MART-MAF: Media file format for AR tour guide service

2009 11th IEEE International Symposium on Multimedia

MART-MAF: Media File Format for AR Tour Guide Service

Ig-Jae Kim[1,2], Injun Song[1], Jane Hwang[1], Sang Chul Ahn[1], Hyoung-Gon Kim[1] and Heedong Ko[1]
Imaging Media Research Center, Korea Institute of Science and Technology[1]
MIT Media Laboratory[2]

ABSTRACT

We are currently developing a new tour guide service on mobile phone using augmented reality. Among the components for the service, in-situ annotation is one of the most important components for capturing of lots of personal experiences in the form of digital multimedia during his/her trip. Our in-situ annotation system provides several pieces of functionality that include recording MART (Mobile Augmented Reality based Tour) media format, transmitting the recorded media and retrieving the related experiences of others. MART media data should be easy for users to store, interchange and manage. Additionally, it can integrate various types of media, such as video, audio, GPS data, motion, 2D/3D graphics and annotated texts. For this, we define a new format for MART media, called MART-MAF, which can compose MPEG standards and non-MPEG standards for various kinds of media.

KEYWORDS: AR tour service, media format, in-situ annotation, MAF, MART.

1 INTRODUCTION

ISO/IEC 23000 is a recent addition to a sequence of standards that have been developed by the Moving Picture Experts Group. This new standard is developed by selecting existing technologies from all published MPEG standards and combining them into so-called "Multimedia Application Formats" or MAFs [1,2].

These MAFs are defined based on ISO/IEC 14496-12 that is ISO Base Media File Format. Some media files widely used in mobile phone, such as MPEG-4 media and 3GPP media, are already following the ISO Based Media File Format. ISO/IEC 23000 standardization has already started to work on several early proposals for MAFs, which are at various levels of maturity in the MPEG standardization process. Figure 1 shows one of MAF examples (ISO/IEC 23000-8: Portable Video). This format contains the content the users generated on their own for portable video player. It contains several kinds of standard media like MPEG4, MPEG7, JPEG and so on. Like this, for AR tour service, we need a new MAF for several kinds of multimedia data including 3D synthetic objects. MART-MAF we propose also should meet the compatibility to each existence standard formats.

Figure 1. An example of MAF : ISO/IEC 23000-8

While travelling, users usually want to archive something related his/her experiences in the form of digital multimedia. Unlike the previous MPEG standard formats that deal with only one or two media types, we need a higher level wrapper including various kinds of media type to capture one's experiences realistically.

A new media format with AR to be applied tour service should be more flexible than the previous ones. For example, unlike the traditional, it may be requested some attributes to support nonlinear playback, because user can access the media interactively according to object or location in which he/she is engaged. Additionally, it needs more layers than before to be assigned for additional metadata associated with the shape descriptors, positional/pose sensor data and so on. If we can provide useful information using augmented reality technique to tourists in right time and right place based on the recorded experiences from others, it could be a new platform of tour guide system.

In this paper, we introduce our new format, MART-MAF to deal with several modalities to cope with digitizing one's experiences. As shown in Figure 2, MART-MAF is defined based on ISO/IEC 14496-12 that is ISO Base Media File Format [1]. User can record their experiences in the form of MART-MAF with their own phone and then the created MAF file is delivered to MART server.

Currently, MART-MAF supports several media types, such as video, audio, visual descriptors, e-compass, GPS, accelerometers, user annotations, still image and 3D

978-0-7695-3890-7/09 $26.00 © 2009 IEEE
DOI 10.1109/ISM.2009.117

synthetic objects. Once MAF files are uploaded to the server, then users are able to download MAF file from MART web server according to his/her preference so that they can be guided to the POIs through the downloaded MAF file with augmented reality.

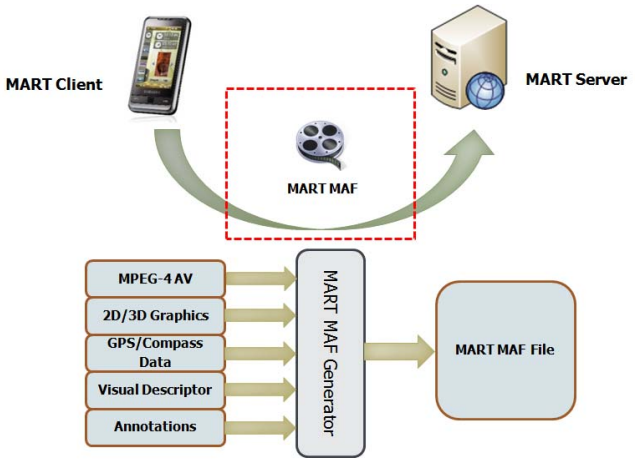

Figure 2. Architecture for in-situ annotation using MART-MAF

2 MART-MAF

We specify MART-MAF as media file format for AR tour guide service. We show the file structure for MART-MAF in Figure 3. As mentioned in previous section, MART-MAF is based on ISO base media file format, which is a general format forming the basis for a number of other more specific file formats. This format contains the timing, structure and media information for timed sequences of media data, such as audio/visual presentations. Besides, MART-MAF includes 3D synthetic objects for AR visualization and scene graph for rendering and user interaction. MART-MAF also includes various sensor data for recognizing POI (Point of Interest), such as GPS, gyro, accelerometer and e-compass data.

The file is structured as a sequence of objects; some of these objects may contain other objects. The sequence of objects in the file shall contain exactly one presentation metadata wrapper (in the movie box). It is usually close to the beginning or end of the file, to permit its easy location. The other objects found at this level may be a filetype box, movie box (moov), metadata or media data boxes. The media data is composed of images, video, audio data, sensor data, visual descriptors and annotations. The track identifiers used in a MART-MAF file are unique within that file; no two tracks shall use the same identifier. A track is allocated to each media type and therefore a POI can be composed of several tracks. These track identifier values and their correspondent POI information is stored in metadata in movie box.

Unlike most conventional mpeg standards which use timed sequences, MART-MAF can support not only timed sequences but also interactive sequences invoked by user's interaction. To consider this characteristic, we want to make good use of metadata box and fill it with scene graph which arranges the logical and spatial representation of a graphical 2D/3D scene. A scene graph is a collection of nodes in a graph or tree structure. A node may have many children but often only a single parent, with the effect of a parent apparent to all its child nodes. For user interaction, we can associate events to some nodes which can be spatially correlated or with author's intention. For the functionalities, we currently consider several xml-based language schema such as OpenSceneGraph[3], KML[4], MPEG4-BIFS[5]. Because of the nature of a tour, we may have several POIs with different position in a tourist place. We should reflect that kind of characteristic so that we take advantage of the concept of Track in MAF. Each track in MAF is used as a media container for each point of interest with different GPS value in tourist place. Track can also contain sub tracks so that each sub track contains each media file for the point of interest. In Figure 3, we show a conceptual content for each track for each point of interest.

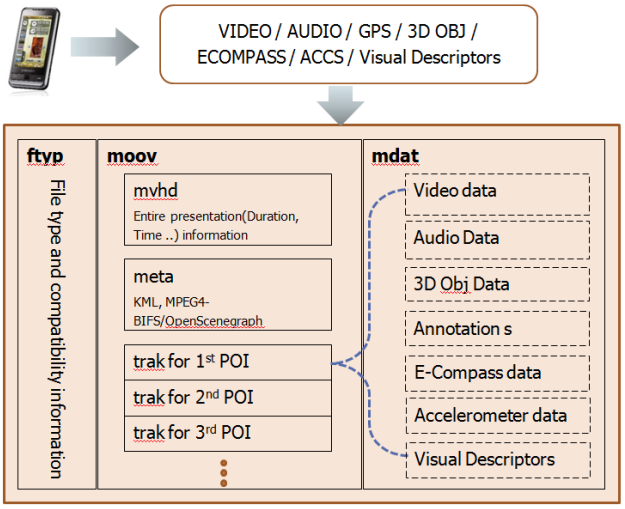

Figure 3. File structure of MART-MAF

3 SYSTEM ARCHITECTURE FOR MART-MAF

After taking a photo or capturing video, we can add some annotations for the POI and explanatory 2D/3D media objects. In addition, we can add sensor data from GPS, gyro, accelerometer, digital compass and visual descriptors. The created MART-MAF supports streaming of media data over a network as well as local playback. The communication protocol between MART server and the client we use is our own one but it is similar to the conventional hypertext transfer protocol. After the server receive MART-MAF, it parse the MAF file into several

media types and metadata is collected into the abstracted database and detailed database separately.

In abstracted metadata database system, we arranged the metadata into multi-level of abstraction to provide user with wider varieties of query methods. We also provide detailed metadata database to store information about conventional metadata using time based annotation model. This combined approach of metadata database enables user both of conventional media retrieval and new high level experience retrieval as well. An abstracted multi-level metadata management system gives information about the whole media file regardless of the temporal occurrence of the metadata. This approach is extremely useful for continuous capture of user's experiences in AR tour service. The detailed metadata database is implemented using eXist NXD [6] and the abstracted metadata database is implemented using PostgreSQL [7].

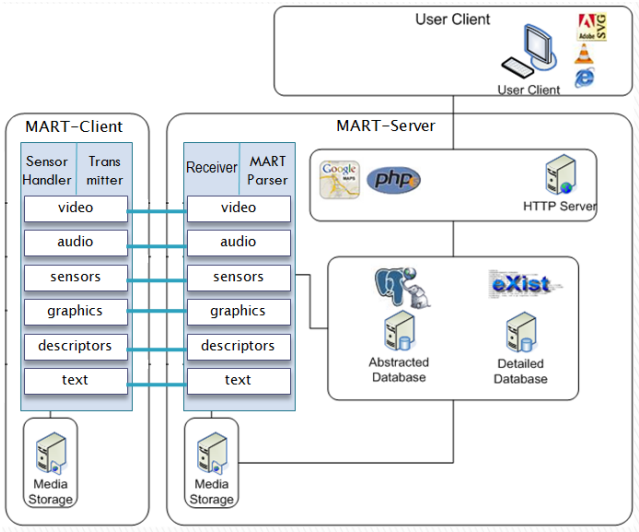

Figure 4. System architecture for MART-MAF

4 EXPERIMENT

We experimented on a Samsung mobile phone (SCH-M490, PXA 312 806Mhz, Windows Mobile 6.1) to evaluate our proposed structure for in-situ annotation. Although the mobile phone we used already includes camera, GPS and Wi-Fi module, since it doesn't have gyroscope and digital compass sensors, we have to use an external sensor module to gather directional information. We therefore used additional IMU sensor module which supports tri-axial accelerometer, gryo and digital compass with Bluetooth transmission. For MART-MAF transmission, we use Wi-Fi module in mobile phone[1].

Since we are currently using Google Earth as a map service, we adopt KML as an xml-based scene graph for rendering image, annotation and 3D object in the form of COLLADA [8]. We show COLLADA is the format of choice for digital content creation tools, and for archiving rich content. COLLADA enables the free exchange of asset data, enabling developers to construct an authoring pipeline from multiple tools. COLLADA can provide the content required for compelling X3D applications, in much the same way that it can carry and process the content for game applications. Moreover, current KML support it and it can be rendered on Google Earth so that we choose a tentative 3D file format for AR visualization. We also may use OpenSceneGraph for rendering our scene graph for more complex 3D scene and user interaction instead of KML.

Figure 5 shows the result of a simple demonstration. In this example, we generate MART-MAF in which we embedded 2D image, text based annotation, mpeg4 video, 3gpp audio and COLLADA for 3D synthetic object. We check the annotated information on the web browser just after annotating because the created MART-MAF can be sent to the MART server in real time.

Figure 5. Example for the visualization of uploaded MART-MAF in Google earth

A small yellow tack in the map in Figure 5 (left) shows that there's annotated information for the POI and if we click the tack, the popup which contains information will appear as shown in the right side of Figure 5.

In Figure 6, we show an internal file structure of MART-MAF that we created in previous simple demonstration. As explained, each media is contained in each track in MAF independently so that it can be accessed by any media player to play an appropriate media. In metadata box, we currently use KML for describing 2D/3D scene to be rendered on user's mobile phone. It can be replaced with user's preferable scene descriptor such as OpenScene Graph script, MPEG4-BIFS, X3D and so on.

[1] Currently we support both Wi-Fi transmission and WCDMA transmission. Although the number of Wi-Fi hotspots is ever increasing, most of regions are still left in Wi-Fi unreachable. If Wi-Fi service is unavailable, our client module sends the recorded MAF file using WCDMA network.

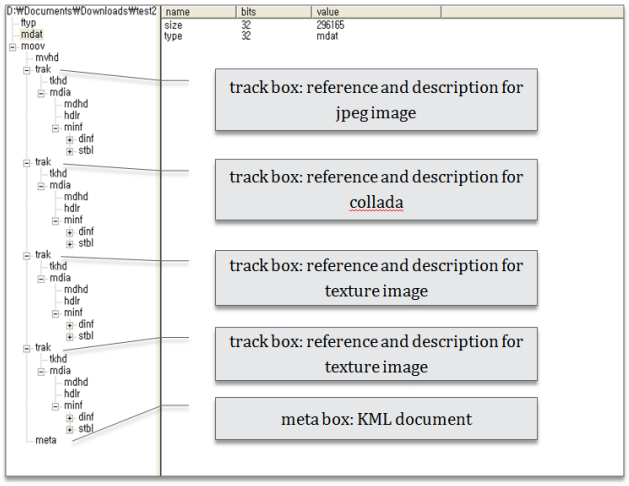

Figure 6. Internal file structure of created MART-MAF

5 CONCLUSION

The conventional guide systems, such as travel brochure and introductory signboard on the spot, may lose their interests with lapse of time unless they are changed. It may cause for tourists not to feel like visiting the same place again. However, users created contents based on their own experiences can give a various and dynamic tour guide for the same attraction so that they will lead to let people revisit the place. From this perspective, we need a new framework for recording a user's raw, unfiltered and vivid experiences. Additionally, we need to a new file format which is able to contain those experiences in the form of digital media.

In this paper, we propose a new file format for mobile tour service using augmented reality. Unlike previous major AV file formats, a new file format for useful and interesting tour guide using AR needs more interactive playback and searching. It can include 3D model for realistic visualization. In addition, it may have various types of objects and events, such as location, direction, altitude, velocity, spatiotemporal information, and etc.

Since MPEG-A aims to serve clearly identified market needs by facilitating the swift development of innovative and standards-based multimedia applications and services, we make a normative specifications of multimedia formats, called MART-MAF, which allow interoperability on an application level. MART-MAF is defined based on ISO/IEC 14496-12 that is ISO Base Media File Format and therefore it has several advantages. Although MART-MAF includes various types of media, it has scalability that supports playback compatible media only in specific players. In addition, since it can handle various media as a unit of track, it is easy to manage mobile tour data which have various kinds of media for POIs and it is suitable format for media streaming. To make good use of annotated information in situ, we implement server-client structure for real time streaming for in-situ annotation. For this, we support wireless transmission on mobile phone using either WiFi or WCDMA. For efficient transmission, we need to consider a new streaming protocol. SOAP with attachment is one of good candidates.

6 ACKNOWLEDGEMENT

This research is supported by Ministry of culture, Sports and Tourism (MCST) and Korea Creative Content Agency (KOCCA) in the Culture Technology (CT) Research & Development Program 2009.

REFERENCES

[1] ISO/IEC 14496-12, Coding of audio-visual objects, Part 12.

[2] ISO/IEC JTC 1/SC 29/WG 11/N6832, Palma de Mallorca, ES, October 2004, MPEG-A Multimedia Application Format Overview and Requirements (v.2), Wo Chang, editor. From

www.chiariglione.org/mpeg/standards/mpeg-a/mpeg-a.htm

[3] B. Kuehne and P. Martz, OpenSceneGraph reference manual, Skew matrix software and Blue newt software, 2007.

[4] KML Tutorial,

http://code.google.com/intl/ko/apis/kml/documentation/.

[5] ISO/IEC JTC1/SC29/WG11 N1683, "MPEG-4 Overview", April, 1997.

[6] eXist Open Source XML Database, Wolfgang Meier, http://exist.sourceforge.net/.

[7] PostgreSQL, PostgreSQL Global Development Group. http://www.postgresql.org

[8] R. Arnaud and T. Parisi, "Developing Web Applications with COLLADA and X3D", A White Paper, March, 2007