Browsing Personal Digital Photograph Collections with Spatial and Temporal Based Ontology and MPEG-7 Dozen Dimensional Digital Content Architecture

Pei-Jeng KUO Terumasa AOKI Hiroshi YASUDA

Yasuda-Aoki Laboratory, The University of Tokyo

Research Center for Advanced Science and Technology

4-6-1 Komaba, Meguroku, Tokyo, 153-8904 JAPAN

Abstract: **The current trend of image retrieval is to incorporate the image visual features used in Content Based Image Retrieval (CBIR) and semantics annotations used in Metadata Based Image Retrieval to enhance retrieval performance. Because of the pervasive of consumer imaging devices, building personal digital photograph libraries became an increasingly interested domain. Personal digital photograph collections have specific characteristics compare to general purpose image databases. Hence, annotation architecture specially designed for that plays an important role in building an interoperatable data repository for future indexing, browsing and retrieving purposes. We propose a MPEG-7 based multimedia content description architecture, Dozen Dimensional Digital Content (DDDC), which annotates multimedia data with twelve main attributes regarding its semantic representation. In addition, we also proposed a machine-understandable "Spatial and Temporal Based Ontology" representation for the above DDDC semantics description to enable semi-automatic annotation process.**

Keywords:

Ontology, Digital Image Database, MPEG-7, Spatial-Temporal, image retrieval, metadata, Semantic Web, semi-automatic annotation

1. INTRODUCTION

While an increasing amount of people are building their online photo albums with the aid of off the shelf digital album tools as well as web album hosting sites, an effective and semantic way of retrieving context relevant images from the large repository of personal digital archives has yet appeared.

Two approaches have been studied in the research community:

1. Content-Based Image Retrieval (CBIR): CBIR research has been on-going for sometime. [14,15,17,18] Most of the Content-based approaches compare images based on their visual features such as color histogram, color layout, texture or shape. However, the retrieval precision has yet to be satisfactory.

2. Metadata-Based Image Retrieval: In Metadata-Based Image Retrieval, external metadata annotations such as keywords or free text descriptions are used when dealing with conceptually higher levels of content. [24]

In this paper, we focus on metadata-based image retrieval with an emphasis on management of personal photograph collections including novel indexing, clustering and retrieving with our proposed architecture.

Typically, individuals can publish their digital photographs online with a few key words annotated. Some users might choose some of their best shots among their digital repository and annotate those photos with semantic descriptions regarding to the context of those images.

There are still some problems which hamper the development of "semantic" level image retrieval given the availability of carefully annotated external metadata [21, 22, 24, 26, 27]:

1. There is lack of common annotation architecture for personal digital image library.

2. Annotations require domain knowledge.

We try to tackle the above two problems with the following steps:

1. Construct common annotation architecture for building personal digital photograph libraries –We proposed The "Dozen Dimensional Digital Content (DDDC)" architecture extended from MPEG-7 Multimedia Description Scheme.

2. Construct a machine-understandable "Spatial and Temporal Based Ontology" representation for the above DDDC semantic description to enable semi-automatic annotation process.

In Section 2, we will describe the special characteristics of personal digital photograph collections. Section 3 summarizes the general concept of our proposed DDDC scheme. Section 4 provides an explanation of our approach in building up the Spatial and Temporal Based Ontology with an example of personal tourist photograph library. Section 5 concludes this paper.

2. PERSONAL DIGITAL PHOTOGRAPH LIBRARIES

2.1 Burst Structure

Consumer photograph collections are different from most general purpose image databases. 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. [8,23] Those personal photograph bursts often occur when the users encounter new events or move to new locations. The burst structure within collections of personal photographs tends to be recursive. This means, 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. [23]

Figure 1 shows a time and location related burst tree structure in a demonstrative "Three Day Trip in Paris" personal photograph collection. The user visited Arc de Triomphe and Eiffel Tower on Day-1, Notre Dame and the Louvre Museum on Day-2, and the bank of Seine River and St-Germain des Prés on Day-3. Bursts of photos related to each location occur in a temporal sequence and hence can be semi-automatically separated given there are noticeable time gaps between two bursts of images. Under the top cluster of the whole collection are three sub-clusters separated by dates.

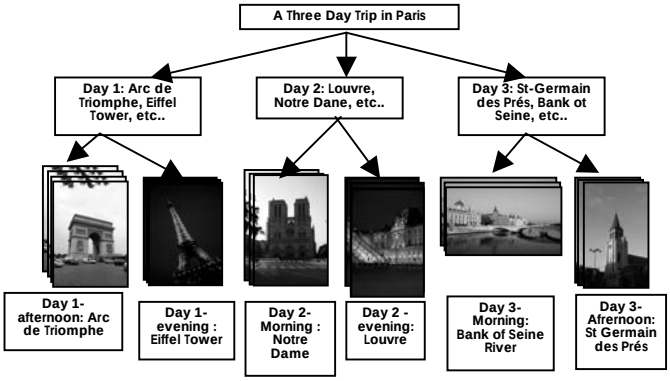

Figure 1 Personal Digital Photograph Clusters Tree Structure

Within each day, sub-clusters can be found according to different time sections of a day. Within the specific time section of a day, we can also define more sub-clusters if needed.

2.2 Proposed MPEG-7 Based Spatial and Temporal Retrieval

In addition to "time", the temporal element, we argue that "location", the spatial element, plays an equally important role as hint to the semantic context of personal photograph collections. One of the most interested topics for personal photographs is tourist photographs. People tend to take a great number of photographs especially during their trips to a new place. Part of the location hints, though poorly addressed, can normally be found from the freely captions users annotated as our example in the previous section. Alternatively, as global positioning system (GPS) is being integrated with digital cameras, the location information can be extracted from the raw image files and serve as a clustering criterion in the future.

To enable a spatial and temporal based personal digital photograph retrieval system, we adopted the MPEG-7 standard as the basis of our proposed annotation architecture, Dozen Dimensional Digital Content (DDDC), which will be explained in the next section.

3. DOZEN DIMENSIONAL DIGITAL CONTENT (DDDC) ARCHITECTURE

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. 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, only brief explanation on the twelve attributes is given as following, and detailed explanation and example codes can be found in [20]:

Who:

The *who* attribute describes animate objects or beings such as "people" and "animals" or "person groups" using Person Description Scheme (Person DS) or free text.

What

The w*hat* attribute describes inanimate object using either free text or a term from the classification scheme.

When

The *when* attribute describes the time point while the specific scene within the digital content happened.

Where: Longitude

Where:Longitude attribute describes the spatial information of the digital content. Here we adopt the GeographicPoint Semantics specified in [12] and hence three attributes longitude, latitude and altitude are required to annotate the location where a specific digital content was taken.

Where: Latitude

The *where: latitude* attribute describes the latitude in degrees. Negative value represents southern latitude.

Where: Altitude

The *where: Altitude* attribute describes the altitude in meters. The reference altitude, indicated by zero, of the measurement is set to the sea level as default.

Why

The *why* attribute describes the purpose that specific digital content such as audio, video or image was recorded.

How

The *how* attribute describes the device condition information while the specific digital content such as audio, video or image was recorded.

Direction: Theta (θ)

The *direction: Theta (θ)* and d*irection: Phi (Φ)* annotations describe relative direction between the recording device and the recorded object.

Direction: Phi (Φ)

As explained above, the second polar angle d*irection: Phi (Φ)* attribute is required for the *direction* annotation.

Distance

The distance between the photographer and the object is determined by the attribute of *distance* d (m) and it can be calculated based on the focal length information provided by most advanced digital recording devices.

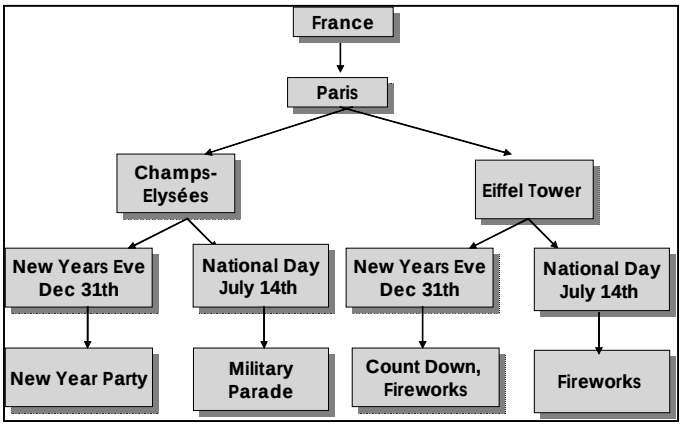

Figure 2 Concept of Proposed Spatial and Temporal Ontology

Duration

For multimedia content, especially audio and video, another attribute, *duration*, is also important when describing its semantic presentation.

4. SPATIAL AND TEMPORAL BASED ONTOLOGY

4.1 Ontology-Based Photograph Annotation

Given the above DDDC architecture, we provide 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 annotations. In [21], several difficulties have been pointed out in terms of the annotation process. First, different annotator might use a 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.

In [22], the idea of Ontology-Based Photo Annotation was described. An ontology is a formal, explicit specification of a domain. Typically, an ontology consists of *concepts*, concept *properties*, and *relationships* between concepts. [26] Ontology concepts are represented by terms, which can help the user in formulating the information needed, the query, and the answers [24]. While images in a content repository are annotated according to specific domain ontology, the same conceptualization can also offer to the users to facilitate focused image retrieval using the right terminology.

4.2 Our Proposed Spatial and Temporal Ontology

Figure 2 illustrates an example of our proposed Spatial and Temporal Ontology built 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.

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 [27], 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 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 Figure 2, we demonstrate event tags come from our Temporal Ontology such as "New Years Party", "Military Parade", "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".

5. CONCLUSION

We have proposed the DDDC architecture which annotates multimedia data with twelve main attributes regarding its semantic representation. In addition, we also proposed a machine-understandable "Spatial and Temporal Based Ontology" representation for the above DDDC semantics description to enable semi-automatic annotation process. As personal digital photograph libraries 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 proposal described in this paper.

REFERENCES

[1] N. Day, "Search and Browsing", *Introduction to MPEG-7* , Ch20, John Wiley & Sons, Ltd, 01.

[2] N. Day, et al., "Mobile Applications", *Introduction to MPEG-7*, Ch21, John Wiley & Sons, Ltd, 01.

[3]ISOIEC 15938-1, "Multimedia Content Description Interface – Part 1: Systems", 2001.

[4] http://elib.cs.berkeley.edu/.

[5] http://www.ctr.columbia.edu/dvmm/

[6] A. B. Benitez, et al., "Semantics of Multimedia in MPEG-7", *Proc. of ICIP-2002)*, 02.

[7] K. Rodden and K. Wood, "How do People Manage Their Digital Photographs?", *ACM CHI 2003*, Apr 03.

[8] J. C. Platt, et al., "PhotoTOC", *Microfort TR*, Feb 02.

[9] A. B. Benitez, et al., "Perceptual Knowledge Construction from Annotated Image Collections", *Proc. of ICME-2002*.

[10] ISO/IEC JTC1/ SC29/WG11 N4980, "MPEG-7 Overview", Jul 2001.

[11] J. Z. Wang, et al., "SIMPLIcity", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 23, no. 9, 01.

[12] ISO/IEC 15938-5:2001, "Multimedia Content Description Interface – Part 5", version 1.

[13] P. Salembier and J. Smith, "Overview of Multimedia Description Schemes and Schema Tools", *Introduction to MPEG-7*, Ch6, John Wiley & Sons, Ltd, 01.

[14] A. Pentland, et al., "Photobook: Tools for Content-Based Manipulation of Image Databases", *SPIE Proc.*, Feb 94.

[15] http://wwwqbic.almaden.ibm.com/

[16] ISO/IEC 1/SC 29/WG 11/N3964, "Multimedia Description Schemes XM", version 7.0, Mar 01.

[17] C. Carson, et al., "Blobworld", *Proc. VIS*, Jun 99.

[18] J. R. Smith and S.-F. Chang, "VisualSEEk", *Proc., ACM Multimedia '96* , Nov 96.

[19] 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.

[20] P. J Kuo, T. Aoki and H Yasuda, "MPEG-7 Based Dozen Dimensional Digial Content Architecture for Semantic Image Retrieval Services", *Proc. EEE-04* ,Mar 04.

[21] V. W. Soo et al., "Automated Semantic Annotation and Retrieval Based on Sharable Ontology and Case-based Learning Techniques", *Proc , JCDL'2003, May 03.*

[22] A. T. Schreiber et al., "Ontology-Based Photo Annotation", *IEEE Intelligent Systems*, May 01.

[23] A. Graham et al., "Time as Essence for Photo Browsing Through Personal Digital Libraries",*Proc.,JCDL'02*, Jul 2002.

[24] E. Hyvönen, et al., "Ontology-Based Image Retrieval", *Proc. of XML Finland 02*.

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

[26] A. Jaimes et al., "Semi-Automatic, Data Driven Construction of Multimedia Ontologies", *Proc.ICME 03*.

[27] N. Noy et al., "Ontology Development 101", *SMI TR*, 01.

[28] J. Hunter, "Enhancing the Semantic Interoperability of Multimedia Through a Core Ontology", *IEEE Transactions on Circuits and Systems for Video Technology*,Jan 03.