An Experiment on Personal Archiving and Retrieving Image System (PARIS)

PEI-JENG KUO, TERUMASA AOKI and HIROSHI YASUDA

The University of Tokyo, Research Center of Advanced Science and Technology
4-6-1 Komaba, Meguro, Tokyo, Japan

Abstract— PARIS (Personal Archiving and Retrieving Image System) is an experiment personal photograph library, which includes more than 80,000 of consumer photographs accumulated within a duration of approximately five years, metadata based on our proposed MPEG-7 annotation architecture, Dozen Dimensional Digital Content (DDDC), and a relational database structure. The DDDC architecture is specially designed for facilitating the managing, browsing and retrieving of personal digital photograph collections. In annotating process, we also utilize a proposed Spatial and Temporal Ontology (STO) designed based on the general characteristic of personal photograph collections. This paper explains PRAIS system.

Keywords— Ontology, Databases and Information Retrieval, MPEG-7, Spatial-Temporal, Digital Library Designs l, metadata, Semantic Web, semi-automatic annotation

I. INTRODUCTION

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. 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. Personal photograph collections have very different characteristics compare with traditional commercial stock image collections, which were normally used for image retrieval researches. Hence, organizing, archiving, and retrieving personal photograph collections also require different approaches. PARIS (Personal Archiving and Retrieving Image System) is an experiment personal photograph library, which includes more than 80,000 of consumer photographs accumulated within a duration of approximately five years, metadata based on our proposed MPEG-7 annotation architecture, Dozen Dimensional Digital Content (DDDC), and a relational database structure. The DDDC architecture is specially designed for facilitating the managing, browsing and retrieving of personal digital photograph collections. In annotating process, we also utilize a

proposed Spatial and Temporal Ontology (STO) designed based on the general characteristic of personal photograph collections. This paper explains PRAIS system.

II. PERSONAL IMAGE ARCHIVING

Personal image archiving has a very different characteristic compare with traditional commercial stock image collections, or web image collections. While the above image collections are normally used for previous image retrieval researches, we think personal image archiving requires a much different strategy in terms of management and retrieving.

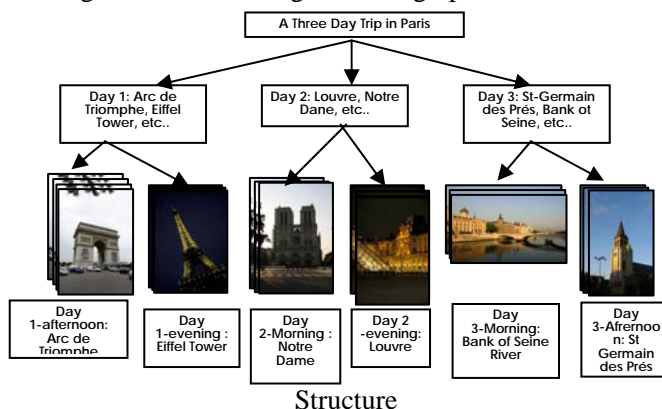Figure 1 Personal Digital Photograph Clusters Tree

Structure

Figure 1 shows a time and location related burst tree structure in a demonstrative "Three Day Trip in Paris" personal photograph collection. Generally speaking, people tend to take personal photographs in bursts. Which means, a group of photos may be taken for a semantic related event, but a few, if any, photos may betaken until another significant event started. [1, 7]

In addition, most people make more photographs while they visit some new locations or during special events. And as a result, the spatial and temporal attributes of personal digital photographs could contain some very relevant context information.

The burst structure within collections of personal photographs tends to be recursive, while small bursts exist within big bursts as shown in Figure 1. And this recursive structure can be represented as a cluster tree, where photographs are stored only at the leaf nodes. [7]

In previous researches [1, 7, 9], semi-automatic event segmentation based on the recorded time tags made possible by most recent image devices were enabled. While our research emphasizes the importance of an integrated approach utilizing spatial information in addition to the ready-to-retrieve temporal information, we have designed metadata description architecture, DDDC (Dozen Dimensional Digital Content), extended from MPEG-7 multimedia description schema for annotating personal digital assets.

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

III. WHY LOCATION MATTERS

In [14], Rodden and Wood concluded that two of the most important features of an efficient, reliable and well-designed system for managing personal photo-graphs 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. Figure 2 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
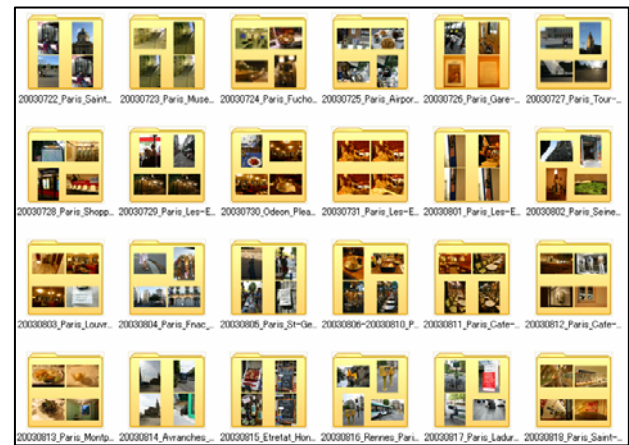

Figure 2 Original Folder Structure of PARIS System

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.

In [15], location has been argued as one of the strongest memory clues 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 [4], utilize GPS information as a retrieving attribute greatly reduce the human effort compare to manual input.

Figure 3 Sample Image with free text annotation "2001_07_14_People and Eiffel Tower in Paris on the National Day of France"

IV.   OUR PROPOSALS

PARIS (Personal Archiving and Retrieving Image System) provides an image annotation methodology based on our proposed MPEG-7 annotation architecture, Dozen Dimensional Digital Content (DDDC). In annotating process, we also utilize a proposed Spatial and Temporal Ontology (STO) designed based on the general characteristic of personal photograph collections.

Extended from the StructuredAnnotation Basic Tool of MPEG-7 Multimedia Description Schemes (MDS), we propose a semantic description tool of multimedia content. The proposed content description tool annotates multimedia data with twelve main attributes regarding its semantic representation. And 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 also define digital multimedia contents including image, video and music embedded with the proposed semantic attributes as Dozen Dimensional Digital Content (DDDC). Figure 3 is a sample image with free text annotation " 2001_07_14_People and Eiffel Tower in Paris on the National Day of France". And Figure 4 shows part of the MPEG-7 based DDDC annotation of Figure 3. Due to limited space, detailed explanation and example codes can be found in [4].

The above DDDC architecture provide a structured methodology to annotate most significant, if not explicit, semantic answers of personal digital photograph collection contexts. However, as shown on Figure 4, some of the DDDC annotations such as the free text part of who, where and what attributes still require manual inputs.

Figure 4 MPEG-7 DDDC Metadata Example

To minimize the frustration of tedious manual inputs, 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.

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.

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 [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 geo-graphic metadata initiatives. A more detailed

explanation of our proposed Spatial and Temporal based Ontology can be found in [13].

V. CURRENT STATUS AND FUTURE WORKS

PARIS system is build in conjunction with a main project called AVR (Advanced Video Retrieval), funded by Japanese Government. The AVR project explores novel methodologies for multimedia content retrieval including video, audio and image. Within AVR project, the PARIS system targets on personal photograph collection retrieval in particular.   In PARIS system, the spatial attributes for DDDC annotation are firstly generated semi-automatically by third-party GPS program and parse to our DDDC annotation schema. And temporal attributes directly come from digital camera records. We are currently investigating on collaborating with GPS receiving device and generate "real-time" geo-referenced photograph data. Currently, PARIS system contains around 80,000 photographs with a time span of around five years. At current stage, around 10,000 of them are geo-referenced and hence related spatial and temporal based annotations are generated based on our "Spatial and Temporal Ontology".

Figure 5 illustrates the structure of AVR testbed. At current implementation, web interface is available for query sentence inputs. After the input, query sentence will be analysized 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. Figure 6 shows the retrieval result interface. 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.

VI. 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 anticipate a
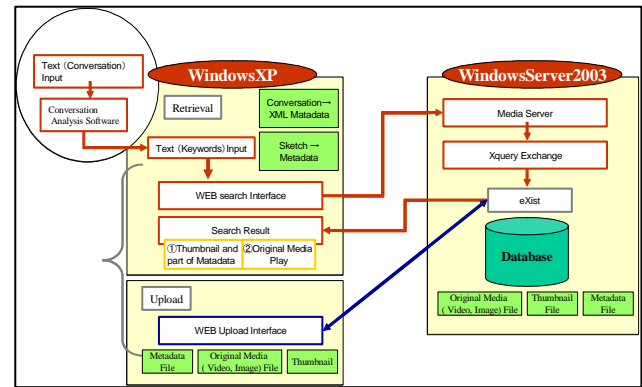

Figure 5 AVR Testbed Structure for PARIS System

Figure 6 Experiment Photograph Library Retrieval Result

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".

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 Im-age 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]   A. T. Schreiber et al., "Ontology-Based Photo Annota-tion", IEEE Intelligent Systems, May 01.

[7]   A. 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]   A. Stent, et al., "Using Event Segmentation to Improve Indexing of Consumer Photographs", Proc. ACM SIGIR'01.

[10] A. 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 Transac-tions 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", Proc. Computer Graphics International-2004, Jun 04.

[14] K. Rodden and K. Wood, "How do People Manage Their Digital Photographs?" ACM Conference on Human Factors in Computing Systems (ACM CHI 2003), Apr 2003.

[15] W. Wagenar, "My memory: A study of autobiographical memory over six year", Cognitive Psychology, Ch 18, 1986.