displayed one-by-one, according to alphabetical or temporal order. Therefore, photos taken in the same scene or having the same topic are separated into different slots, and the browsing experience is cut off. The proposed system that automatically generates music-driven photo slideshows, in which photos having similar characteristics would be displayed in the same frame, and the demonstration of photos proceeds as the pace of the incidental music. Tiling multiple photos into the same frame emphasizes the atmosphere of viewing experience, because the coherence of a frame is elaborately maintained. Collaborative photo presentations that are synchronous to music beats even improves the enjoyment of photo browsing. Because a frame is tiled by multiple photos, we called the proposed presentation *tiling slideshow*.

MM'06, October 22–28, 2006, Santa Barbara, USA.

2. System Overview

The system consists of three main stages, as shown in Figure 1. In the preprocess stage, we first perform orientation correction based on the metadata stored in EXIF (exchangeable image format) [3]. Visual quality is then estimated via the clues of motion blur [2] and underexposure/overexposure. The photos with serious quality degradation are filtered out.

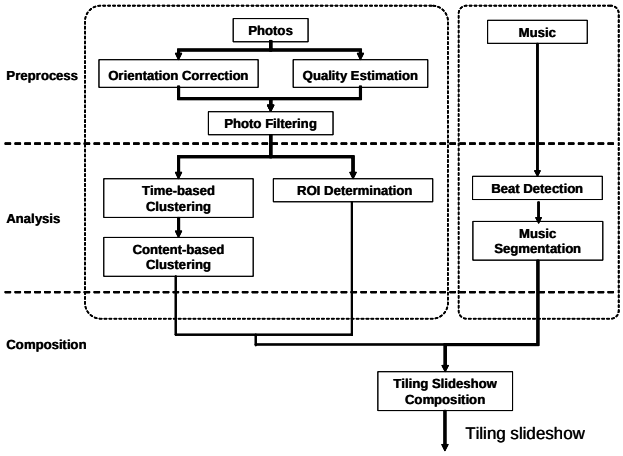

Figure 1. System flowchart of the proposed titling slideshow.

In the analysis stage, photos are first clustered according to temporal context [1]. Based on the results of time-based clustering, visual features such as dominant color and color layout are used to perform finer content-based clustering. Moreover, because the frame space is smaller than that of multiple photos, it is inevitable to shrink photos into smaller tiles to fit into a frame. Therefore, it is essential to find the most attentive region of a photo for later composition. We use face detection and contrast-based attention

model [4] to perform region-of-interest (ROI) determination. In music content analysis, we perform beat detection [5] to find the pace of music. Beat information would be the basis of the timing for photo displaying and frame switching.

Two phases, say spatial and temporal compositions, are in the final stage. In spatial composition, we perform some manipulations on a cluster of photos and tile them into a frame. The manipulations, such as scale down, cropping, and location assignment, are elaborately designed according to the content-based importance metrics of photos. In temporal composition, occurrence of parts of a frame and switching between frames are determined by the detected music beats. They are temporally synchronized to make coordinate effects.

3. Tiling Slideshow Generation

According to the guidelines of writing, a solid paragraph contains a topic sentence, which identifies the main idea of this paragraph, and several supportive sentences, which provide supportive details of the main idea. Many paragraphs are therefore concatenated to convey the whole narration of an article. Likewise, we advocate that a journey or an event can be reproduced by many *photographic paragraphs*, which are composed of at least one topic photo (with larger size) and several supportive photos (with smaller size).

- **Template Determination**

Given a photo cluster, we should select appropriate layouts for presentation. Intuitively, if the number of photos in a cluster is four, we just select the templates with four cells. To enrich the variety of displaying layout, several templates with four cells are designed, as shown in Figure 2. To choose a suitable tiling template, we define importance metrics for each photo based on attention values [4] and face information. Importance values of a cluster of photos are packed as a vector in descending order. Similarly, we define an importance vector for each template based on the occupied areas of its cells. Based on this information, the template with the importance vector that has the smallest included angle to the photo-based importance vector is the most suitable one. After this process, because both importance vectors are sorted in descending order, which photo should be put into which cell is also determined. That is, more important photos should be put into larger cells.

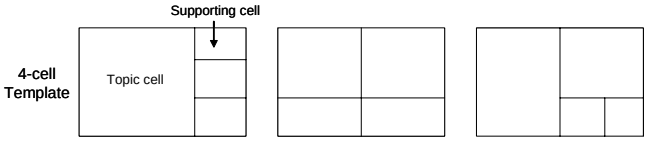

Figure 2. Examples of 4-cell templates.

- **Composition**

The final task for generating the tiling frame is to put photos into the cells of the determined template. Because the resolution of photos is significantly larger than the targeted resolution (720 × 480), it's unavoidable that we should resize or crop photos to fit into the layout. Moreover, we have to guarantee that the aspect ratio of the cropping region is the same as that of the targeted cell. Therefore, smart cropping and resizing based on the evaluation results of ROI detection are applied to shrink photos, and then we can tile photo into the corresponding layout template.

After determining the timing for displaying a photo or switching frame by the music beats, a tiling slideshow is finally generated. To facilitate more gorgeous presentation, we also include transition effects such as fade-in to display photos. Some sample results are illustrated in Figure 3.

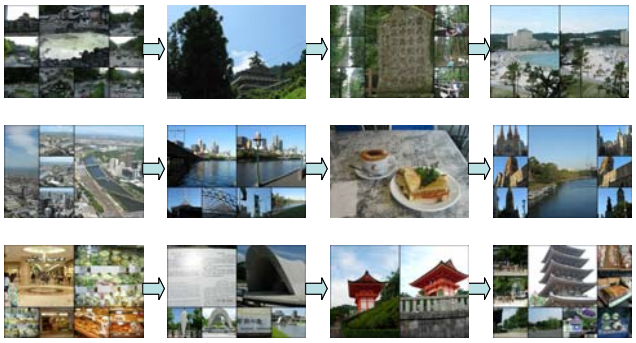

Figure 3. Some snapshots of the evaluated tiling slideshows.

4. CONCLUSION

The proposed tiling slideshow system automatically generates a composite audiovisual presentation. Photo collections are first organized according to temporal and spatial contexts, which are the units for agglomerative presentation. Photos in the same cluster are elaborately manipulated, such as ROI determination and smart cropping/resizing, to tile into the same displaying frame. Accompanying with incidental music, the tiling slideshow not only switches frames with music pace, but also displays each "tile" according to beat information. This kind of presentation brings delights in photo browsing and helps us enjoy the recall of trips.

5. REFERENCES

[1] Luo, J., Boutell, M., and Brown, C. Pictures are not taken in vacuum – an overview of exploiting context for semantic scene content understanding. *IEEE Signal Processing Magazine*, vol. 23, no. 2, 2006.

[2] Tong, H., Li, M., Zhang, H.-J., and Zhang, C. Blur detection for digital images using wavelet transform. In *Proceedings of IEEE International Conference on Multimedia & Expo*, pp. 17-20, 2004.

[3] Digital Still Camera Image File Format Standard. Japan Electronic Industry Development Association, 1998.

[4] Ma, Y.-F., Hua, X.-S., Lu, L., and Zhang, H.-J. A generic framework of user attention model and its application in video summarization. *IEEE Transactions on Multimedia*, vol. 7, no. 5, pp. 907-919, 2005.

[5] Scheirer, E.D. Tempo and beat analysis of acoustic musical signals. *Journal of Acoustical Society of America*, vol. 103, no. 1, pp. 588-601, 1998.