

Semantic Photo Album Based on MPEG-4 Compatible Application Format

Seungji Yang, Sihyoung Lee, Yong Man Ro*, *Senior member, IEEE*, and Sang-Kyun Kim

Abstract—In this paper, we propose a promising photo album system that enables augmented use of digital photos over a wide range of multimedia devices and semantic consumption of photos as well. The proposed system contains an album engine to create metadata and a photo player to consume the photos. It operates based on a standardized application format to encode a group of photos and associated metadata. Experiments showed that the proposed system achieved reasonable albuming performance.

I. INTRODUCTION

In recent decades, our daily life has been captured in digital due to the proliferation of digital camera. But, as the volume of personal photo collections is continuously increasing, many of them are often entangled without any useful keyword. This makes users difficult to browse or search for desired photos. Unfortunately, almost none of the existing home photo albums allows users to consume photos in semantically meaningful order without manually selecting, sorting, and annotating the photos. Many researchers have studied to reduce the semantic gap between human perception and low-level features [1][2]. However, reducing the semantic gap is still a challenging task. Although a few applications support semantic-like metadata, e.g., IPTC in EXIF [3], the metadata is unstructured, hard to be interpreted due to uncertainty of free text annotation. Any application-oriented metadata is dysfunctional or even useless in different applications or devices due to lack of compatibility. In addition, although the Internet provides a convenient means to share photos, large amounts of photos are still difficult to be shared at once.

This paper describes a promising photo album system that enables not only augmented use of digital home photos over a wide range of multimedia devices, but also semantic photo consumption as minimizing user's manual tasks. The photo album operates based on a standardized application format, called Photo Player MAF (PPMAF, where MAF stands for multimedia application format) [4]. Photo albuming engine to create PPMAF file and photo player to consume it have been implemented in our prototype system.

II. THEORY

In this paper, three sorts of photo semantics are proposed as semantic metadata: situation, category, and person. Situation is a semantic unit to group visually similar photos taken in close proximity of time. Category is a semantic unit to group photos in some visual concepts such as indoor and landscape. Person-identity is a semantic unit to group photos in some

specified persons. The proposed system consists of albuming engine and photo player. In order to build the albuming engine, the photo semantics are modeled with visual and syntactic features in a composite way. Syntactic features are obtained from existing photo metadata such as camera metadata, user preferences, keyword, and free text annotation. Visual features are obtained from content itself such as color and texture.

Fig. 1 presents semantic metadata extraction from visual and syntactic features in the albuming engine. The albuming engine first constructs a new feature set that combines visual and syntactic features of an input photo. Given a series of photos, situation semantic is determined by situation changes detected on measure of taken-time and visual differences [5]. Category semantic is determined by multiple visual concepts detected from small photo regions represented by visual and syntactic features [6]. Person semantic is determined based on identity features that incorporates visual features of face and associated clothes in a given situation [7]. Then, the albuming engine generates a set of possible photo semantics as semantic metadata.

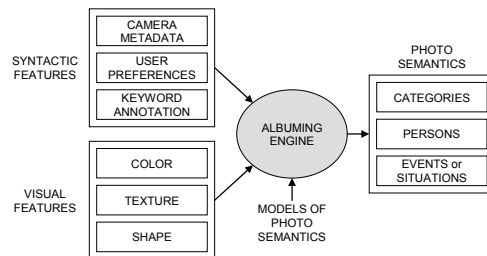


Fig. 1. Semantic metadata extraction in the albuming engine

After photo albuming, album metadata is generated to be conformed to the MPEG-7 MDS (multimedia description scheme, ISO/IEC 15938 Part 10). Thus, it allows compliant devices to support enhanced album functionality, such as intelligent browsing, content-based retrieval and automatic categorization. For hierarchical representation of individual photo and its collection, the schema of the album metadata is designed in two levels: collection- and item-level descriptions. The collection-level description is metadata related to group of photos, each of which has item-level description that contains its detailed information. The set of metadata includes semantic description such as visual features and keyword annotation, as well as creation description such as taken date/time and camera settings.

A group of photos and associated metadata are integrated into one single file that conforms to PPMAF. The PPMAF offers a standardized solution for the carriage of JPEG photos and associated metadata based on MPEG-4 file format. Fig. 2 shows the PPMAF structure. As shown in Fig.2, each file has a file type box. Photo resources are stored in media data box

*Corresponding author: Prof. Yong Man Ro is with Dept. of Electrical Eng., IVY Lab., ICU, Daejeon, South Korea; e-mail: yro@ic.u.ac.kr

and their associated metadata is stored in movie box, where one or more track box can be presented. In track box, media box contains a pointer to JPEG code stream that can be physically located in file or logically referred to additional entity out of file. Meta box contains metadata description.

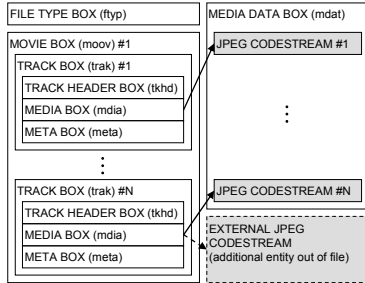


Fig. 2. The PPMF structure

Fig. 3 shows systematic process of the albuming engine and photo player. Albuming engine first initializes model DB with a few set of photos. If new photos are inputted to the albuming engine, its metadata is generated and encoded to a MAF file with the photos. The metadata can be used to refine existing model DB. Users consume MAF files on photo player. If a MAF file input to the photo player, its metadata is parsed by ISO-based atom parser. For photo browsing or retrieval, user query is matched with the metadata. A part of matched photos can be encoded into a MAF file for sharing it.

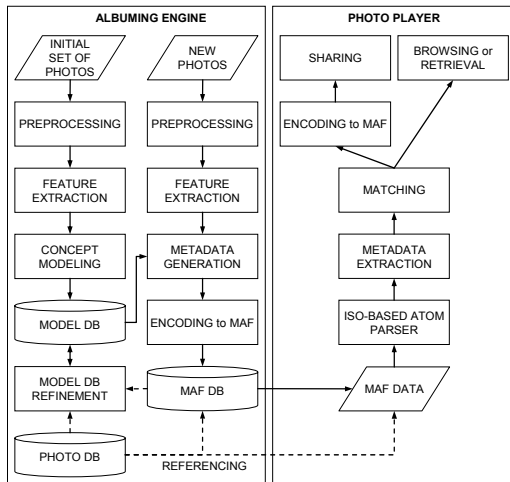


Fig. 3. Systematic process of albuming engine and photo player

III. EXPERIMENTS

Experiments were performed with about ten thousands home photos that come from official DB of the MPEG-7 visual core experiment (VCE). The test DB is accompanied with official ground truth set. Resolution of all the photos is more than 800 x 600. Table I shows albuming performance at each semantic and its processing time. The performance stands for average of recall and precision. The situation semantic of the 3828 photos is correctly determined about 80%. The overall performance of the 7 category semantics is about 75% that is the weighted-average in terms of the number of photos at each

category. For 27 persons in 1120 photos is correctly classified about 92%. Although average processing time seems slightly long, it does not matter since, once metadata has been generated, they do not have an affect on browsing time. Fig. 4 shows a layout of the proposed system. It contains detailed view of the selected photo on the left side, thumbnail view of photos in the selected semantics on the right side, and semantic tree view in-between them.

TABLE I Albuming performance and its processing time

Photo semantics	Num. of photos	Performance (%)	Processing time (sec/photo)	
Situation	4409	80.80	0.92	
Category	Architecture	66.72	2.78	
	Indoor	86.73		
	Terrain	72.80		
	Night-scene	92.11		
	Snowscape	82.55		
	Sunset	79.07		
	Waterside	70.88		
Person	27 persons	1120	92.23	1.47

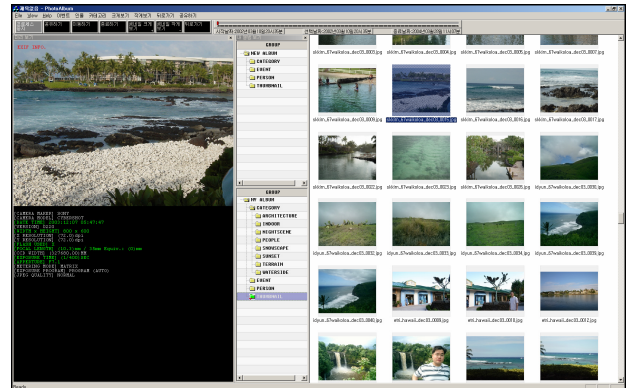


Fig. 4. Layout of the photo album system

IV. CONCLUSIONS

In this paper, we propose a promising photo album system that enables not only augmented use of digital photos over a wide range of multimedia devices, but also semantic photo browsing. Albuming engine and photo player have been implemented in the system. The proposed system has been operated with about 80% performance of semantic albuming.

V. REFERENCES

- [1] J.H. Lim et al, "Home photo content modeling for personalized event-based retrieval," *IEEE Trans. on Multi.*, vol. 10, no. 4, pp. 24-37, 2003.
- [2] A.C. Loui et al., "Automated event clustering and quality screening of consumer pictures for digital albuming," *IEEE Trans. on Multi.*, vol. 5, no. 3, pp. 390-402, 2003.
- [3] Exchangeable image file format for digital still cameras: Ver.2.2 (JEITA CP-3451), *Standard of Japan Electronics and Information Technology Industries Association*, 2002.
- [4] ISO/IEC FCD 23000-3: MPEG photo player application format, N8032, April 2006.
- [5] Y.M. Ro et al., "Automated situation change clustering of home photos for digital albuming," *SPIE Electronic Imaging on Storage & Retrieval*, vol. 5682, pp. 212-223, Jan. 2005
- [6] Y.M. Ro et al., "Home photo categorization based on photographic region templates," *LNCS*, vol. 3689, pp. 328-338, Oct. 2005
- [7] Y.M. Ro et al., "Automatic photo indexing based on person identity," *LNCS*, vol. 3768, pp. 877-888, Nov. 2005