Handbook of Research on User Interface Design and Evaluation for Mobile Technology

Volume I

Joanna Lumsden
National Research Council of Canada
Institute for Information Technology – e-Business, Canada

Information Science REFERENCE

INFORMATION SCIENCE REFERENCE
Hershey • New York

Acquisitions Editor:	Kristin Klinger
Development Editor:	Kristin Roth
Senior Managing Editor:	Jennifer Neidig
Managing Editor:	Sara Reed
Copy Editor:	Joy Langel, Katie Smalley, and Angela Thor
Typesetter:	Jeff Ash
Cover Design:	Lisa Tosheff
Printed at:	Yurchak Printing Inc.

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue, Suite 200
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com

and in the United Kingdom by
Information Science Reference (an imprint of IGI Global)
3 Henrietta Street
Covent Garden
London WC2E 8LU
Tel: 44 20 7240 0856
Fax: 44 20 7379 0609
Web site: http://www.eurospanonline.com

Library of Congress Cataloging-in-Publication Data

Handbook of research on user interface design and evaluation for mobile technology / Joanna Lumsden, editor.

p. cm.

Summary: "This book provides students, researchers, educators, and practitioners with a compendium of research on the key issues surrounding the design and evaluation of mobile user interfaces, such as the physical environment and social context in which a device is being used and the impact of multitasking behavior typically exhibited by mobile-device users"--Provided by publisher.

Includes bibliographical references and index.

ISBN 978-1-59904-871-0 (hardcover) -- ISBN 978-1-59904-872-7 (ebook)

1. Mobile computing--Handbooks, manuals, etc. 2. Human-computer interaction--Handbooks, manuals, etc. 3. User interfaces (Computer systems)--Handbooks, manuals, etc. I. Lumsden, Joanna.

QA76.59.H36 2008

004.165--dc22

2007024493

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

Chapter V
Interaction Design for Personal Photo Management on a Mobile Device

Hyowon Lee
Dublin City University, Ireland

Cathal Gurrin
Dublin City University, Ireland

Gareth J.F. Jones
Dublin City University, Ireland

Alan F. Smeaton
Dublin City University, Ireland

ABSTRACT

This chapter explores some of the technological elements that will greatly enhance user interaction with personal photos on mobile devices in the near future. It reviews major technological innovations that have taken place in recent years which are contributing to re-shaping people's personal photo management behavior and thus their needs, and presents an overview of the major design issues in supporting these for mobile access. It then introduces the currently very active research area of content-based image analysis and context-awareness. These technologies are becoming an important factor in improving mobile interaction by assisting automatic annotation and organization of photos, thus reducing the chore of manual input on mobile devices. Considering the pace of the rapid increases in the number of digital photos stored on our digital cameras, camera phones and online photoware sites, the authors believe that the subsequent benefits from this line of research will become a crucial factor in helping to design efficient and satisfying mobile interfaces for personal photo management systems.

INTRODUCTION

Long before digital technology came into everyday use, people have been managing personal photos with varying degrees of effort. Individuals' photo management strategies ranged from stacking photos in shoe boxes to carefully placing them into a series of photo albums with detailed notes of where and when each photo was taken or a witty caption beside it. Reminiscing and story-telling past events that have been visually recorded in personal photos is a highly-valued activity for many people. This gives meaning to the person's past events and also works as a socially-binding and relationship-enhancing device at gatherings of family or friends. With the Internet revolution, and the arrival of inexpensive digital cameras, people's photo organizing and sharing behavior has been evolving as new technologies allow different ways of managing photo collections. This is exemplified with online photoware applications such as Flickr[1], with which people can now upload personal photos taken from their digital cameras onto a shared web space on which collaborative annotation, browsing and sharing photos with other people is possible.

Another aspect of the development of digital photography is that people's behavior in capture of photos is changing as well. In particular, due to the low cost and ease of capture nowadays people are taking many more photos than in the past. This is possibly best illustrated by the ubiquity of camera phones, mobile devices that can be used as digital camera as well as a phone. Many people carry their phone with them at all times meaning that they can capture their everyday lives and holiday scenes whenever they want. This change in capture behavior can also have a significant impact on people's personal photo management activity. Once captured, the phone can be used to send photos to a friend's mobile phone or to upload them to a public Website for instant sharing and receiving comments back. This means that when designing personal photo management tools, we should consider the implications of the changes in user photo capture behavior arising from the emergence of the ubiquitous availability of the means of photo capture. For example, there is a need to design specific user-interfaces for photo management on a camera phone itself. A camera phone may be used merely as a capture device that takes photos and stores them, to be copied later to a PC for further photo management. However, the quality of screens now commonly available on mobile phones means that it is quite reasonable to look to design tools that enable users to organize, annotate and browse photos *on* the mobile phone itself. Between these two extreme cases, there is a spectrum of varying degrees to which a camera phone or other mobile device can be integrated into overall photo management functions and tasks, effectively a continuum of trade-off among technological resources and the user's effort and time. For example, due to the difficulty of text input on a camera phone arising from physical constraints, it may be easier for the user to fully annotate photos after copying them on to a desktop PC at home. Even so, a user in some situations might still want to make the effort to annotate their photos using the mobile and to send them to a friend for the benefit of its immediacy and not having to do the extra work of copying photos to a PC at home in the evening before performing the annotation. On occasion some users will want to bulk-upload a large number of photos taken at a party directly to a website without any annotation, e.g., to share with close friends. Depending on the design decisions on the allocation of photo management tasks for different devices, the optimal user-interface for such tasks on the mobile device will vary. Currently available interfaces on camera phones and digital cameras for photo management illustrate this possible diversity of user task requirements.

Particular challenges faced in designing and evaluating mobile interfaces for personal photo management arise due to, among other things, the following:

- New technology regularly emerges and applications constantly evolve.
- Mobile users are difficult to observe.

Consequently, it is difficult to rely on the traditional system development cycle of user study, user needs, and requirements establishment, followed by prototyping and evaluation. By the time this established process has been completed for an application, a new technological innovation may have appeared bringing in a new possible line of products to be developed. Conducting a user study on mobile devices can pose significant problems compared to desktop-based systems because a proper observation of users using a mobile device is technically difficult. When the user is on the move, we are simply unable to hook up an observation camera and recorder, and keep following the user while he or she is out and about using the device.

However, recently study methodologies have been developing to cope with these difficulties. For example, quicker and cheaper prototyping techniques are starting to be adopted to keep pace with changing technology (e.g., PC-based simulated prototypes or using a general mobile platform (Jones & Marsden, 2006, p. 179)). In another example, user studies are emerging which adopt light ethnographic and indirect observations using diaries and self-reporting to cope with test users on the move (Palen & Salzman, 2002; Pascoe et al., 2000; Perry et al., 2001). More specifically on the use of *camera phones*, some early longitudinal user studies have appeared (Kindberg et al., 2005; Sarvas et al., 2005; Van House et al., 2005) that aim to better understand the different motivations and current practices in camera phone use. These studies use diaries and interviews to capture usage data from 30-60 users within a period ranging from one to two months.

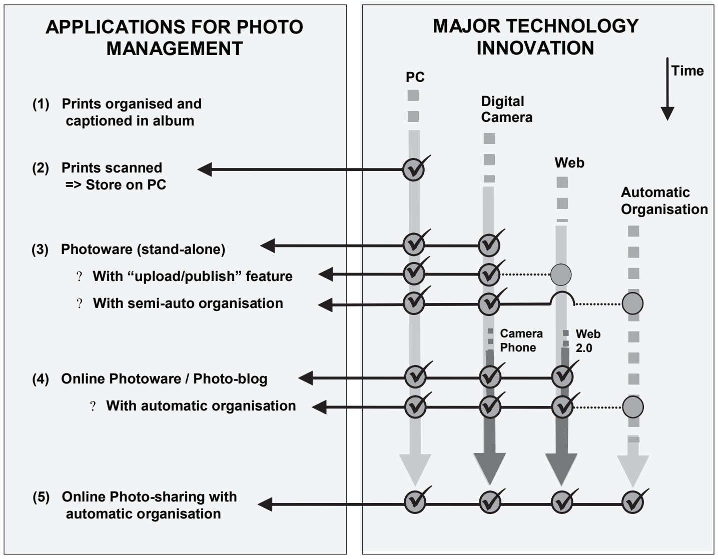
In understanding the current status of mobile interaction design for personal photo management and in setting the right direction for future applications, we need to look at the way such applications have been developed so far. As will be described in more detail in the following sections, applications for personal photo management have been incrementally shaped by the major technology innovations that have appeared during the last two decades or so. By looking at other new technologies, which are likely to be available in the near future, we can roughly determine what kind of personal photo applications or requirements will emerge in which mobile devices are an important component. The primary technology that we anticipate will appear is the *automatic organization* of the photos. The need for and adoption of this technology will be driven by the very rapid increase in the number of digital photos that users will accumulate on camera phones, desktop PCs, and online photoware. The sheer volume of photos means that it is increasingly difficult and frankly becomes unrealistic to manually annotate these photos.

Fortunately, by leveraging context data such as the time and location of photo capture, the bulk of the organization task can be done automatically. In addition, content-based image analysis techniques, although considered still not mature enough for many other applications, are also proving a very promising element that can further contribute to effective automatic organization and annotation of large photo collections. Use of these organization automation tools reduces the user's annotation burden to such an extent that a user working with large digital photo collections can focus on enjoyable browsing, searching and sharing tasks, rather than the nuisance of file extension and ongoing manual annotation. In this chapter we explore how mobile interaction design for personal photo management can take advantage of these emerging technological factors to overcome potential interaction design problems for photo management applications on a mobile device.

BACKGROUND: TECHNOLOGY TREND AND PHOTO MANAGEMENT ON MOBILE

Starting from physically printed photos, which are manually organized in an album, moving to current camera phones to capture, annotate, and browse digital pictures, the way we manage personal photos is evolving very rapidly. Current emerging applications have largely been geared to the major technological innovations that have occurred. These include the wide uptake of PCs

Figure 1. Supporting personal photo management at different stages of technological development



to the general public since the mid-80s, the World Wide Web and inexpensive digital cameras that have become commonplace since the mid- and late-90s. Each of these has brought a set of new ways of managing personal photos and the applications (with their associated features) have been quickly developed and used. This is illustrated in Figure 1.

On the right side of the figure, four vertical arrows indicate the major technologies: PCs, digital cameras and camera phones, Web and Web 2.0 technology, and finally a newly appearing technology, Automatic Organization. The first three of these have already shaped the way people manage their personal photos, in effect the *enabling technology* for personal photo management. These technologies were not mature when first emerging (thus the dotted line at the start of each of the vertical lines), but in time have become reliable enough to be taken up as full applications. The rows on the left side of the diagram show the way people manage their photos taking advantage of these technologies. For each row the cross point with the vertical lines indicates which technology has been used (circle with a tick mark). Circles without a tick mark indicate cases where the uptake of the technology has only been partial or done in a work-around way due to insufficient development of this particular technology, or of our understanding of its potential at that time. The last row is the envisaged future photo applications that take full advantage of the four technologies including Automatic Organization.

In this section we briefly go through each of the rows as numbered in the figure, highlighting how the enabling technologies influenced its development, what specific features became possible due to such technologies, and where mobile device and associated user-interface design issues arise from this.

Prints Organized and Captioned in an Album

Since the camera became a common household gadget, the most traditional way to manage personal photos was to get film developed and then select well-taken prints to insert into an album. Because the user selected only good photos for inclusion in a photo album, it was most likely to contain high quality or useful photos (Rodden, 1999). The major feature of an album is the grouping or organization into pages and the *captioning*, often appearing beside the chosen photos where the user adds a short description of the photo, often with a humorous comment. When people meet at a family event or party, they might gather around the album and flip through pages discussing and appreciating photos and their captions; good photos were also ordered for reprints to be given or posted to families and friends for sharing. These social activities have been popular as the main ways of browsing and sharing personal photos (Frohlich et al., 2002) and have often been adapted in photoware as a metaphor.

Prints Scanned and Stored on a PC

Before digital cameras became inexpensive and commonplace, those who owned PCs could scan prints of physical photos to store them digitally on their PCs. Preservation, digital enhancement or novelty were more likely motives rather than as a true replacement of chemical photos. Sharing was possible by copying files onto a disk and passing it to somebody who also owned a PC or more recently sending the photos as an email attachment (Frohlich et al., 2002; Rodden & Wood, 2003), but the PC did not come to be used as a tool for photo management until people started taking photos with digital cameras.

Stand-Alone Photoware

The availability of inexpensive digital cameras meant people could now capture their interesting events directly into digital format and copy them onto their PC. As the cost of capture is virtually zero and unwanted photos are easily deleted, the average number of photos taken increased dramatically. However, digital photos stored in PC directories often have cryptic file names such as "1430XX23-02.jpg" which have been generated automatically by the camera. The large quantity of such photos means that users often do not attempt to make the effort to rename them with more meaningful titles. It is thus not possible for users to find individual photos based on filenames.

As the Web started to become more common, some users manually created Web pages containing their photos to share with family and friends. However, this required much time and effort to find and select photos from within their collections and then to generate the Web pages, and only carefully selected photos were made available for sharing by users interested in both web and digital photography technologies.

This situation improved with the introduction of *photoware* software that imports photos from a digital camera and supports easy management of photos by allowing the grouping, sorting by date, annotation, and subsequently allowing searching and browsing. As people started accumulating large numbers of photos the utility of photoware grew. Examples of popular photoware include Photoshop Album[2], ACDSee[3] and Picasa[4].

Some experimental photoware systems support automatic grouping of photos based on time/date as the photos are imported from the camera (O'Hare et al., 2006), removing some of the organizing burden from the user while at the same time motivating higher quality annotation (Kustanowitz & Shneiderman, 2004). Other systems feature a convenient "upload" or "publish" feature whereby after organizing photos on a PC, the system generates attractive Web pages with the selected photos and captions.

Currently a large number of similar photoware tools are available each providing some combination of features including photo manipulation (rotation, brightness/contrast change, sharpening, red-eye removal, adding visual effects, etc.), organization (adding titles and descriptions, fast photo tagging, searching for duplicate photos, synchronising directories, etc.) and browsing and searching (thumbnail and full-screen views, slideshows, easy zoom-in/out, search by annotation, view colour histogram, etc.).

Online Photoware for Sharing and Photo-Blogging

With the advent of Web 2.0 in which the web is itself a platform, a highly interactive user-interface can be realized directly within the Web browser. Conventional stand-alone software applications such as word processors, time schedulers and e-mail clients are now available online, and photoware applications can be deployed in this way as well. Organization, annotation and other useful photo manipulation such as rotation and cropping, as well as various photo collection visualizations are now featured in online photoware in which the user directly interacts with a Web-based interface.

In addition, the proliferation of *camera phones* has resulted in ubiquitous use of capturing and sharing of digital photos, spurring ever higher quantities of digital photos to be taken. In parallel, some mobile phone services have started allowing digital photos taken on a camera phone to be directly uploaded to online photoware applications and shared with other users, bypassing the step of copying photos onto a PC. Use of a camera phone for saving, organizing, annotating, browsing, searching, and sharing photos raises multitudes of user interaction issues, as we will see in more detail in the next section.

Finally, some users have started regular online posting of their daily photos with annotations allowing a community of online citizens to comment on their photos and annotations, referred to as *photo-blogging*, merging the uploading of digital photos with the features of text blogging.

Emerging from the combination of these digital and online technologies, we have popular online photoware such as Flickr[5] and Yahoo! 360[6] providing highly interactive photoware features, while focusing on the online community and sharing of photos with such features as searching over community photos, popular photos of the day, commenting and photo grouping across users. Although, not yet a mature genre, the early design guidelines for online photoware can already be found drawn from general design principles and developers' experiences (Frohlich et al., 2002).

Online Photo-Sharing with Automatic Organization

Manually organizing photos into groups and subgroups and annotating them one by one is at times pleasant, but as our photo-taking habits change from being selective to taking as many photos as possible, creating meaningful captions for individual photos becomes more and more time consuming. However, to be able to subsequently access the photos, appropriate organization and annotation of each photo is crucial. While the subjective nature of indexing a visual medium is a problem and in itself an important research area (Enser, 1995), to help a user index hundreds or thousands of personal photos for efficient searching and browsing is a challenging problem both in terms of design and technique.

Photoware which *automatically* organizes and annotates photos for the user is an attractive possibility, and as with any other digital library project where a system automatically indexes the documents in the database, is becoming more and more feasible with technical advancements that are happening today. As will be described in detail in the following sections, much of the automatic organization and annotation for personal photos can be achieved by recording context information at the time of photo capture (such as time and location) in conjunction with content-based image analysis techniques to detect, for example, the existence of faces and buildings in the photos. These features, although promising, require further research and development to be able to deliver robust performance in real-life photo applications. At this point other semi-automatic or work-around schemes to leverage a user's manual annotation input have been proposed. For example, an initial annotation by the user at the time of photo capture can then be used later by the system to suggest more annotation options to the user (Wilhelm et al., 2004); a user does manual bulk annotation of faces appearing in a group of photos, after which the system automatically assigns the annotation to the faces in each photo (Zhang et al., 2004); and how a system could motivate its users to do enjoyable annotation is also considered (Kustanowitz & Shneiderman, 2004).

As seen in this section, photo management applications have been developed incrementally adding feature after feature whenever a new technology allowed it. In drawing up the last row in Figure 1 for a truly online automatic photo-sharing application for personal photo management, the role of mobile devices should be considered carefully especially in the light of the upsurge in the use of camera phones.

From this progression, we can now envisage a scenario in which a user takes photos at a party with her camera phone, these are instantly uploaded to an online photo server where they are fully and automatically processed and annotated, then a few days later, the user visits her uncle's house and searches on her phone for those photos taken at the party, and shares them by passing her phone around or by playing a slide show on the TV screen or an interactive wall in the house.

DESIGNING MOBILE INTERFACE FOR PHOTO MANAGEMENT

From the foregoing discussions we can see that personal photo management with a mobile device could involve activities such as:

- Capturing (taking photos)
- Storing and/or uploading
- Organizing and annotating
- Browsing and searching
- Sending and sharing of personal photos

However, not all of these activities need to be conducted on the mobile device itself. We need to consider ways in which a mobile device is best used in conjunction with other technology in varying degrees of division and overlap of task. In considering *mobile information ecologies* (Jones & Marsden, 2006, pp. 280-286) such as how a mobile device's usage should fit with other devices, physical resources, network availability and other context sources should be considered. For example, where processing power is not sufficient on the mobile device which captured a photo, uploading to a server which can index the photo more quickly and then send the result back to the mobile device could be a better solution. It is often more convenient to enter long textual descriptions for a photo using a desktop PC when the user returns home, while a short annotation at the time of capture might be still useful for facts that could have otherwise been forgotten by the time the user returns home. Viewing slide shows of photos at a gathering of relatives may be best served with a TV screen rather than viewing on the mobile device.

Depending on how a particular service or product has been designed in its use of resources in the chain of activities from capturing to storing to searching to sharing with friends, the user may need to interact with different types of interfaces, or this could be to some extent transparent and hidden from the user. For example, photos stored on a remote server or on somebody else's mobile device (in the case of peer-to-peer architecture in resource allocation) could be downloaded to the user's mobile device in the background while the user is browsing photos without them having to use a separate interface to download or browse photos stored on the server. In either case, the often quoted problems of user-interface design for mobile devices seem to remain true for personal photo management. These include details such as;

- Limited screen space
- Limited input mechanism especially awkward text input
- Potentially distracting usage environments

Many interaction and visualization related issues are raised due to these limitations. Unfortunately, we have to live with them because these natural limitations arise from the fact that most mobile devices need to be, by definition, small and mobile. Studies on Web page searching on mobile devices (Jones et al., 1999; Jones et al., 2003), though not specifically on photo searching, form a useful starting point for searching within a mobile photo management context; ideas for visualization on mobile devices have been proposed, especially for displaying interactive maps for location-based navigation (Chittaro, 2006).

The minimal attention user interface (MAUI) (Pascoe et al., 2000) tries to design a mobile interface that requires minimal user attention. This is especially intended to assist field workers who use a PDA during their physically demanding tasks. In a similar vein, use of "push" technology has been proposed to reduce the amount of user interaction on a mobile device by shifting the user's interaction burden to background processing by the device (or the server the device is connected to). For example, if the device can predict which photos the user wants to see on the mobile screen at this moment, the system can display those photos without the user needing to make the effort to enter the search query, possibly delivering the relevant photos as soon as the user turns on the device. How to accurately predict which photos the user will want to view at a given time is of course the main challenge of such an approach. Examples of such technologies include the use of collaborative filtering with the data collected from explicit preference indications (Gurrin et al., 2003), and use of attention data collected from the user's daily web browser log data (Gurrin et al., 2006). The key point of this approach is to reduce the frequency and amount of the user's interactions with the device. We will see an example of how a small number of selective personal photos are displayed before a user's query input on the first browsing screen in the next section.

The following are some of the photo browsing techniques that could be suitable for a mobile interface:

- **Thumbnail browsing:** Spatially presenting multiple miniaturized photos allows easy browsing that leverages the efficient human visual system, and has been used in almost all desktop photo management systems. Although the screen is much smaller on a mobile device, it is still a useful technique and widely used in mobile interfaces for photo browsing.
- **Smart thumbnail view** (Wang et al., 2003)**:** A photo usually contains a main focus of interest (for example, the face of a friend) as well as unnecessary visual elements (for example, strangers in the background or a large background area). By automatically determining the "regions of interest" within a photo, the interface can crop the photo to show only the area that is pertinent to the viewer. By identifying multiple regions of interest in a photo, the interface can guide the user, automatically moving the view window over a photo from one region to another in the order of importance of the regions.
- **Rapid, serial, visual presentation (RSVP)** (De Bruijn & Spence, 2000)**:** Temporally presenting multiple photos one by one as in a slide show seems particularly suitable for a small screen (De Bruijn, Spence & Chong, 2002), although having to keep focusing on the flipping-through of the images requires continuous user attention, and is thus a disadvantage when using a device with such an interface if the user needs to check their environment frequently (for example, while walking or waiting for a bus).
- **Speed Dependent automatic zooming (SDAZ):** When scrolling a page, the photos become smaller (zoomed-out) showing more of them, while it is zoomed-in when scrolling speed is reduced or ceases. This technique attempts to use the context of screen browsing. Some variations of this idea have been evaluated on a mobile device with promising results (Patel et al., 2004).
- **Key photo selection:** When multiple photos need to be displayed on a small screen, the system can determine one of those photos that is most representative of the photos and simply show the one chosen photo (while indicating that there are more to be viewed if desired). In this way, the screen space is saved and the user can browse more photos one by one if they wish to. Of course, selecting one representative photo from a group of photos taken at a particular event is an interesting research question in itself.

Some of these techniques were originally developed for desktop interfaces while others originated in PDA interfaces. These techniques and variants of them are currently being investigated for mobile interfaces. We can expect to see some of these techniques appearing in mobile photo management applications in the near future. Two strong features for enhancing photo management on mobile devices come from the technology camp, and are the subject of the remainder of this chapter. These are:

- **Content-based image analysis:** Computer vision techniques can be used to analyze the image content to classify, label, or identify something meaningful in photos for searching and browsing.
- **Context-awareness:** Context such as time, location (from GPS) and people present at the time of capture can be recorded and used to enhance metadata for searching and browsing.

As we will see in the following section, leveraging these technical elements can significantly enhance mobile interaction for photo management by enriching metadata. They can also be used subsequently to derive other useful metadata, which can in turn reduce the user's photo organization and annotation effort, and possibly enable the use of simple yet powerful time- and map-based interfaces suitable for mobile devices.

ENHANCING INTERACTION FOR MOBILE PHOTO MANAGEMENT WITH CONTENT AND CONTEXT

Content-Based Image Analysis

The use of content-based image analysis techniques for indexing and retrieving images has been an active area of research in the field of computer vision and information retrieval for many years and is neatly summarized in Smeulders et al., (2000), although even in the intervening years there have been further developments. Current approaches to content-based image retrieval can broadly be divided into three different approaches, namely using low-level features, using high-level semantic features and using segmented objects, which we now describe in turn.

Analyzing visual features of an image into low-level features such as color, shape, and texture has been the major building block for indexing image databases in order to classify and retrieve images in terms of their visual characteristics. This approach can be characterized as computationally efficient and undemanding, since these image features can be identified directly from the encoded (compressed) form of the images. Similarity between the low-level features of images can be computed simply based on intersecting histograms representing color or texture bands, where these histograms are derived for the entire image or for regions within the image. While they are computationally efficient and scalable to replicate, low-level representations of images most often do not correspond to high-level, semantic concepts that humans use when we see images. We can say that color, texture and shape are a crude first approximation to semantic image content, but very often they do not satisfy our requirement for recognizing, understanding, searching and browsing images. This difference between what low-level features offer, and what users require, is known in the literature as the "semantic gap," and has been a difficult research problem to tackle (Har et al., 2006; Smeulders et al., 2000).

The second general approach to image retrieval addresses the semantic gap head-on by trying to automatically detect semantic units directly from image content. Such semantic concepts can include almost anything, but generic concepts such as faces, buildings, indoor/outdoor, and landscape/cityscape are often used because they give general applicability. The set of possible semantic features we could detect is influenced by the use that individual detected features can offer, and as this is mostly in image classification and image retrieval, the set of possible features which we *could* calculate is enormous. In order to provide some structure and to limit the set of semantic features to use in image retrieval, we usually arrange features into an ontology which is a hierarchical arrangement of semantic topics, like the LSCOM ontology (Naphade et al., 2006). The LSCOM ontology has just under 1,000 concepts taken from the domain of broadcast TV news, but most of these concepts could be applied to any visual media, including personal photos.

The main challenge with using semantic features in applications such as personal photo management is in building classifiers to automatically detect the features. Semantic features are usually detected based on an analysis of low-level features like color, texture and shape. They are usually constructed by using a machine learning algorithm to *learn* the presence and absence of features associated with individual semantic concepts based on some training set. In the early days of using semantic features where the number of features was of the order of dozens, this was a scalable approach, but as we move towards detecting several hundred features or more, then the approach of building and training individual feature detectors does not scale, and this is one of the main challenges facing the field currently.

A second major challenge in automatically detecting features is improving the accuracy and reliability of the feature detection. Performance assessment of feature detection is carried out as part of the annual TRECVid evaluation benchmarking campaign[7] in which many (70+) participating research groups from around the world benchmark the performance of their systems for automatically detecting high-level concepts appearing in video sequences. In particular, and what makes this activity relevant to content-based image retrieval tasks such as photo managements, is that TRECVid feature detection is mostly based on shot keyframes, which are still images taken from within video shots[8]. At TRECVid 2006 benchmarking the performance of only 39 feature detectors including the presence of buildings, desert, roads, faces, animals, airplanes, cars, explosions and so on, was a significant activity for the participating groups. It is believed that building semantic feature detectors which depend upon each other, in the same way that concepts in the LSCOM ontology are arranged in a hierarchical dependency, will lead to improved feature detection accuracy (Naphade et al., 2002; Wu et al., 2004).

The final approach to content-based image retrieval that we will mention is to detect, and then use, *objects* that appear in an image as the basis for retrieval. In the approaches described so far, the processing is done on the entire image whereas in this approach we seek to identify and segment the major objects that appear within an image and to use them, rather than the whole frame, for retrieval. For example if we seek to find photos of boats then we can use a segmented image of a boat object taken from an image, independent of the background, and retrieve other objects from a photo collection based on their color, texture and/or shape. As an example of this, Sav et al. (2006) describe a system to allow manual segmentation of semantic objects from query images which are then matched against segmented objects in database images. A similar approach, albeit applied to video rather than to image retrieval, is reported by Sivic et al. (2006) used for the Google Video Search Engine[9]. What these, and a number of object-based retrieval applications which are experimental in nature, have in common is that they use segmented objects as the basis for retrieval, yet the task of automatic or semi-automatic image segmentation remains one of the most challenging image processing tasks and represents a significant hurdle towards making object-based image retrieval more widespread.

Applying content-based analysis methods of these types can enable the use of photos as queries to search for similar ones or objects within one photo to find similar objects in other photos. However, as is made clear from the above, these technologies are either rather unreliable, since they lack the power to capture perceived semantic features because they are too low-level, such as the color-based features, or they are rather specialized, in the form of learned features of particular objects. These methods thus have considerable possible utility, but are not, at present at any rate, suitable for robust and reliable photo management, but they do offer interesting potential when used in combination with context features associated with photo capture as explored in the next section.

Context-Awareness

Context information recorded at the time of photo capture can be used to assist with photo management on a mobile device in a number of ways. Particularly important given the high volume of photos often taken with devices such as camera phones and the interaction issues reviewed earlier in this chapter, a key feature in the use of the context of photo capture is that it can generally be used entirely automatically. In this section we examine the following easily captured and used context features: time, location, lighting levels and weather.

Time

Chronology is one of the most important clues when a user is looking for photos (Rodden, 2002). We know that users often remember at least the rough date/time of photo capture, even if they cannot remember exact details. Digital cameras routinely record the time of photo capture in the EXIF header of photos captured. This data allows photos to be indexed using fields such as year, month, day of the month, day of the week, hour of the day. In addition it is possible to derive more descriptive fields, such as season, weekday, or weekend, which will further aid user interactions. Indexing by time along multiple dimensions like this is useful when the user only remembers certain facets of the temporal context surrounding photo capture. For example, they may remember only that a photo was taken in the summer, in the evening, on a certain day of the week, or on the weekend.

Location

The integration of a location capture device and a camera provides the ideal scenario for location stamping of digital photo collections. However, at this point consumer digital cameras do not have integrated location stamping capabilities. While awaiting the arrival to the market of cameras which incorporate this capability, it is possible to utilize a separate GPS device to record the location at which photos are taken, and then via a timestamp matching process, incorporate the location of photo capture into the EXIF header. In our own work, we capture the locations using a small portable GPS device tracklog stored every 10 seconds, and utilize this tracklog in the location stamping process. In order to map raw GPS coordinates to real world locations, we utilize a gazetteer which typically allows the indexing of each photo at three separate levels: country, city and state, and town. The level and accuracy of location stamping depends on the granularity of the available gazetteers.

The key benefits of labeling digital photos with their location are that it enables us to support a number of access methodologies: search by actual location (country, city,town, even street), search by proximity to a location, or by proximity to other photos. By using such information the browsing space (number of photos that a user has to browse through) when seeking a particular photo can be drastically reduced.

In addition, it is possible to present a user with a map-based interface to their photo collection, with photos, or icons, plotted on a map. For example the Microsoft WWMX system (Toyama et al., 2003) takes this approach, while Google Maps[10] allows its map-specific APIs to be easily incorporated into a Web-based personal photo application thus saving development effort.

Previous research (Gurrin et al., 2005) shows that the integration of location context into a time-context based system reduces mean time to locate a given photo within an experimental collection of 8,000 photos from 32 seconds to 18 seconds, and reduces the mean number of query iterations required to locate the given photo from 3.7 to 2.8.

Other Context Issues

However powerful time and location are individually at supporting user search of digital photo collections, by combining these two contextual features, one can derive additional contextual features, such as *lighting levels* and *weather*. Standard astronomical algorithms (Meeus, 1999) allow us to calculate the environmental lighting level at the time and location of photo capture. A photo taken at 10 a.m. will be in daylight in most parts of the world, but this is not always the case, for example, in parts of Scandinavia and similar high-latitude locations this time could signify dawn, or even darkness, depending on the time of year. We use astronomical algorithms to calculate sunrise and sunset times for any location on any date, and using these algorithms we can associate a daylight status (daylight, darkness, dawn or dusk) with each photo based on its time and GPS location of capture, and thereby automatically annotate each photo with this information. When searching for a photo it is probably more likely that a user will remember that it was dark when a particular photo was being taken, than the exact time that they took the picture.

Another feature that can be used to annotate each photo is the prevailing weather conditions. There are 10,500 international weather stations dotted all across the globe which log weather data a number of times each day. Given this information, and readily available access to the weather data logs via the Web, one can annotate each photo with the weather data (clear, cloudy, rainy, or snowy) from the closest international weather station at the time the photo was taken.

Finally, people present at the time of photo capture could be yet another potentially useful context that can be captured. By using a Bluetooth device, people nearby who have Bluetooth-enabled devices can be picked up and recorded, and this information can complement other methods such as face recognition (Davis et al., 2005) effectively combining context with content-based techniques.

Content-based analysis and context-awareness as discussed so far can be applied to user access to photo collections via a mobile device to significantly enhance the user interaction on such a device. The next section introduces a prototype of such a system under development in our laboratory.

Mobile Photo Access: An Example

The MediAssist mobile interface (Gurrin et al., 2005) to personal digital photo libraries has been designed to minimize user input and proactively recommend photos to the user. Consequently, it supports the following three access methodologies from a mobile device:

- ‘**My Favorites’:** The first screen a user sees when accessing their archive using a mobile device (see Figure 2a) is a personalized thumbnail listing of the top 10 most popular photos based on a user’s history of viewing full-size photos, where this history data is gathered both from mobile devices and conventional desktop device access.
- **Search functionality:** Primarily based on location and the derived annotations. The aim is to reduce the level of user interaction required to quickly locate relevant content. In order to maximize screen real-estate available for browsing the photo archives, search options are hidden in a panel that slides into view when a user wants to search (see Figure 2b) and then disappears afterwards until required again (Figure 2a).
- **Browsing the collection by events:** Even by supporting the two access methods on a mobile device a user may still end up having to spend time scrolling through screens of photos if many were taken at the same time and place. To address this issue, the interface presents results to the user clustered into *events* and ordered by date and time. Events are logical combinations of photos taken in close proximity of location and time. Event clustering of photos can either be a rule-based process (e.g., no photos taken for a period of 90 minutes signifies the end of an event), or a clustering process where photos are grouped together based on location and/or time and the unique clusters extracted to comprise events in a personal photo collection.

A user accessing the photo archive is immediately presented with the ‘My Favorites’ screen, of their most accessed photos, helping to reduce user interaction. If the required photo is not in the favorites, the user engages in a process of searching, followed by browsing of the search results, so in effect it is a two-phase search. The search options are: three level location (country, state, and city or town), season, weather, and lighting status, as shown on the sliding panel in Figure 2b. The contents of the location drop-down boxes are personalized to the user’s collection to minimize user input. Season, weather, and lighting status are included to filter the search results thereby reducing the amount of browsing effort required to locate the desired photos.

Figure 2. MediAssist mobile interface takes advantage of context information to automatically organize personal photos

(a) 'My favorite photos' with search panel down

(b) Searching the archive with a representative photo results as event summary

Simply presenting a (potentially long) list of photo thumbnails in response to a query is not an ideal interaction scenario for the user of a mobile device. A more 'mobile friendly' technique is to group photos into events and using a single 'key' thumbnail which represents an event, as shown in Figure 2b (one event displayed). This is done by automatically grouping photos together into logical sets by examining when and where clusters of photos co-occur and choosing a single representative photo to represent the whole cluster. Typically the photo chosen to represent the cluster is the middle photo from a temporally organized listing of the cluster photos. In future work we will focus on judiciously choosing the most representative photo in a query-biased manner taking account of context and content data associated with photos in the event. These clusters are then presented to the user, ordered by time and date. Tapping on a thumbnail photo on screen presents the user with a full-screen photo, and it is this detailed viewing of a photo that is used to support the 'My Favorites' access method. Associated with each thumbnail is a small arrow button on the right side of the thumbnail. Tapping on this arrow brings the user to a screen showing all photos from that particular event, once again organized by date and time. In an experiment, the broad context searching capabilities of the mobile MediAssist system were shown to clearly outperform a more conventional time-only based system (Gurrin et al., 2005).

CONCLUSION

Interesting avenues for application scenarios are already appearing in literature which leverage context and/or content analysis for mobile photo annotation and searching. A mobile photo management system (Sarvas et al., 2004) records location, time and user data at the time of photo capture and then compares this with other already annotated metadata from other users and presents an inferred annotation for the new photo to the user. The Photo-to-Search system (Fan et al., 2005) allows a user to take a photo with a camera phone, the system then searches for visually similar images from the web and returns the result on the mobile device. While these make interesting applications and their evaluation with users will be highly important for developing future mobile systems, we need further application ideas and testing in other tasks of photo management to be able to explore more diverse application possibilities and new kinds of functions.

As we have seen in this chapter, as the number of digital photos each person needs to manage in their collection continues to grow, it will be inevitable that some form of automatic management is used as an integral module of photo management systems to help the user cope with the number of photos, even if running only at the background. Leveraging context data at the time of photo capture and use of steadily improving content-based image analysis will form a crucial part in making automatic organization of personal photos feasible. The role of mobile devices such as camera phones in this application area is also growing very rapidly as such devices become more powerful technically and more ubiquitous and more accepted socially. However, a mobile device will not, in itself, be designed to do every task of photo management. They will be designed to be optimally used in conjunction with other devices such as desktop PCs and laptops, TVs, and other information appliances, depending on technical and social situations. Automatic organization, then, is a vital back-end technology in the chain of personal photo management tasks of which mobile interaction is a part.

ACKNOWLEDGMENT

We gratefully acknowledge support from the aceMedia Project under contract FP6-001765, Science Foundation Ireland under grant number 03/IN.3/I361, and Enterprise Ireland under grant number CFTD-03-216.

REFERENCES

Chittaro, L. (2006). Visualizing information on mobile devices. *IEEE Computer*, *39*(3), 40–45.

Davis, M., Smith, M., Canny, J., Good, N., King, S., & Janakiraman, R. (2005). Towards context-aware face recognition. In *Proceedings of the 13th Annual ACM International Conference on Multimedia '05* (pp. 483-486). Singapore

De Bruijn, O., & Spence, R. (2000). Rapid Serial Visual Presentation: a Space-Time Trade-off in Information Presentation. In *Proceedings of the Working Conference on Advanced Visual Interfaces AVI '00* (pp.189-192). Palermo, Italy.

De Bruijn, O. Spence, R., & Chong, M. Y. (2002). RSVP browser: Web browsing on small screen devices. *Personal and Ubiquitous Computing*, *6*(4), 245-252.

Enser, P. (1995). Pictorial information retrieval. *Journal of Documentation, 51*(2), 126-170.

Fan, X., Xie, X., Li, Z., Li, M., & Ma, W.-Y. (2005). Photo-to-search: Using multimodal queries to search the web from mobile devices. In *MIR '05 - Proceedings of the 7th ACM SIGMM International Workshop on Multimedia Information Retrieval* (pp. 143–150). Singapore.

Frohlich, D., Kuchinsky, A., Pering, C., Don, A., & Ariss, S. (2002). Requirements for photoware. In *CSCW'02 -Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work* (pp. 166–175). New Orleans, LA.

Gurrin, C., Brenna, L., Zagorodnov, D., Lee, H., Smeaton, A. F., & Jahansen, D. (2006). Supporting mobile access to digital video archives without requiring user queries. In *MobileHCI '06 - Proceedings of the 8th International Conference on Human-Computer Interaction with Mobile Devices and Services* (pp. 165-168). Espoo, Finland.

Gurrin, C., Jones, G., Lee, H., O'Hare, N., Smeaton, A. F., & Murphy., N. (2005). Mobile Access to Personal Digital Photograph Archives. In *MobileHCI'05 - Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices and Services* (pp. 311-314). Salzburg, Austria.

Gurrin, C., Smeaton, A. F., Lee, H., Donald, K. M., Murphy, N., O'Connor, N., & Marlow, S. (2003). Mobile access to the Físchlár-news archive. In F. Crestani, M. Dunlop, & S. Mizzaro (Eds.), In *Mo-*

bileHCI'03 – Proceedings of the 5th International Symposium on Human Computer Interaction with Mobile Devices and Services, Workshop on Mobile and Ubiquitous Information Access (LNCS Vol. 2954) (pp. 124-142), Berlin/Heidelberg, Germany: Springer.

Har, J. S., Lewis, P. H., Enser, P., & Sandom, C. J. (2006, January15-19). Mind the Gap: Another Look at the Problem of the Semantic Gap in Image Retrieval. In *Proceedings of the SPIE Multimedia Content Analysis, Management, and Retrieval, 6073(1).* San Jose, CA.

Jones, M., Buchanan, G., & Thimbleby, H. (2003). Improving web search on small screen devices. *Interacting with Computers*, *15*(4), 479-495.

Jones, M., & Marsden, G. (2006). *Mobile interaction design*. Wiley.

Jones, M., Marsden, G., Mohd-Nasir, N., & Boone, K. (1999). Improving Web Interaction on Small Displays. In *Proceedings of the 8th World Wide Web Conference* (pp. 1129-1137). Toronto, Canada.

Kindberg, T., Spasojevic, M., Fleck, R., & Sellen, A. (2005). The ubiquitous camera: An in-depth study of camera phone use. *IEEE Pervasive Computing*, *4*(2), 42-50.

Kustanowitz, J., & Shneiderman, B. (2004). Motivating Annotation for Digital Photographs: Lowering Barriers While Raising Incentives. In *Technical Report, HCIL-2004-18, CS-TR-4656, ISR-TR-2005-55, HCIL University of Maryland, 2004.*

Meeus, J. (1999). *Astronomical Algorithms, 2nd ed.* Willmann-Bell.

Naphade, M., Kozintsev, I., & Huang, T. (2002). A factor graph framework for semantic video indexing. *IEEE Transactions on Circuits and Systems for Video Technology*, *12*(1), 40-52.

Naphade, M., Smith, J. R., Tesic, J., Chang, S.-F., Hsu, W., Kennedy, L. et al.(2006). Large-scale concept ontology for multimedia. *IEEE Multimedia*, *13*(2), 86-91.

O'Hare, N., Lee, H., Cooray, S., Gurrin, C., Jones, G., & Malobabic, J. (2006). MediAssist: Using content-based analysis and context to manage personal photo collections. In H. Sundaram, M. R. Naphade, J. R. Smith, & Y. Rui: (Eds.), *CIVR2006 - 5th International Conference on Image and Video Retrieval* (Vol. LNCS 4071) (pp. 529-532). Tempe, AZ.

Palen, L., & Salzman, M. (2002). Beyond the handset: Designing for wireless communications usability. *ACM Transactions on Computer-Human Interaction*, *9*(2), 125–151.

Pascoe, J., Ryan, N., & Morse, D. (2000). Using while moving: HCI issues in fieldwork environments. *ACM Transactions on Computer-Human Interaction*, *7*(3), 417–437.

Patel, D., Marsden, G., Jones, S., & Jones, M. (2004). An Evaluation of Techniques for Browsing Photograph Collections on Small Displays. Lecture Notes in Computer Science (LNCS). In *MobileHCI'04 - Proceedings of the 6th International Conference on Human-Computer Interaction with Mobile Devices and Services* (pp. 132-143). Glasgow, Scotland.

Perry, M., O'hara, K., Sellen, A., Brown, B., & Harper, R. (2001). Dealing with mobility: Understanding access anytime, anywhere. *ACM Transactions on Computer-Human Interaction*, *8*(4), 323–347.

Rodden, K. (1999). How do people organise their photographs? In *IRSG'99 - Proceedings of the BCS IRSG 21st Annual Colloquium on Information Retrieval Research,* Glasgow, Scotland.

Rodden, K. (2002). Evaluating Similarity-based Visualisations as Interfaces for Image Browsing. In *University of Cambridge Technical Report, UCAM-CL-TR-543.*

Rodden, K., & Wood, K. (2003). How do people manage their digital photographs? In *CHI'03 - Proceedings of CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*, (pp. 409-416). Fort Lauderdale.

Sarvas, R., Herrarte, E., Wilhelm, A., & Davis, M. (2004). Metadata Creation System for Mobile Images. In *MobiSys '04 - Proceedings of the 2nd International Conference on Mobile Systems, Applications, and Services* (pp. 36–48). Boston, MA.

Sarvas, R., Oulasvirta, A. and Jacucci, G. (2005). Building social discourse around mobile photos: a systematic perspective. In *MobileHCI' 05 – Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices and Services* (pp. 31-38). Salzburg, Austria

Sav, S., J. F. Jones, G., Lee, H., O'Connor, N. E., & Smeaton, A. F. (2006). Interactive experiments in object-based retrieval. In H. Sundaram, M. R. Naphade, J. R. Smith, & Y. Rui: (Eds.), *CIVR 2006 – Proceedings of the 5th International Conference on Image and Video Retrieval* (LNCS 4071). Tempe, AZ.

Sivic, J., Schaffalitzky, F., & Zisserman, A. (2006). Object level grouping for video shots. *International Journal of Computer Vision*, *67*(2), 189–210.

Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A., & Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(12), 1349–1380.

Toyama, K., Logan, R., Roseway, A., & Anandan, P. (2003). Geographic location tags on digital images. In *MULTIMEDIA'03 - Proceedings of the 11th ACM International Conference on Multimedia* (pp. 156-166). Berkeley, CA.

Van House, N., Davis, M., Ames, M., Finn, M., & Viswanathan, V. (2005). The uses of personal networked digital imaging: an empirical study of cameraphone photos and sharing. In *CHI'05 - Proceedings of the CHI 2005 Extended Abstracts on Human Factors in Computing Systems* (pp. 1853-1856). Portland, OR.

Wang, M.-Y., Xie, X., Ma, W.-Y., & Zhang, H.-J. (2003). MobiPicture: Browsing Pictures on Mobile Devices. In *MULTIMEDIA'03 - Proceedings of the 11th ACM International Conference on Multimedia* (pp.106–107). Berkeley, CA.

Wilhelm, A., Takhteyev, Y., Sarvas, R., House, N. V., & Davis, M. (2004). Photo annotation on a camera phone. In *CHI'04 Extended Abstracts on Human Factors in Computing Systems* (pp.1403-1406). Vienna, Austria.

Wu, Y., Tseng, B., & Smith, J.R. (2004). Ontology-based multi-classification learning for video concept detection. In *ICME 2004 - Proceedings of the IEEE International Conference on Multimedia and Expo* (pp. 1003-1006). Taipei, Taiwan.

Zhang, L., Hu, Y., Li, M., Ma, W., & Zhang, H. (2004). Efficient propagation for face annotation in family albums. In *MULTIMEDIA'04 - Proceedings of the 12th annual ACM international conference on Multimedia* (pp. 716-723). New York, NY.

KEY TERMS

Blog: A type of website in which the user adds regular written contributions on his/her own life or thoughts, as in a journal or diary. Contracted from *weblog,* usually the entries are in reverse chronological order, and readers are allowed to add their own comments.

Context-Awareness: A system that can use information about the circumstances under which it is being used. For example a context-aware device will use the current time and location where it is being used to infer what would be the most beneficial piece of information to display for the user.

Content-Based Image Retrieval (CBIR): An application of computer vision to image retrieval, in which an image's content (its color, texture, shapes, objects or faces in it, etc.) is automatically analyzed to index the image for subsequent retrieval.

Information Retrieval (IR): An interdisciplinary field of study that deals with searching for information in documents (papers, books, pictures, video clips, or any other item that contain useful information). IR systems seek to return to users documents which satisfy their current information need as expressed through some form of search request which may comprise components in one or more media.

Global Positioning System (GPS): A satellite navigation system in which more than two dozen satellites broadcast precise timing signals by radio, allowing any GPS receiver device to accurately determine its location.

Photoware: A software application used for personal photo management. Although the term emerged when online sharing of photos became common in personal photo management software, in this chapter we use the term in a more general sense.

Web 2.0: The second generation of Internet-based services in which the Web itself is a platform for users to directly use and share information on the Web, often characterized by its highly-dynamic and highly-interactive Web interfaces and pulling together the distributed resources from independent developers of contents.

ENDNOTES

1. Flickr. Available at http://www.flickr.com/ (Retrieved January 2007)
2. Adobe Photoshop Album. Available at http://www.adobe.com/products/photoshopalbum/starter.html (Retrieved January 2007)
3. ACDSee. Available at http://www.acdsee-guide.com/ (Retrieved January 2007)
4. Google Picasa. Available at http://picasa.google.com/ (Retrieved January 2007)
5. Flickr. Available at http://www.flickr.com/ (Retrieved January 2007)
6. Yahoo! 360. Available at http://360.yahoo.com/ (Retrieved January 2007)
7. TRECVid 2006 Guideline. Available at: http://www-nlpir.nist.gov/projects/tv2006/tv2006.html (Retrieved January 2007)
8. A “video shot” is an unbroken sequence of frames taken by a single camera. Shot boundaries occur at camera changes.
9. The Google Video Search Engine. Available at: http://video.google.com/ (Retrieved January 2007)
10. Google Maps. Available at http://maps.google.com/ (Retrieved January 2007)