Intelligent Multimedia Content Management on Mobile Devices

Bhavan Gandhi, Alfonso Martínez, and Frank Bentley
Motorola Labs
gandhi@labs.mot.com, martinez@labs.mot.com, fbentley@labs.mot.com

Abstract— **The increasing capability of mobile handsets, advancing multimedia processing, improving codec technology (e.g., mp3, JPEG, MPEG-4 SP, H.264), and bigger communication pipes have the potential for providing increasing volumes of multimedia data to mobile users. The problem of intelligently managing multimedia content is becoming increasingly prevalent in this space. For example, applications that archive and retrieve personal content as well as those that search and stream commercial content over wired and wireless channels need to provide a compelling user experience while transparently and efficiently handling the vast amounts of underlying data. This paper provides a user-intuitive, standards-based approach for managing multimedia content on mobile handsets.**

I. Introduction

UNTIL recently, the problem of creating, consuming, and communicating multimedia content on handheld cellular phones has been constrained by the physical limitations of those devices. Increasingly, the cellular handset is becoming a viable platform for creating and consuming multimedia content (photos, music, videos, etc). The limitation still remains on how to manage these multimedia entities while providing a meaningful user experience. Intelligent content management technology facilitates managing the plethora of multimedia content on mobile handsets. This is facilitated by creating, storing, and manipulating its associated metadata. Metadata is able to represent semantic information that a user finds meaningful about the multimedia content. Allowing a user to manage his multimedia based on semantic attributes provides an easy to use environment, especially on mobile handsets.

This paper provides an overview of current handset capabilities and metadata management technologies (section II), describes the functional software architecture of a content management system (section III), and provides an example photo management application (section IV). Finally, section V presents conclusions and provides insight into future directions of research in this area.

II. Handset Platform Capabilities And Content Management Technology

The capability of handset platforms to create and consume multimedia and the emergence of content management technologies enable an intuitive environment for users to manage (i.e., capture, store, search, retrieve, etc.) their multimedia.

A. Mobile Platforms - Current Capabilities

Mobile handset manufacturers are introducing products that are more multimedia capable. State-of-the-art handsets are shipping with integrated cameras, color displays, and support for polyphonic sounds. Platforms support capturing and consuming photos (JPEG) and video clips (H.263 and MPEG-4 SP), as well as rendering music (mp3). Handsets also provide a communication means for transferring media over a variety of channels such as Bluetooth and IrDA. Table I [1] [2] provides a snapshot of current capabilities of multimedia-enabled handsets. It is reasonable to expect that processing speed, image capture, and display capabilities will continue improving.

Even though handsets are increasingly more capable, they are still limited by their physical size and interface. As such, it is one thing to consume multimedia on a handset, and it is another to manage the multimedia content intuitively. Currently, media management on mobile handsets is counterintuitive, file-based, and is not conducive to managing varied or large media collections. The user has the burden of creating static file structures and then recreating that structure in his mind to look for specific content. As explored by Lansdale, humans are poor at conceiving static organization systems [3]. Depending on the context of the situation that they are in when they access stored information in the future, the static groupings they create may not aid them in finding the content. Metadata associated with multimedia content can capture semantically meaningful information, which can be employed by a content management system to allow users to find content in many dynamic ways.

B. Metadata Management Technology

An effective content management method is necessary to facilitate a user to create, consume, archive, search, and retrieve his media. This requires that handsets be capable of managing and manipulating the metadata associated with the multimedia content.

There are many formats for metadata including MPEG-7 [4], Dublin Core [5], EXIF [6], and ID3 [7]. MPEG-7 and Dublin Core define XML-based syntax for describing multimedia content, whereas EXIF and ID3 provide specific file tags for describing the content of images and music respectively.

A user may typically exchange his media with other mobile devices, or synchronize her multimedia content with

0-7803-8603-5/04/$20.00 ©2004 IEEE

Handset	CPU (MHz)	Display	Imager	Multimedia Support
Motorola A760	200	320 x 240, 12 bpp	640 x 480	image/audio/video
Nokia 3650	200	176 x 208, 12 bpp	640 x 480	image/audio/video
Samsung i519	400	320 x 240, 16 bpp	640 x 480	image/audio/video

TABLE I
MOBILE HANDSET CAPABILITIES - 2003

a central server. This makes it necessary to have a common format for exchanging the multimedia and its associated metadata to ensure interoperability. To this end, an emphasis is placed on consuming, creating, and managing metadata descriptions using attributes from MPEG-7 MDS [8]. Additional support is also provided for consuming EXIF and ID3 descriptions, since they are accepted industry formats.

III. METADATA MANAGEMENT ARCHITECTURE

Figure 1 shows the software architecture supporting multimedia content management on mobile handsets. The architecture was designed as a layered approach to hide physical file organization from the user. The core of the architecture is the Metadata Services Layer, which translates higher semantic meaning tasks handed to it from the Application Layer into sets of lower semantic meaning operations to be processed by the Device Interface Layer. The Metadata Services Layer is stateless in nature, and therefore communicates with the Application Layer through a Session Manager Layer that maintains state of a user session.

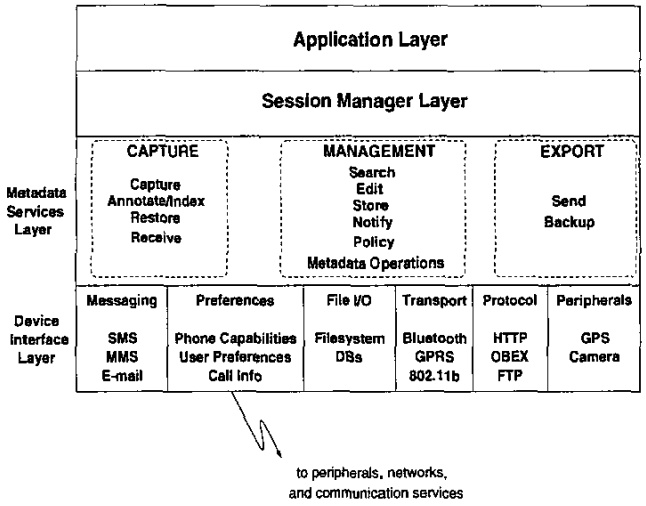


Fig. 1. Organization of the Architecture's Layers and Services

The Metadata Services Layer contains services that can be classified into three functionalities: importing media into the phone, organizing the media on the phone, and exporting the media off the phone (see Fig. 1). Capture services pull in media and create, parse or extend its metadata. Media management services are used to organize digital items (i.e., media and its associated metadata) that reside on the mobile handset. Media export services prepare metadata associated with an outgoing media instance, bind the metadata and the media into a digital item, and manage the export protocols and transports. Examples of importing media include receiving purchased media, capturing media using the phone, or downloading personal media from another device. Examples of exporting or communicating the media off the phone include sharing it with another device using MMS or e-mail, or uploading the media to another device. A service from one category may rely on services from other categories to perform its task. E.g., when exporting media, Send (in the Media Export category) calls Metadata Operations (in the Media Management category) to search for metadata, extend it with current usage information, encode it, and bind it to the media.

The Metadata Operations service shown within the Metadata Services Layer of Fig. 1 is critical to this architecture. These operations are required during the capture, management, and export of metadata-enriched media. They include a number of specialized lower-level services falling into two categories. The first category includes services to extract, translate, encode, and transcode metadata. Media may be annotated with different types of metadata, e.g., EXIF [6], id3 [7], or MPEG-7 [8]. Those metadata formats are parsed and stored in a common standard format (MPEG-7) preserving the original metadata. When exporting media, the original metadata may need to be attached (e.g., to preserve rights information), along with any updates and transcodings into a common metadata format to ensure interoperability across devices and networks. The second category of low-level metadata operations includes the ability to search, retrieve, generate, and delete metadata, verify existence of metadata and media bindings, manage metadata collections, and provide elementary search and query operations.

The Metadata Services Layer is stateless, so it is accessed by the Application Layer through the Session Manager Layer, where the overall state of the session is maintained. State or context information on the use of the media can be put to many practical uses (e.g., usage pat-

terns, preferences). Operations on non-metadata-extended media may be logged for similar purposes but operations on metadata-extended media provide rich information of a higher semantic level that is useful for media distribution and syndication, not to mention to the user himself.

The Session Manager Layer coordinates the calling flow into the Metadata Services Layer, simplifying the job of application developers. The Session Manager Layer knows the sequence in which it needs to call metadata services to fulfill the task handed to it from an application. For example, a picture capture application would call the Session Manager Layer. The Session Manager then calls the Capture service. The Capture service retrieves the image from the camera drivers and calls the Store service with the data of the newly captured image. The Store service saves the image according to the Policy service and calls the Annotate service. The Annotate service uses the current time stamp, GPS location, PIM information, and data from other images (using the Search service) to annotate the image. Then it calls Metadata Operations to save annotations in MPEG-7 format, performing transcoding from the native JPEG (EXIF) metadata. In summary, the Session Manager Layer maintains global state and performs common functionality for all applications, automatically annotates usage information as items are accessed and used, and provides a simplified API to hide the Metadata Services Layer for developers to write applications.

Lastly, the Device Interface Layer, which is present in current state-of-the-art mobile handsets, provides local connectivity for advanced features (e.g., Bluetooth, camera access, global positioning system, messaging services) as well as traditional features.

IV. Example Applications

We approached the problem of managing content on limited, mobile devices from the perspective of potential users. The true value of metadata management is in allowing users to find the content that is important to them using cues that are relevant to them at the time of the search. As these cues are different for different media types and the context of access, it is important to understand the user's own model of the item's context.

Our services support management of content of any type. Music files can be managed in a way similar to the currently popular iTunes [9] and Winamp [10] products which allow users to search by song title, album, and artist. Photos and short video clips can be managed in a similar way. In this section, we present an application for managing digital photos on a device with limited screen resolution, input capabilities, and processing power.

In order to define functional requirements for a photo-management system, we performed a series of rapid ethnographic interviews on six diverse members of the Motorola community focused on how consumers share pictures today, both digital and print. We used observations from these interviews to guide the design of the services described above as well as the Media Assistant application described in this section.

A. Platform

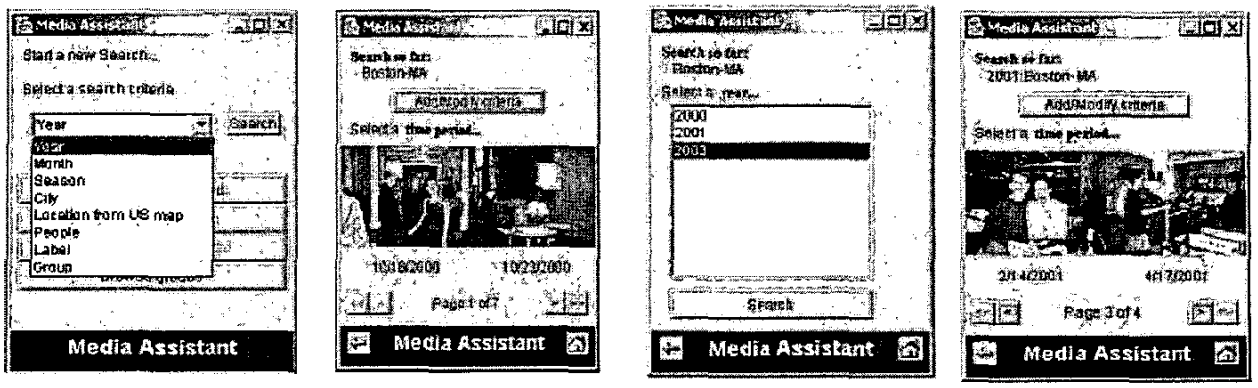
Our photo-finding application, named Media Assistant, is implemented in Java and runs on a limited device platform running a 400MHz Intel XScale processor with 64 MB of memory. The application runs on top of the services described above which are also implemented in Java. The user interface is provided via a touch screen on a 320x240 pixel display.

B. Managing Digital Photos

The Media Assistant application allows users to find digital photographs from their collections. It provides metadata in the form of time and location cues for users to find photos in the system, as these are easy to determine without user intervention. Based on our user research, these attributes relate to how people remember the context of their pictures. Other metadata cues can optionally be provided by the user such as a semantic title or the names of the people in the photo. We use event groupings similar to those in [11] to allow users to view events as an intermediate search result before being presented with all of the pictures in the result sets. Systems by Loui [12] and Platt [13] also explored using time-based groupings to aid in automatic organization of personal photo collections.

As pictures are taken, they are saved and annotated using the Annotate service mentioned above. This attaches the current time (the time of capture) and the current location of the camera to the photo in the metadata database. A simple lookup table allows for GPS locations to be converted to semantic region names such as "Chicago" or "San Francisco." More sophisiticated net-based services could provide a finer degree of specification for this location. This automatic categorization will allow the user to locate the image in the future without having to manually store the file in a set folder in a hierarchy.

When a user is searching for a picture, he simply uses the information that is salient to him at the time. For example, imagine a group of pictures taken on a family trip to Boston for Christmas in 2002. A user could remember this as "A trip to Boston," or "December 2002," or "A trip to Boston sometime after 2000," or countless other ways. Our interface allows the user to select any criteria that he remembers and choose from a list of values for that criteria, saving him from typing into the limited device. After selecting a criteria, he is shown thumbnails with dates under them representing the events at which he took pictures that match the search criteria. For example, if the user



Original choice of attributes | **Viewing 14 events from Boston, MA.** | **Selecting an available year** | **Viewing 8 events from Boston, MA in 2001.**

Fig. 2. Screen Shots from our demo Media Assistant application demonstrate the power of flexible metadata-based searching for digital photographs.

picked 2002, all events from 2002 would be displayed. The thumbnail is chosen as the first picture from that event as this photo represents the stimulus that made the user take out the camera.

At this point, the user can choose to view pictures from one of those events, or he can refine the search by selecting another available criterion, such as location, and picking a value from those represented in the set of search results currently displayed. For example, if in 2002 pictures were only taken in Boston and Chicago, these would be the only choices available even if there was a trip to San Francisco in 2003, since the search was already limited to only those pictures in 2002. This search flow is illustrated in Figure 2.

After the user has found a photo or set of photos, he is able to transfer it to a remote device. When files are transferred, they are sent along with their MPEG-7 metadata so that remote devices can provide the same searching experience as the originating device.

C. User Validation

After iterative testing with other participants, a rapid assessment [14] of our services was conducted with four participants outside of Motorola Labs on photo sets from several years' worth of personal digital pictures. Users were able to navigate the interface to find their pictures in the complete absence of any manual organization on their part. For the purposes of the study, we manually entered the locations for each of the pictures in their collection to simulate how the live system would work with newly captured pictures.

V. Conclusion

Using underlying metadata associated with multimedia content to represent semantically meaningful information from a user's perspective is essential to providing an intuitive experience in managing media on mobile devices. A photo management application built on top of metadata-based content management service classes was presented. The application flow and the set of identified attributes were derived from user centered research. The underlying content management architecture and software modules were developed to interoperate with MPEG-7 MDS, EXIF, and ID3. We expect to build upon this research to allow for more seamless exchange of multimedia and its associated metadata within a heterogeneous network of devices.

References

[1] "Mobile Phones News and Reviews (MobileBurn)," http://www.mobileburn.com/.
[2] "infoSync World," http://www.infosyncworld.com/.
[3] M. Lansdale, "The Psychology of Personal Information Management," *Applied Ergonomics*, vol. 19, no. 1, 1988.
[4] S.-F. Chang, T. Sikora, and A. Puri, "Overview of the MPEG-7 Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 6, pp. 688–695, June 2001.
[5] "Dublin Core Metadata Initiative," http://dublincore.org/.
[6] Japan Electronics & Information Technology Industries Association Technical Standardization Committee on AV & IT Storage Systems & Equipment, "Exchangeable image file format for digital still cameras: Exif version 2.2," *JEITA CP-3451*, Apr. 2002.
[7] "ID3v2," http://www.id3.org/.
[8] ISO/IEC JTC 1/SC29 N4242, "Information Technology – Multimedia Content Description Interface – Part 5: Multimedia Description Schemes," *ISO/IEC FDIS 15938-5*, July 2001.
[9] "iTunes," http://www.apple.com/itunes/.
[10] "Winamp," http://www.winamp.com/.
[11] A. Graham, H. Garcia-Molina, A. Paepcke, and T. Winograd, "Time as Essence for Photo Browsing Through Personal Digital Libraries," in *ACM Joint Conference on Digital Libraries*. 2002, ACM.
[12] A.C. Loui and M.D. Wood, "A Software System for Automatic Albuming of Consumer Pictures," in *ACM Multimedia '99*. 1999, ACM.
[13] J.C. Platt, M. Czerwinski, and B. Field, "PhotoTOC: Automatic Clustering for Browsing Personal Photographs," in *Microsoft Technical report MSR-TR-2002-17*. 2002, Microsoft Research.
[14] James Beebe, *Rapid Assessment Process: An Introduction*, Alta Mira Press, Walnut Creek, 2001.