Continuous Archiving of Personal Digital Photograph Collections with a MPEG-7 Based Dozen Dimensional Digital Content Architecture

Pei-Jeng Kuo, Terumasa Aoki and Hiroshi Yasuda
The University of Tokyo
Tokyo, Japan

Abstract

Lately, growing research efforts are emphasizing on the semantic level image retrieval based on annotated metadata or surrounding information. We are now moving towards an era that any individual can easily archive thousands of photographs, thus fast and precise consumer level image retrieval implementations have become keenly essential. This paper explains our novel approaches towards building up a lifelong personal photograph archive in an experiment database called PARIS.

Introduction

People are taking an increasing amount of photographs while they migrate from traditional cameras to digital ones. The proliferation of image devices such as digital video camera, digital camera and mobile phone camera enable individuals to capture continuous shots for life events with almost no additional cost. Consequently, consumers began to accumulate hundreds of photographs when travel to new locations or during special events. In the past, most image retrieval researches focused on stock image collections or internet image repertoire where thousands of images are indexed and compared based on their signal features such as color histogram or color layout. Lately, growing research efforts are emphasizing on the semantic level image retrieval based on annotated metadata or surrounding information. We are now moving towards an era that any individual can easily archive thousands of photographs, thus fast and precise consumer level image retrieval implementations have become keenly essential. Unfortunately, satisfactory personal image retrieval system has yet appeared.

We argue that people make more photographs while they visit some new locations or during special events. And hence spatial and temporal attributes of personal digital photographs could contain the most relevant context information. Because of the very different characteristic compare with traditional commercial stock image collections, which normally used for previous image retrieval researches, organizing, archiving, and retrieving personal photograph collections requires also different approach. To tackle this problem, we utilize the MPEG-7 multimedia description standard, which describes not only the signal low-level audio-visual content features such as color, texture, motion, audio energy as well as high-level features of semantic objects, event, content management related information and so forth. In this paper, we focus on our novel approaches towards building up a lifelong personal photograph archive.

Consumer photograph collections contain a burst structure. In other words, a group of photos may be taken for a semantic related event, but a few, if any, photos may be taken until another significant event started.[1,7] In addition, the burst structure within collections of personal photographs tends to be recursive, as small bursts exist within big bursts. This recursive structure can be represented as a cluster tree, where photographs are stored only at the leaf nodes.[7]

While previous research[1,7,9] tried to enable semi-automatic event segmentation based on the recorded time tags made possible by most recent image devices, our research emphasis the importance of an integrated approach utilizing spatial information in addition to temporal information. We have designed metadata description architecture, DDDC (Dozen Dimensional Digital Content), extended from MPEG-7 multimedia description schema for annotating personal digital contents. And we also proposed the concept of "Spatial and Temporal Based Ontology", constructed based on the special pattern of personal photograph collections as we argue that time and location are two most important attributes in terms of personal photograph retrieval.

The integrated system constructed with the above proposals is named as PARIS, Personal Archiving and Retrieving Image System. In Sec 2, we discuss the importance of geographical information in terms of personal photograph collection retrieval. Sec 3 elaborates our proposed DDDC architecture and Sec 4 explains the concept of "Spatial and Temporal Based Ontology". The structure of our prototype system is illustrated in Sec 5 and Sec 6 describes the potential application based on our system and concludes this paper.

Continuous Archiving of Geo-Coded Personal Photograph

In Ref. [14], Rodden and Wood concluded that two of the most important features of an efficient, reliable and well designed system for managing personal photographs are: automatically sorting photos in chronological order, and displaying a large number of thumbnails at once. While people are familiar with their own photographs, laborious and detailed keywords annotating are not specifically motivated for most people. Digital camera has become popular in the past few years, which means, a majority of people have only digital photograph collections accumulated within a limited time span of a few years. If we envision a continuous lifetime digital photograph archiving process, one might recalls, for example, his or her several trips to Paris a few years ago and one trip to Tokyo during Christmas season, but could not clearly memorize the exact year or dates.

While most Image retrieval researches focus on general image collections and a lot of prototype systems were developed with part of the COREL photo CDs , we would like to emphasize our research on consumer photograph collections where time and location count for important clues toward image contents. Fig. 1 shows an example of the raw data folder structure of our PARIS system. Those raw image data are continuously archived by one of our authors with a period of around three and half years and contain around 700 folders. Those folders were named with a brief line starting with the date when the groups of photographs were taken and following by a brief indication of the event, subject or location related to the content. The annotations convention of those folder names were done and followed constantly without previous knowledge of this study and hence are unbiased in terms of serving as a preliminary indication of the normal consumer naming tendency towards their digital photographs. Figure 2 shows the folder name naming example in our experiment database.

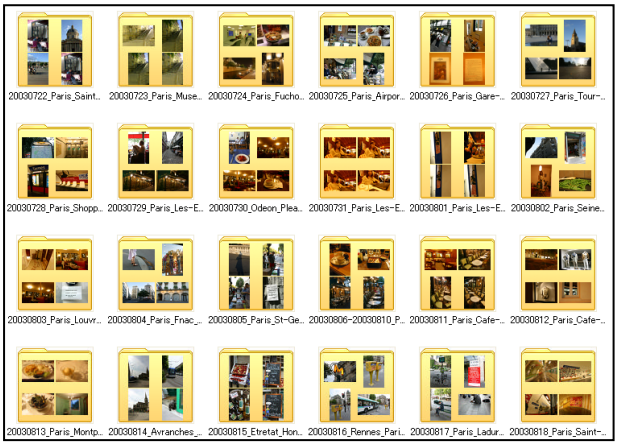

Figure 1. PARIS System Raw Data Structure

```
20020525-ASAKUSA
20020526-MEIJIGIKU
20020601-NIBOSHI
20020602-KAMAKURA
20020607-DAIBA-SHINJUKU
20020608-ASAKUSA-GINZA
20020609-YOYOGI-
20020615-OZE
20020623-LITTLE FLOWER
20020626-YAKINIKU
20020628-DOCOMO
20020629-DOCOMO NICE PHOTO
20020629-KARAOKE
20020630-NOBI-YOKOHAMA
20020701-20020704-NOMIKAI-MOVING
20020706-NYC-SOHO-GREENWICH-HOBOKEN
20020708-NYC-CENTRAL PARK-WORLD TRADE CENTER
20020710-MIT
20020711-BOSTON TOUR-CRAB-WODEN
20020713-NYC-TIME SQUARE-FIFTH AVE-CENTRAL PARK
20020715-SISIHAHA
20020716-0806-SISIHAHA-TAIWAN
20020810-HANABI
20020813-OMODESANDO HANABI
20020824-25-FUKOKA-NAGAZAKI
20020828-SAPPORO-BIEI-FURANO
20020901-INOKASHIRA KOEN
20020915-KICHIJOJI AUTHUM FESITIVAL
20020916-KAMAKURA-MANY CATS-ENOSHIMA
20021013-14-INOKASHIRA KOEN-SHIBUYA-NOKENDAI
```

Figure 2. Folder Naming Structure Example of Experiment Database

In Ref. [15], location has been argued as one of the strongest memory cues when people are recalling past events. In analyzing our prototype database raw data, we also find that most folder names contain words related to geographic information such as “Asakusa”, “Yokohama”, “Boston”, NYC”, “Fukoka”, and “Paris”, as well as temporal information such “France National Day”, “Christmas”, or “ Winter”. All of the folder names start with temporal information and 673 out of the 711 folders contain geographical related words, which is about 93% in percentage.

While our prototype system only contains images from a single person, we also conducted oral interviews with morn than 10 active digital image users to ensure an unbiased presumption towards common user naming convention. The interview results show that all of the interviewees store their digital photographs in different folders roughly named by location, event, and time period, for example, “A trip to Europe_June_2004”, “Birthday Party with Nico_20040125”, etc.

According to the above discussion, we see a strong association between the image context and its respective spatial and temporal clues for personal digital photograph collections. With the aid of recent GPS receiving devices, it is also possible to embed geographic information in GPS format onto digital image data. Although the recorded device GPS information does not guarantee the exact location of subjects appeared in the photographs as we had discussed in Ref. [4], utilize GPS information as a retrieving attribute greatly reduce the human effort compare to manual input. In Sec 4, we will further discuss the semi-automatic geo-information generation methodology conducted in PARIS system. Before that, we will briefly explain the Dozen

Dimensional Digital Content (DDDC) description architecture utilized in our annotation process in the following section.

Continuous Archiving With DDDC

Extended from the Structured-Annotation Basic Tool of MPEG-7 Multimedia Description Schemes (MDS), we propose a semantic description tool of multimedia content. The proposed content description tool enables geo-referenced multimedia data annotation with twelve main attributes regarding its semantic representation. The twelve attributes include answers of who, what, when, where, why and how (5W1H) the digital content was produced as well as the respective direction, distance and duration (3D) information. We define digital multimedia contents including image, video and music embedded with the proposed semantic attributes as Dozen Dimensional Digital Content (DDDC). Due to limited space, detailed explanation and more example codes can be found in Ref. [4].

Continuous Archiving with Spatial and Temporal Ontology

The above DDDC architecture provides a structured methodology to annotate most significant, if not explicit, semantic answers of personal digital photograph collection contexts. However, some of the DDDC annotations such as the free text part of who, where and what attributes still require manual inputs. In Ref. [6], several difficulties have been pointed out in terms of the annotation process. First, different annotator might use different terms to annotate the same concept. Second, the users who do not have specific domain knowledge might not be able to input the right keywords or natural language query for semantic image retrieval. And third, the manual annotation of a large amount of personal digital photograph collections, if not impossible, is a laborious task.

We propose to build location specific "Domain Ontology" for popular tourist stops such as the city of Paris, Tokyo and New York based on their respective spatial and temporal attributes. Illustrations in Fig. 3 and Fig. 4 present example of spatial and temporal based hierarchical ontology for the city of Paris.

In building Spatial Ontology, we firstly separate Paris into several popular tourist districts such as "The Latin Quarter", "The Eiffel Tower Quarter", "Champs-Elysées" and "St-Germain des Prés". Under each district, we again separate it into sub-districts or point of interests such as "Café de Flore", "The Eiffel Tower" and "Café les Deux Magots". Each node of the sub layer inherits the properties of their upper layers; therefore, when we annotate a photograph with "Café de Flore" metadata, upper layer properties of "St-Germain des Prés" and "Paris", "France" would also be included.

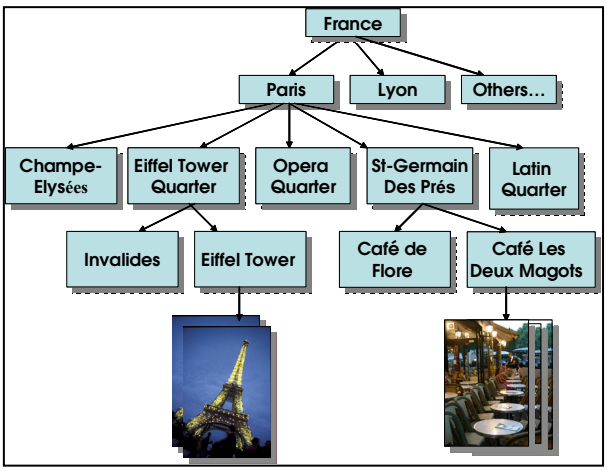

Figure 3. Concept of Proposed Spatial Ontology

The construction of Temporal Ontology requires more domain knowledge of the specific location. For example, the seasonal events periodically happen in the area, or special event occurs on specific date. As suggested in Ref. [12], there is no single correct class hierarchy for any given domain. And the ontology should not contain all the possible information about the domain but only specific enough for what you need in the application. We suggest building up the location specific "Spatial and Temporal Ontology" according to the photographer's personal interest and experience. In addition, we can also construct that with the aid of third party databases such as travel information portals or existing geographic metadata initiatives. In Fig. 4, we demonstrate event tags come from our Temporal Ontology such as "New Years Party", "Military March", "Fireworks" and "Count Down", which are associated with different image groups that were taken at the location of "Champs-Elysées" and "The Eiffel Tower" at special time such as "New Year's Eve" or "National Day".

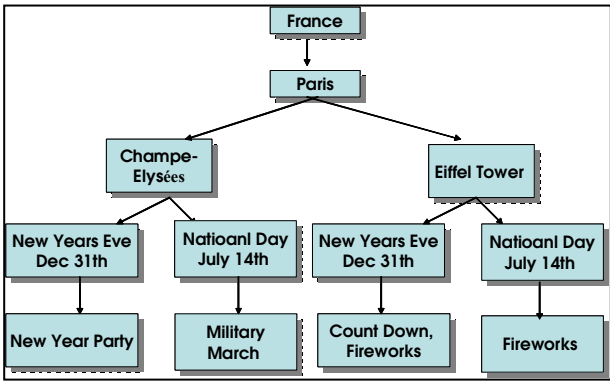

Figure 4. Concept of Proposed Spatial and Temporal Ontology

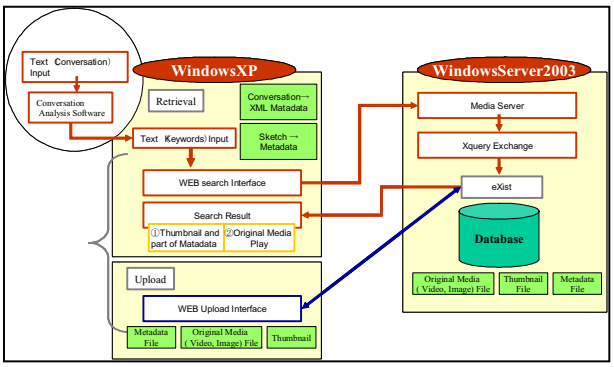

Figure 5. Testbed Structure used for constructing PARIS System

Continuous Archiving Experiment Database

We are building up our personal photograph collection database based on the DDDC architecture described above, which annotates multimedia data with twelve main attributes regarding to its semantic representation. And we utilize third-party software in generating our proposed "Spatial and Temporal Based Ontology" for popular tourist cities such as Tokyo, New York and Paris. Different from previous researches, our "Spatial and Temporal Based Ontology" are designed based on the special pattern of personal photograph collections while we argue that time and location are two most important attributes in terms of personal photograph retrieval. The PARIS system are build in conjunction with a main project called AVR (Advanced Video Retrieval), funded by Japanese Government. Figure 5 illustrates the structure of the AVR testbed. At current implementation, web interface is available for query sentence inputs. After the input, query sentence will be analyzed by conversation analysis software called "Sen" and be converted into xQuery inputs. Unfortunately, the current input analysis is only available for Japanese query inputs; however, we will work on providing a more multilingual compatible solution in the future. The spatial and temporal associated MPEG-7 annotations files are stored in a XML database called eXist and query results and thumbnails will be returned to our web-based interface. While AVR project explores novel methodologies for multimedia content retrieval including video, audio and image; PARIS system targets on personal photograph collection retrieval and is built under the whole AVR testbed. Raw image files, thumbnails and XML based DDDC MPEG-7 annotations are stored in the database. The spatial attributes are firstly generated semi-automatically by third-party GPS program and parse to our DDDC annotation schema. We are currently investigating on collaborating with GPS receiving device and generate "true" geo-referenced photograph data. However, at this point, we utilize image data that were taken before the availability of GPS receiving device and semi-automatically annotate the location information with the aid of third-party software. Currently, the PARIS system contains around 80,000 photographs with a time span of around three years, and the number of images keeps increasing with a constant rate. Around 20,000 of the photographs were scanned from traditional films and hence to not contain exact temporal data. The other around 60,000 photographs were taken from contemporary digital cameras and all contain detailed temporal information. At current stage, around 10,000 of them are geo-referenced and hence related location annotations are generated based on our "Spatial and Temporal Ontology".

Conclusion

PARIS, Personal Archiving and Retrieving Image System, is designed and implemented specially towards the emerging trend of continuous capture and storage for personal experience. While digital recording devices become pervasively available, consumers can easily accumulate thousands of digital photographs within a very short time period. Thus, image archiving, retrieving and managing methodology specially tailored for consumers are keenly needed. We have designed a MPEG-7 based DDDC multimedia description schema which annotates multimedia data with twelve main attributes regarding to its semantic representation. As we argue that location and time contains the most curial context information for personal photographs, we proposed a machine-understandable "Spatial and Temporal Based Ontology" representation for the above DDDC semantics description to enable semi-automatic annotation process particularly for enabling spatial and temporal based image retrieval for geo-referenced personal photographs.

Different from previous researches, the DDDC architecture complements the attributes specifically important for personal photograph collections. Besides, our "Spatial and Temporal Based Ontology" is a new concept designed based on the special pattern of personal photograph collections. The PARIS system is implemented is in conjunction with a three-year long project called AVR (Advanced Video Retrieval) funded by Japan Government and currently a text-based web retrieving interface has be realized. We anticipate a more sophisticated demonstration system to be presented by the end of year 2005.

As personal digital photograph collections have specific characteristics and are particularly spatial and temporal associated, we envision various novel browsing possibilities at semantic level can be developed based on the proposed DDDC architecture with the aid of our domain specific "Spatial and Temporal Based Ontology". While this paper explains the underly structure of PARIS system constructed based on our proposed methodology, we are also conducting experiments on our designed spatial and temporal image retrieval. We expect to publish the evaluation results on future publications.

References

1. J. C. Platt, et al., "PhotoTOC", Microfort TR, Feb 02.
2. ISO/IEC 15938-5:2001, "Multimedia Content Description Interface – Part 5", version 1.
3. P. J Kuo, T. Aoki and H Yasuda, "Semi-Automatic MPEG-7 Metadata Generation of Mobile Images with Spatial and

Temporal Information in Content-Based Image Retrieval", Pro. SoftCOM 2003, Oct 03.
4. P. J Kuo, T. Aoki and H Yasuda, "MPEG-7 Based Dozen Dimensional Digital Content Architecture for Semantic Image Retrieval Services", Proc. EEE-04, Mar 04.
5. V. W. Soo et al., "Automated Semantic Annotation and Retrieval Based on Sharable Ontology and Case-based Learning Techniques", Proc , JCDL'2003, May 03.
6. T. Schreiber et al., "Ontology-Based Photo Annotation", IEEE Intelligent Systems, May 01.
7. Graham et al., "Time as Essence for Photo Browsing Through Personal Digital Libraries", Proc.,JCDL'02, Jul 2002.
8. E. Hyvönen, et al., "Ontology-Based Image Retrieval", Proc. of XML Finland 02.
9. Stent, et al., "Using Event Segmentation to Improve Indexing of Consumer Photographs", Proc. ACM SIGIR'01.
10. Jaimes et al., "Semi-Automatic, Data Driven Construction of Multimedia Ontologies", Proc.ICME 03.
11. N. Noy et al., "Ontology Development 101", SMI TR, 01.
12. J. Hunter, "Enhancing the Semantic Interoperability of Multimedia Through a Core Ontology", IEEE Transactions on Circuits and Systems for Video Technology,Jan 03.
13. P. J Kuo, T. Aoki and H Yasuda, Building Personal Digital Photograph Libraries: An Approach with Ontology-Based MPEG-7 Dozen Dimensional Digital Content. In Proc. IEEE Computer Graphics International-2004 ,Jun 04.

Biography

Pei-Jeng Kuo received her B.S. in Physics from National Taiwan University, Taiwan and M.E. from Massachusetts Institute of Technology, Cambridge, USA. She has interned with the Ericsson Eurolab Multimedia Application Research Department, French Telecom Creative Research Lab and NTT DoCoMo Wireless Laboratory. She is currently a Ph.D candidate in the University of Tokyo and her research topics include the indexing, archiving, delivering, and retrieving of multimedia contents with MPEG-7 technology.
{peggykuo,aoki,yasuda}@mpeg.rcast.u-tokyo.ac.jp