A MOBILE PLATFORM FOR ON-THE-MOVE USER-GENERATED TOURIST CONTENTS

Ander Garcia, Maria Teresa Linaza, Imanol Arbulu, Isabel Torre, Yolanda Cobos

Department of Tourism, Heritage and Creativity, Visual Communication Technologies VICOMTech

ABSTRACT

The paper presents the implementation of a platform which allows the generation, annotation and retrieval of on-the-move multimedia tourist contents while on a tourist destination. The platform toggles the different stages of the processes involved on the cycle of life of the content: creation, indexation, storage and publication. The application has been implemented using the J2ME standard to take advantage of the standard APIs implemented on the last generation mobile phones. Content annotation is based on multilingual folksonomies, allowing a widespread approach to describe the content and strengthening the social nature of the on-the-move multimedia user-generated content. An authorized person or body filters it before being published. Once user-generated multimedia tourist contents are accepted, they are published using a tool into a web page. This GUI allows browsing for contents provided by individual users or those including a tag or keyword. Finally, the paper provides a brief summary of the evaluation of the prototype.

1. INTRODUCTION

Prior to the advent of digital cameras, an average camera user was severely limited both in the amount of photos that could be taken (limited mainly by the financial and temporal cost of purchasing as well as of developing films), and in the ways those photos could be organize (the prevailing method was albums, usually arranged more or less chronologically). Digital camera quickly broke down the barriers. However, manual classification and organization of the generated pictures is still tedious. Fortunately, digital cameras provide useful metadata automatically such as time, location and EXIF parameters, including focal length, exposure time or flash-light status.

Innovations in consumer photography have made it exceedingly simple for people to capture images, which they do at an ever-growing rate. This growing capture rate is driven by the proliferation of capture devices (such as digital cameras and camera phones) together with decreasing storage costs.

At the same time, creation of semantic metadata about photo content remains an elusive goal. Some amount of annotation can significantly improve the usefulness of such photo collections as they grow into the thousands.

However, algorithms for semantic interpretation and annotation of image content are far from reach by any automated system. Providing tools for annotation of media is therefore an active field of research in Human-Computer Interaction. These research efforts, mainly focused on desktop-based tools, attempt to ease the annotation task as well as maximize the benefits for annotating content. While some of this research has been incorporated into commercial photo browsing systems, most people still do not bother about annotating their photos, though they think it would be useful for later photo retrieval.

Folksonomies represent one of the most successful annotation systems within the emerging Web 2.0 technologies. A folksonomy is composed of tags or labels, usually one single-word, that attach metadata to objects (text, photos, videos, …) in order to add contextual meaning to the items themselves. These tags are not taken from a limited vocabulary, but each user can select the preferred taxonomy. Then, this content can be published in a server accessible by other users through user-generated content platforms, such as YouTube or Flickr.

This paper focuses on the vision of digital personal tourist experience, providing mobile support for recording, tagging, storing and sharing their experiences in destinations, just when the event they want to share or record is happening. Content-on-the-move is a special type of user-generated content: as soon as personal experiences are captured from mobile devices, users can archive and edit them from the same mobile devices that are used for capturing. It is no longer necessary for tourists to download the content to their desktop to share it, but the process can be directly fulfilled through their mobile devices.

This paper has been organized as follows. Section 2 presents a brief summary of the related work, including several aspects related to mobile content capturing and annotation processes, as well as several existing applications. The main objectives of the project are presented in Section 3. Then, Section 4 focuses on the description of the technical components of the system. A small assessment of the prototype is described in Section 5. Finally, Section 6 includes some conclusions and further work.

2. RELATED WORK

2.1. Mobile content capture

Recent advances in multimedia hardware manufacturing technologies have led to a growing consumer market of affordable camera-equipped mobile devices such as smart phones or PDAs. For many users, these devices have become ubiquitous as part of their everyday inseparable items. Their great market success can be attributed to convenience brought by the combination of a digital camera and a radio communication within one small mobile device. Its benefits include the communication capability to share and distribute these recorded personal experiences on the Internet, and the ability to record everyday personal experiences. People can become content producers of their own personal experiences. The ideas that "everyone can be a content producer" and "everyone has a content-producing mobile device" are expected to bring a fundamental change in the type of future digital contents.

In the capturing phase, many research activities have focused on context-aware multimedia, which includes techniques for intelligent metadata acquisition at the capturing moment rather then later. Life log agent created by Aizawa *et al* is a system that can capture videos and audio from a wearable camera [1]. As they are being captured, contents are automatically annotated with context metadata from a GPS receiver and an accelerometer. These annotations are used as index keys, allowing a user to input queries in the form of who, what, where and when.

Another system, the MMM system, can automate content metadata extraction on camera phones using available context information, such as location and time [2]. When a photograph is taken at a location, the system can reuse metadata from previous photographs taken at the same location. This approach requires a centralized repository to store the shared metadata.

Finally, Wu *et al* have designed, implemented and evaluated a mobile authoring tool called mProducer that enables everyday users to effectively and efficiently perform archiving and editing of digital personal experiences from their camera-equipped mobile devices [3]. This point-of-capture capability is crucial to enable sharing of digital personal experiences anytime, anywhere.

2.2. Annotation approaches

To effectively manage, access and retrieve multimedia data, a widely adopted solution is to associate the image content with semantically meaningful labels, annotating the content. There are two types of image annotation approaches available: automatic and manual. The former, which aims at automatically detecting the visual keywords from image content, has attracted a lot of attention from researchers in the last decade. Although automatic annotation approaches have achieved notable success recently, it remains a challenge for them to accurately annotate other more specific and less visually similar keywords.

Along the latter, recent years have seen a proliferation of manual image annotation systems for managing online/personal multimedia content. Examples include PhotoStuff and Aria for personal archives, or Flickr and ESP Game for online content. The rise of manual annotation partially stems from its high annotation quality for self-organization/retrieval purpose, and its social bookmaking functionality that allows public search and self-promotion in online communities.

Manual image annotation approaches can be categorized into two types: tagging and browsing. The former allows users to annotate images with a chosen set of keywords (tags) from a controlled or uncontrolled vocabulary. For example, Flickr encourages users to create free-text tags for each uploaded image. It views tags as the central component for sharing, retrieval and discovery of user-generated content. On the other hand, the latter requires users to sequentially browse a group of images and judge their relevance to a pre-defined keyword.

Efficient labelling of photos has been an active research field since 1999. The latest photo browser commercial packages, such as Adobe Photoshop Album, adopted methods to support easy labelling of photos. More recent efforts utilize temporal and spatial context to assist in labelling photos in personal collections.

Vartiainen's research [4] features a shared semantic ontology for mobile image annotation, but it does not leverage social, temporal and spatial contextual metadata to make inferences about media content.

Moreover, MobiCon [5] is a video production tool for mobile camera phones, which integrates video clip capture with context-aware, personalized clip annotation- supporting automatic annotation suggestions based on context data and efficient manual annotation with user-specific ontologies and keywords- and clip sharing secured by digital rights management techniques. Thus, the system allows users to inexpensively create metadata-annotated video clips for a better management of their clip collections and keeps them in control of the clips they share.

However, recent studies on the management of personal photo collections have determined that users often use only event-based organization for their photos and apply little or no additional annotation or organization.

2.3. Applications in the tourism domain

Within the tourism sector, recent studies have shown that tourists' travel decisions are increasingly being determined by content and reviews generated by other "peer" tourists [6]. Combining this fact with the growing willingness of tourist of sharing their experience with others [7], it is not surprising the numerous examples of user created and managed travel web sites and applications available (travelblog, wikitravel, …). Even the official tourism organizations are successfully starting to include user-generated content on their sites [8] as a way of promoting their destination.

Nowadays, this type of user-generated tourist content is created once the day or the trip is over. The next step is to allow tourists to directly generate content while they are on the destination, just when they are experiencing the object they want to share. In this situation, content on-the-move takes importance, sharing the experiences as they have just been lived and capturing their essences.

There are some applications from generic fields that have worked on this direction. For instance, ZoneTag is a camera phone application used to upload photos taken by the phone to Flickr [9]. ZoneTag is loosely based on a system designed to capture, annotate, store and share photos from the phone. More importantly, ZoneTag attempts to encourage annotation on the phone at the time of capture by providing tag suggestions.

Furthermore, the LOCALE system at Stanford allows devices and users to share location information and labels for photographic images [10]. It uses location to determine what labels other photographs taken in a similar location should have. LOCALE uses free text annotations of location.

Finally, LocoBlog [11] tested a space-time photo travel blogging based on geo-referenced content. Using the system it was possible to upload multimedia entries to each own blog.

3. OBJECTIVES

The objective of this project is the implementation of a platform which allows the generation, annotation and retrieval of on-the-move multimedia tourist contents while on a tourist destination. The platform targets four major phases on the digital content production, as shown in Figure 1. The personal experiences are first captured as digital content using cameras on mobile devices. The digital content is indexed or annotated using multilingual tags and then stored and archived on the mobile device and on the remote server database. The digital content is then supervised, and edited if necessary, and finally it is published.

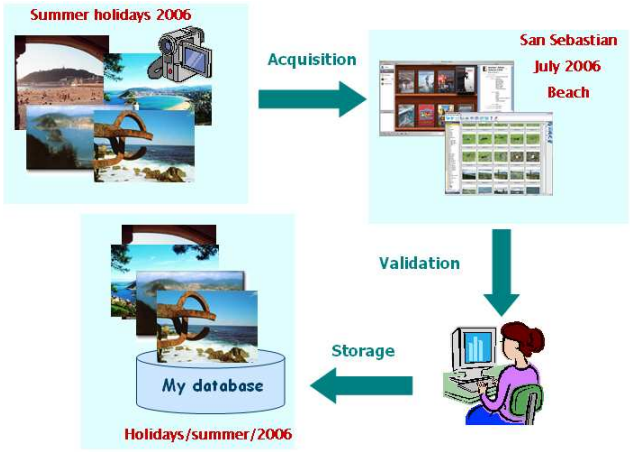

Figure 1. Graphical representation of the system.

In general, most users do not carry PCs in the mobile environment, especially when they are on holidays. Thus, content archival and editing are usually delayed until users return home or to their office. This production delay can reduce the value of personal experiences that are time sensitive, meaning that audience interest in the content decreases with the time delay. Thus, the implemented application should allow the selection of a multimedia content file and its annotation, taking into account the limitations of mobile devices.

Moreover, user-generated content management should take care about the storage and retrieval of those contents for several purposes. A special effort has been devoted to implement a multilingual system for content sharing among tourists, no matter the language used in the annotation process.

The application should allow also filtering the user-generated contents before publishing them, for example, in a tourist destination web site, in order to avoid undesired contents. An additional benefit of the mobile editing is that it allows a user to operate a single device during the entire content production lifecycle, therefore, saving user effort in transferring content between devices.

Finally, and related to the publishing of the files in the remote server, a functional web will be implemented based on Web 2.0 concepts so that users can access the existing contents.

4. TECHNICAL COMPONENTS OF THE APPLICATION

Figure 2 displays a graphical approach to the system architecture. The tourist takes a picture with the camera of the mobile device. After annotating it and automatically including the GPS data if available, the content and its associated metadata are sent to the server and also stored on the device. The server receives this content and stores it with the metadata on the system database.

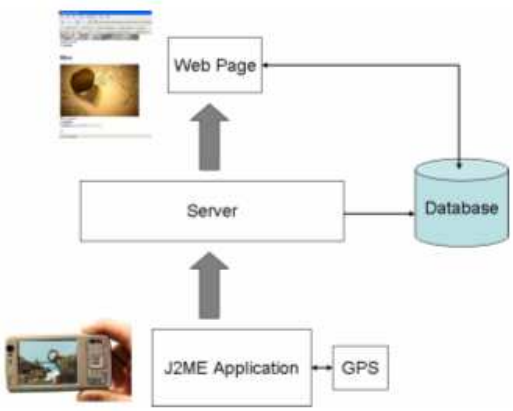

Figure 2. Architecture of the system.

Finally, after the filtering process has been completed, the user-generated tourist content is published through a Web page.

Although optional, the filtering capacity has been implemented so that contents could be revised and controlled before publishing them. This functionality is sometimes demanded by some of the tourist destination offices in order to control the suitability of comments and pictures about their destination.

The following sections describe the components included in the application.

4.1. Content database

Contents are a key issue in the platform as they are the elements that will be stored, retrieved and displayed to the user. Therefore, the success of the prototype will depend on their attractiveness, as more people would like to join the community if they are engaged with added-value contents. Although the platform is flexible enough to handle any multimedia format, the contents for this first prototype will be based on photographs, as they are one of the most popular ways for tourists to remind their trips.

4.1.1. Definition of the database

The system includes three types of data that should be stored in the database: the user-generated tourist content, tags for semantic retrieval and data about the users.

User-generated content. A model for describing the on-the-move multimedia tourist content has been defined, including the name or title and short description of the tourist Point of Interest, and geo-references if automatically available, as well as the owners and associated rights of the contents.

Tags for semantic retrieval. Tags or keyword used for semantic retrieval of contents are stored independently in order to manage multilingualism. Tags in several languages referring the same concept are stored as a unique tag, relating directly the content to the tag in all the languages. Although the current prototype supports three different languages (English, Spanish and Basque), it can be easily extended to include new languages. Following the theory of the long tail applied to folksonomies, which states that few tags are used most of the times, a tag will not be created if it is already stored on the system.

Data about the users. User models also need including structures to describe personal demographic profiles and preferences. Demographic data about users such as name, surname, telephone or e-mail address, as well as user ID and password for the server connection will be stored by the system.

Figure 3 shows a simplified view of the implemented database, which has been implemented using MySQL. The number of metadata associated with the multimedia content has been restricted as the main objective of this preliminary prototype is to analyze the viability of the system. Therefore, it is important to retrieve contents as dynamically as possible, taking into account the limitations imposed by mobile devices.

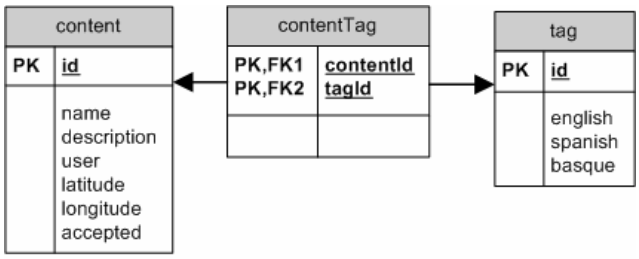

Figure 3. Simplified view of the database.

4.2. Mobile application

It must be mentioned that the platform is not devoted to the generation of personal albums, but user-generated contents for the promotion of the tourist destinations while on-the-move. Therefore, the application should be dynamic enough so that a tourist touring in a city with the application installed a mobile device can execute the application, capture and annotate the Points of Interest, and send the recorded pictures in a user-friendly way.

4.2.1. Design of the application

The requirements of the application have been mainly determined by the limitations of the mobile devices. The application should be user-friendly and intuitive, so that it can be easily used by occasional users. The limitations of the mobile devices regarding accessibility should be taken into account, including the difficulties of writing large texts. Moreover, the limitations of the screen of the mobile devices are crucial due to the variable size.

The design of the prototype can be described using the flow chart shown in Figure 4.

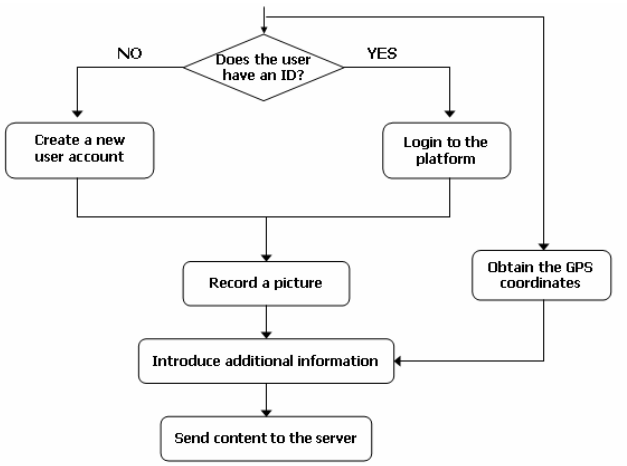

Figure 4. Flow chart of the application.

The system interface has been implemented in a simple and user-friendly style. Figure 5 displays some screenshots corresponding of the different steps of the application. The first step is the establishment of the connection between the mobile device and the server. Therefore, the first screen includes the necessary field to authenticate the user through the ID and password. The first time a tourist enters the application, a further GUI will appear in order to introduce his/her new data and create a new account.

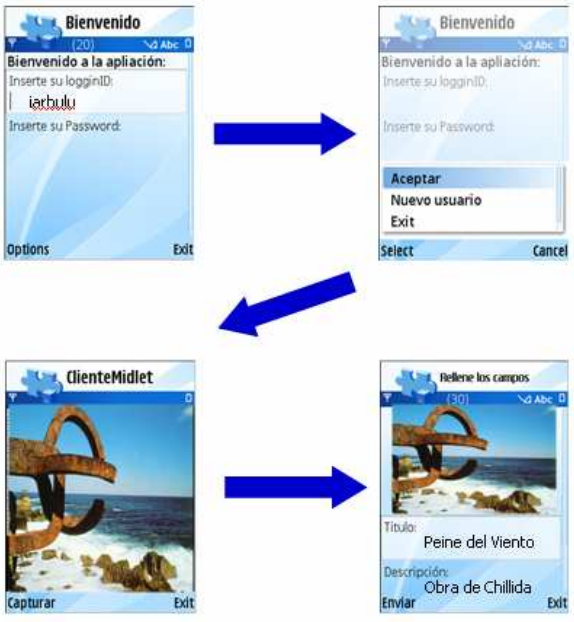

Figure 5. Screenshots of the annotation prototype.

Once the user has logged in, the application connects with the camera of the mobile device. At the beginning of a new content recording, in case available, the system queries the GPS receiver to obtain the location of the device through an independent thread. Thus, the user is not aware of any possible delays on acquiring the GPS position.

After the service has been accessed, the application records the image of the camera and displays the content that is being captured. The user only has to click when the proper picture is displayed. The file is then stored in the image folder of the memory card of the mobile device before sending it to the server. The following step includes another GUI for the addition of further information as name, description and tags. Before uploading, when available, the GPS coordinates provided by the mobile device are added to the tags. Once the fields have been filled out, the information is sent to the server.

4.2.2. Selection of the mobile device

The mobile device requires some type of connection to Internet, such as GPRS, 3G or WLAN. However, the user has to initialize the connection before starting the application. Regarding the GPS, the user should also manage the availability of the GPS if the geo-reference can be added as metadata.

Due to the technical possibilities offered by the device, the prototype has been implemented on a Nokia N95, which has built-in WiFi and Bluetooth modules, a 5 MP digital camera, a GPS receiver and 3G and GPRS connectivity. Therefore, all the proposed functionalities have been tested.

4.2.3. Implementation of the application

The application has been implemented using Java both for the server and the mobile client, using the J2ME standard to gain minimum interoperability. The application takes advantages of the standard APIs implemented on the last generation mobile phones, such as the location API JSR-179 to manage the GPS connection, the Mobile Media API JSR-135 to control the camera, and the Web Service specification JSR-172 to communicate with the server using Web Services.

The implemented application has been designed for last generation mobile devices, as it is compulsory the use of the API JSR-172 for Web Services. Therefore, it can be run no matter the underlying machine that works as a client, as long as the mobile device supports sending and receiving SOAP messages.

Although it could seem to be a very restrictive limitation, it is expected that mass market mobile devices will include these requirements in short term.

However, the API JSR-172 does not allow adding attachments while running the Web Services, Therefore, the communication between the server and the mobile device is managed using a servlet when uploading the content.

During the development process, a simple PC-based client has also been developed. This way, the server has been tested and debugged using different mobile devices. Finally, it must be mentioned that it is expected that future mobile APIs will allow attaching inside calls to Web Services, overcoming current limitations.

4.4. Publishing and browsing over a Web page

Once the user-generated tourist content has been stored, an authorized person or body filters it before being published. While editing the content, metadata can be changed, including modifying or deleting tags, or the name or description of the contents. Moreover, tags must be translated from the user language into the different supported languages if multilingual content retrieval is wanted.

Once user-generated tourist contents are accepted, they are published using a tool into a Web page. This GUI allows browsing for contents provided by individual users or those including a tag or keyword.

When search through tags is performed, the system automatically retrieves the related content, even if they have been originally annotated in a different language. Current mobile browsers are not able to properly handle AJAX due to the reduced JavaScript version they support. Therefore, a servlet has been developed for the mobile clients. Figure 6 shows a screenshot of this Web page.

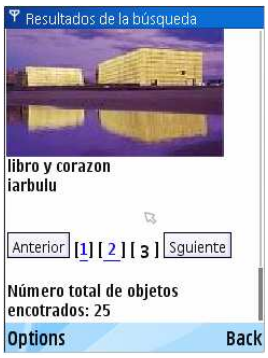

Figure 6. GUI for browsing on the mobile device.

5. EVALUATION OF THE PROTOTYPE

Once the application has been implemented, a small assessment campaign has been conducted in order to mainly analyze the functionalities of the prototype from a technical point of view. It must be mentioned that further evaluation with non-expert users is expected to happen in the forthcoming months.

Quantitative methodology based on questionnaires was selected for the evaluation. Questions were mainly clustered among three subgroups: general performance of the prototype, including aspects such as the user-friendliness of the menus, the GUI or the processing time gaps; existing functionalities as the performance of the camera or the automatic capturing of the content; and possible future extensions to include other multimedia formats (audio, video) and the handling of the Intellectual Property Rights. Finally, there was also the possibility of add text-free comments.

The sample has been selected among the scientific personal of VICOMTech due to the main goal of this evaluation. The average age of the users was between 20 and 30 years old, graduated and with the same percentage of female and male.

The test has been performed with the same Nokia N95 used during the development phase. The location chosen for the test has been the Technological Park of the city of Donostia-San Sebastian. Although 3G is available in this area, the local WiFi network was used for the tests. Moreover, all the contents generated during the tests were geo-tagged.

Before starting the tests, a brief introduction (around three minutes) was given to the users. As all of them knew how to use the device, the introduction was devoted to explain the way of creating and accessing the user-generated multimedia content.

Regarding the general performance of the prototype, users were quite satisfied with the current version, although processing time should be improved in order to shorten waiting times. These delays are very significant when the system searches for existing wireless networks to upload the user-generated tourist contents. However, the acquisition of the GPS coordinates does not cause any trouble when interacting with the system.

In relation to future extensions, most of the users agree on the possibility of introducing other types of multimedia contents such as videos or audios.

6. CONCLUSIONS AND FUTURE WORK

We describe our design and implementation of a mobile annotation tool that enables tourists to capture and edit their personal experiences at the point of capture while on the destination from a mobile device. The system can transform our everyday camera-equipped, mobile devices from simply content capturing devices to content producing devices. Its unique aspect is that it enables immediate point-of-capture editing and archiving from a mobile device, so that users can quickly distribute digital personal tourist experiences.

The implemented system includes three types of data that are stored in the database: the user-generated tourist content, tags for semantic retrieval and data about the users. Moreover, the number of metadata associated with the multimedia content has been restricted as the main objective of this preliminary prototype is to analyze the viability of the system.

The requirements of the application have been mainly determined by the limitations of the mobile devices. The system interface has been implemented in a simple and user-friendly style. The application has been implemented using Java both for the server and the mobile client, using the J2ME standard to gain minimum interoperability. The application takes advantages of the standard APIs implemented on the last generation mobile phones. Once user-generated tourist contents are accepted, they are published using a tool into a Web page. This GUI allows browsing for contents provided by individual users or those including a tag or keyword.

A small assessment campaign has been conducted in order to mainly analyze the functionalities of the prototype from a technical point of view. The sample has been selected among the scientific personal of VICOMTech due to the main goal of this evaluation. It can be concluded that users were quite satisfied with the current version, although processing time should be improved in order to shorten waiting times. Finally, most of the users agree on the possibility of introducing other types of multimedia contents such as videos or audios.

Several improvements have been proposed for further works. Current metadata will be extended in order to be MPEG-7 compliant. However, it must be mentioned that the standard will be simplified in order to handle only fields that are supposed compulsory for the proper semantic retrieval.

The translation process of tags is going to be automated, using automatic translation systems that have a public API, in order to ease the work of the person in charge of the filtering process. Finally, the final interface of the Web page is going to be updated using Google Maps to take advantage of the geo-reference data.

7. ACKNOWLEDGEMENTS

The authors would like to thank the Basque Government for partially funding this work through the eTourgune project.

8. REFERENCES

[1] K. Aizawa, T. Hori, S. Kawasaki, and T. Ishikawa, "Capture an efficient retrieval of life log," *Pervasive 2004 Workshop on Memory and Sharing of Experiences,* 2004.

[2] R. Sarvas, E. Herrarte, A. Wilhelm, and M. Davis, "Metadata creation system for mobile images," *MobiSYS'04: Proceedings of the 2nd International Conference on Mobile Systems, Applications and Services,* ACM Press, pp. 36-48, 2004.

[3] C. Wu, C.J. Teng, Y. Chen, T. Lin, H. Chu, and J.Y. Hsu, "Point-of-capture archiving and editing of personal experiences form a mobile device," *Pers. Ubiquit. Comput. 11*, Springer-Verlag, London, pp. 235-249, 2007.

[4] Vartiainen, P., *Using Metadata and Context Information in Sharing Personal Content of Mobile Users*, Master's Thesis, University of Helsinki, Finland, 2003.

[5] J. Lahti, U. Westermann and M. Pakola, "MobiCon- Integrated Capture, Annotation and Sharing of Video Clips with Mobile Phones," *Proc. of the MM'05*, Singapore, pp. 798-799, November 6-11, 2005.

[6] P. O'Connor, "User-Generated Content and Travel: A Case Study on Tripadvisor.com", *Proc. of the Information and Communication Technologies in Tourism Conference ENTER 2008*, Innsbruck, Austria, pp 47-58, January 21-24, 2008.

[7] Y. Kang, J. Stasko, K. Luther, A. Ravi, Y. Xu, "RevisiTou: Enriching the Tourism Experience With User-Generated Content", *Proc. of the Information and Communication Technologies in Tourism Conference ENTER 2008*, Innsbruck, Austria, pp 59-69, January 21-24, 2008.

[8] M.T. Linaza, F. Lolhoffel, A. Garcia, C. Lamsfus, A. Alzua and A. Lazkano, "Mash-up Applications for Small Destination Management Organizations Websites", *Proc. of the Information and Communication Technologies in Tourism Conference ENTER 2008*, Innsbruck, Austria, pp 130-140, January 21-24, 2008.

[9] M. Ames and M. Naaman, "Why we tag: Motivations for Annotation in Mobile and Online Media,", *CHI 2007 Proceedings-Tags, Tagging and Notetaking*, San José, California, USA, pp. 971-980, 2007.

[10] B.N. Lee, W.Y. Chen, and E.Y. Chang, "A Scalable Service for Photo Annotation, Sharing and Search," *MM'06*, ACM, Santa Barbara, California, USA, pp. 699-703, October 23-27 2006.

[11] W. Bamford, P.Coulton, and R. Edwards, "Space-time travel blogging using a mobile phone", Proceedings of the International Conference on Advances in Computer Entertainment Technology, ACM Press New York, NY, USA, pp. 1-8, 2007.