Mobile Media Metadata for Mobile Imaging

Bibliographic Reference:

Marc Davis and Risto Sarvas. “Mobile Media Metadata for Mobile Imaging.” In: *Proceedings of IEEE International Conference on Multimedia and Expo (ICME 2004) Special Session on Mobile Imaging in Taipei, Taiwan*, IEEE Computer Society Press, 2004.

Mobile Media Metadata for Mobile Imaging

Marc Davis
Garage Cinema Research
School of Information Management and Systems
University of California at Berkeley
http://garage.sims.berkeley.edu
marc@sims.berkeley.edu

Risto Sarvas
Helsinki Institute for Information Technology (HIIT)
Helsinki University of Technology
http://www.hiit.fi/risto.sarvas/
risto.sarvas@hiit.fi

Abstract

Camera phones offer a new platform for digital imaging that integrates capture, programmable processing, networking, and rich user interaction capabilities. This platform can support software that leverages the spatio-temporal context and social community of image capture and (re)use to infer media content. Our Mobile Media Metadata prototype (MMM) uses a small custom client on the camera phone and its XHTML browser along with a metadata server to enable annotation at the time of image capture, leverage contextual metadata and networked metadata resources, and use iterative metadata refinement on the mobile imaging device.

1. Introduction

Mobile phones with media creation capabilities are rapidly entering the marketplace in the USA and already have significant market presence in Asia and Europe. While camera phones offer a tremendous opportunity to create and share images, sound, and video, they also bring with them the inherent problem of media management. As we capture more and more media every day, the pressing need for solutions to manage media grows ever greater. With hundreds of pictures, human memory and browsing can manage content; when one has many thousands of pictures, the individual pictures one may want are effectively lost. The solution is *metadata* that describe the content of mobile media. However, we face a conundrum: automatic media analysis cannot represent media content in ways that address human concerns and purely manual media annotation is too time-consuming for consumers. A third way is to leverage the spatio-temporal *context* and social *community* of media capture in mobile devices to infer media content.

While useful work has been done on consumer image annotation [1], the vast majority of prior research on personal image management has assumed that image annotation occurs *after image capture in a desktop context*. Time lag and context change affects the likelihood of the annotation task being performed as well as the photographer's accurate recall of the contextual information to be assigned to the photograph [2]. Importantly, the devices and usage context of consumer digital photography are undergoing rapid transformation from the traditional camera-to-desktop-to-network image pipeline to an integrated mobile imaging experience. The availability of mobile, networked digital imaging devices with operating systems (e.g., Symbian) and support for high level programming languages (e.g., Java, C++) using open standards and accessible APIs means that multimedia researchers (and consumers) now have a new platform for the development of digital imaging applications that can leverage:

1) Programmable processing at the point of media capture
2) Device and network supplied temporal, spatial, and social metadata
3) Network resources for processing and communication
4) Rich interaction with the camera user

We have explored the integration of capture, processing, and interaction at the point of media capture in our research on "Active Capture" [3]. Camera phones bring a new dimension to our research by enabling the automated gathering of contextual metadata—temporal, spatial, and social (e.g., username and presence)—to create, infer, and learn annotations of media content.

Related work has looked at leveraging temporal [4] and spatial [5] information for organizing captured images. Furthermore, by utilizing the networking,

interaction, and contextual metadata capabilities of camera phones, mobile image annotation applications can not only use temporal and spatial metadata, but also enable and leverage the creation, sharing, and reuse of media and metadata among communities of users. We developed our MMM ("Mobile Media Metadata") system [6] independently of, but around the same as, related approaches [7, 8], and moving beyond these systems, we leverage social, temporal, and spatial contextual metadata together as well as interaction at the point of capture to make inferences about media content. In our approach we:

- Gather all automatically available information at the point of capture (time, spatial location, phone user, etc.)
- Use metadata similarity and media analysis algorithms to find similar media that has been annotated before
- Take advantage of this previously annotated media to make educated guesses about the content of the newly captured media
- Interact in a simple and intuitive way with the phone user to confirm and augment system-supplied metadata for captured media

As a result of this approach, we believe we will solve a fundamental problem in consumer adoption of mobile media services—the need to have content-based access to the media consumers capture on their mobile phones and devices.

2. Campanile scenario

To understand our approach to mobile media metadata, creation, sharing and reuse, consider the following scenario. Everyday hundreds of visitors photograph the Campanile tower at the University of California at Berkeley (See Figure 1). With current digital imaging technology this is a solitary and highly inefficient process. Although thousands of people every year are basically taking the same photo in the same place, they are unable to easily share any metadata one or more of them might create about their common photographic subject. Furthermore, after capturing the image in the camera, the process of transferring, storing, and potentially sharing the photograph is cumbersome, time-consuming, and requires interfacing with a desktop computer and network. Digital images are identified by cryptic sequential file names (e.g., "pic0047.jpg") and the software available for manually adding metadata requires more time and commitment than most consumers have or care to offer.

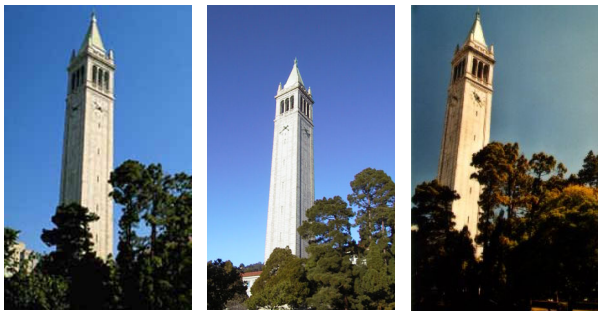

Figure 1. The Campanile at UC Berkeley

With our camera phone annotation prototype we leverage both the spatio-temporal *context* and social *community* of capture to radically simplify and automate mobile media metadata creation and (re)use. Imagine that when a person takes a picture of the Campanile, the mobile phone uploads the image, spatio-temporal metadata, and user name to a metadata server. The metadata server analyzes both the media content and metadata to find media and metadata that are similar to the captured media—i.e., media that were captured at the same place and time and that resemble one another. By looking at the metadata associated with the media captured by others at the same place and time, the metadata server can suggest additional metadata that the user can approve or not for the captured media. Metadata ascribed to the pictures above (and thousands like them taken near the Campanile in the day time) would enable the server to indicate to the mobile phone user that the picture they have just taken is likely "an outdoor picture of the Campanile." Camera phones enable us to apply this approach to media analysis and management by using regularities in the media capture context to infer media content, i.e., to leverage the coherence of the world and of our collective process of recording it.

3. System description

Our developed prototype, MMM, offers unique opportunities for consumer photo management by enabling annotation at the time of image capture, leveraging contextual metadata and networked metadata resources, and enabling iterative metadata refinement on the mobile imaging device. While we have not yet integrated media signal analysis into our system, our metadata server enables us to make inferences about media content based on spatial, temporal, and social metadata that is interactively suggested to and refined (i.e., confirmed, corrected, or augmented) by the user.

MMM combines a GSM/GPRS camera phone and a remote web server in a client-server architecture (See

Figure 2). Using our client software on the phone, the user captures the image and selects the *main subject* of the image (*Person*, *Location*, *Object,* or *Activity*) before uploading it to the server. The server receives the uploaded image and the metadata gathered at the time of capture (*main subject*, *time*, *date*, network *cellID*, and *user name*). Based on this metadata, the server searches a repository of previously captured images and their respective metadata for similarities. The images and metadata in the repository are not limited to the user's own images and metadata, but contain every user's annotated media to leverage the advantages of shared metadata.

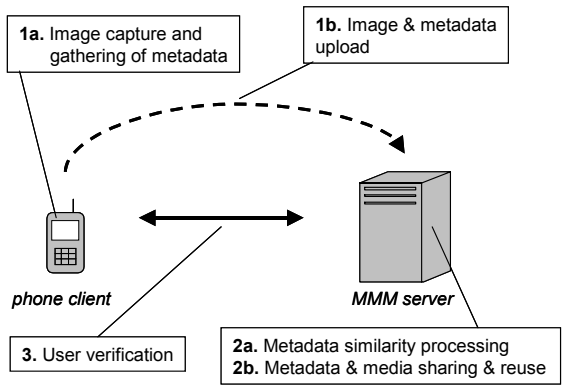

Figure 2. Mobile Media Metadata (MMM) system overview

Using the information from previously captured images that have similar metadata, the server program generates educated guesses for the user (i.e., selection lists with the most probable metadata first). The user receives the server-generated guesses for verification, and confirms, corrects, or augments the metadata. Below we describe the system implementation in more detail by dividing it into the main parts of the metadata creation process described above.

3.1. Image capture and metadata gathering

The client-side image capturing, user selection of main subject, automatic gathering of metadata, and communication with the server were implemented in a C++ application named *Image-Gallery*. It was developed in cooperation with Futurice (www.futurice.fi) for the Symbian 6.1 operating system on the Nokia 3650 phone. The user captures the image using *Image-Gallery* which automatically stores the *main subject*, *time, date,* GSM network *cellID*, and the *user name*. The image and metadata upload process was implemented in *Image-Gallery* and on the server side using the Nokia Image Upload API 1.1.

3.2. Metadata similarity processing

The server side metadata similarity processing was implemented in a Java module that provides a set of algorithms for retrieving metadata using the metadata of the image at hand and the repository of previously annotated images. The values returned by the metadata processing and retrieval are the guesses sorted in order of highest probability. In the MMM system we implemented two main sets of algorithms: location guessing and person guessing which leverage the various intersections of spatial, temporal, and social similarity as illustrated in Figure 3.

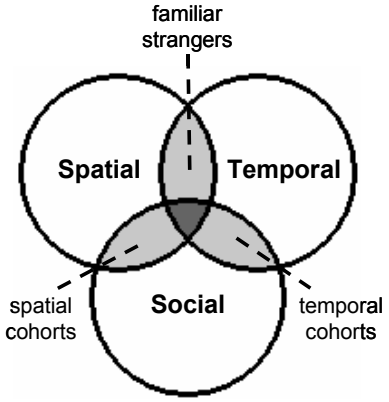

Figure 3. Connecting spatial, temporal, and social metadata

The patterns in where, when, and with and of whom individuals, social groups, and cohorts take photographs have discernible regularities that we use to make inferences about photo content. Our location guesser uses a weighted sum of interrelated spatial, temporal, and social features: most recently visited location by the user; most visited location by the user in this CellID around this time; most visited location by the user in this CellID; most visited location by other users in this CellID around this time; and most visited location by other users in this CellID. Interestingly, notions of what it means to "visit" a location and "other users" are complex and rely on spatial, temporal, and social congruities. For example, I can "visit" a location by taking a photograph there and/or by being photographed there. "Other users" may be connected to a given user in a variety of intersecting spatial, temporal, and social relations.

3.3. Metadata and media sharing and reuse

One of the main design principles in the MMM system is to have the metadata shared and reused among all users of the system. This means that when processing the media and metadata, the system has

access not only to the media and metadata the user has created before, but the media and metadata everyone else using the system has created. While this sharing may raise privacy concerns, several factors can reduce these risks: anonymization of the metadata; aggregate use of the metadata; and leveraging social networks for metadata sharing. Metadata sharing enables us to leverage a variety of shared spatial, temporal, and social relations across many users (See Figure 3) to improve metadata guessing. For example, the probability that a given MMM user has taken a photograph at a given place and time (e.g., at home on the weekend) can be increased by another MMM user having photographed the MMM user before in that place at a similar time.

The images and their respective metadata are stored in an open source object-oriented database (Ozone 1.1) on the server. The metadata is stored in a faceted hierarchical structure. In our structure the facets were the main subjects of the image: *Person*, *Location*, *Object*, and *Activity*. The objective of the faceted structure is for the facets to be as independent of each other as possible, in other words, one facet can be described without affecting the others. Facetted hierarchical structures enable vocabulary control during annotation and precise similarity processing for retrieval.

3.4. User verification

The user verification and system responses were implemented in XHTML forms. After uploading the image and metadata, the client-side *Image-Gallery* program launches the phone's XHTML browser to a URL given by the server during the uploading. After the server creates the metadata guesses to facilitate the user's annotation work, it creates XHTML pages from the guesses for the client-side browser to present to the user. The dialog between the server and the user is then implemented in the form data sent from the phone to the server, and the XHTML pages created by the server that are rendered by the phone's browser.

3.4. System deployment

MMM has been deployed since September 2003 and was used by 40 graduate students and 15 researchers at UC Berkeley's School of Information Management and Systems in a required graduate course "IS202: Information Organization and Retrieval" co-taught by Prof. Marc Davis and Prof. Ray Larson. Students used the MMM prototype and developed personas, scenarios, storyboards, metadata frameworks, and presentations for their concepts for mobile media and metadata creation, sharing, and reuse applications.

4. Future work

In our future work, we will be integrating media analysis algorithms with our metadata inferencing algorithms which we are further testing and refining based on the datasets from the 4 month trial of our Mobile Media Metadata prototype.

5. Acknowledgements

The authors would like to thank British Telecom, AT&T Wireless, Futurice, and Nokia for their support of this research and the members of the Mobile Media Metadata project in Garage Cinema Research at the School of Information Management and Systems at the University of California at Berkeley.

6. References

[1] A. Kuchinsky, C. Pering, M. L. Creech, D. Freeze, B. Serra, and J. Gwizdka, "*FotoFile*: A Consumer Multimedia Organization and Retrieval System," presented at SIGCHI Conference on Human Factors in Computing Systems (CHI '99), Pittsburgh, PA, 1999.

[2] D. Frohlich, A. Kuchinsky, C. Pering, A. Don, and S. Ariss, "Requirements for Photoware," presented at 2002 ACM Conference on Computer Supported Cooperative Work (CSCW '02), New Orleans, LA, 2002.

[3] M. Davis, "Active Capture: Integrating Human-Computer Interaction and Computer Vision/Audition to Automate Media Capture," presented at 2003 IEEE Conference on Multimedia and Expo (ICME2003) Special Session on Moving from Features to Semantics Using Computational Media Aesthetics, Baltimore, MD, 2003.

[4] M. D. Cooper, J. Foote, A. Girgensohn, and L. Wilcox, "Temporal Event Clustering for Digital Photo Collections," presented at ACM Multimedia 2003, Berkeley, CA, 2003.

[5] K. Toyama, R. Logan, and A. Roseway, "Geographic Location Tags on Digital Images," presented at 11th ACM International Conference on Multimedia (MM2003), Berkeley, CA, 2003.

[6] R. Sarvas, E. Herrarte, A. Wilhelm, and M. Davis, "Metadata Creation System for Mobile Images," presented at Second International Conference on Mobile Systems, Applications, and Services (MobiSys2004), Boston, MA, 2004.

[7] P. Vartiainen, "Using Metadata and Context Information in Sharing Personal Content of Mobile Users," in *Department of Computer Science*. Helsinki, Finland: University of Helsinki, 2003, pp. 67.

[8] M. Naaman, A. Paepcke, and H. Garcia-Molina, "From Where to What: Metadata Sharing for Digital Photographs with Geographic Coordinates," presented at 10th International Conference on Cooperative Information Systems (CoopIS 2003), Catania, Sicily, 2003.