

Snap and Share Your Photobooks

Niels Henze
OFFIS - Institute for Information Technology
Escherweg 2
Oldenburg, Germany
henze@offis.de

Susanne Boll
University of Oldenburg
Escherweg 2
Oldenburg, Germany
susanne.boll@uni-oldenburg.de

ABSTRACT

The sharing of photos with others, friends and family, has become even more popular with digital photography and Internet applications such as email or Web albums. At the same time, the physical touch of printed photos is still appreciated and customers use different services to print their photos on post cards, calendars or photobooks, often to give them as a present or to create a physical souvenir. Once printed, however, the sharing of photos with others becomes difficult as there is no link back from the now physical item to its digital counterpart. With Bookmarkr, we developed a system that employs a mobile camera phone to bridge the gap between the printed photo and its digital counterpart. The user takes an image of a photo in a photobook with the mobile phone's camera. The image is transmitted to a photobook server, which employs image analysis techniques to retrieve the corresponding photo. The photo is sent back to the user and establishes the digital-physical match. Bookmarkr allows the user to interact with printed photos in a similar way as interacting with digital photos by using a point-and-shoot metaphor. Our performance evaluation shows that the system is able to return the correct photo in up to 99% of all cases. A conducted user study revealed that the designed interaction is suitable for the participants and their printed photobooks.

Categories and Subject Descriptors

H.5.1 [Multimedia Information Systems]: Artificial, augmented, and virtual realities; H.5.2 [Interfaces and Presentation]: User Interfaces - Interaction styles

General Terms

Algorithms, Design, Human Factors, Experimentation

Keywords

photo sharing, mobile interaction, image analysis, photobook, contextual bookmark

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MM'08, October 26–31, 2008, Vancouver, British Columbia, Canada.

Copyright 2008 ACM 978-1-60558-303-7/08/10 ...\$5.00.

1. INTRODUCTION

In the last hundred years consumer photography became a widely used tool to document events and share personal memories. Before emergence of digital photography most photos had to be printed to paper or reversal film. With the advent of digital photography the way people use photos changed dramatically. The number of photos taken increased enormously and left the user with the problem of how to manage these collections and which ones to share with friends using which media. One can observe that after a few years in which consumers mostly shared their photos digitally by uploading them to Web albums or sending them by mail; the actual physical print is again in fashion. Europe's largest photofinisher CeWe Color¹, for example, produced 2.8 billion colour prints in 2007 in total. On the edge from analogue to digital photography, the amount of digital photos printed by CeWe Color increased by 35.9% to 1.5 billion digital photos in 2007 [8].

From the market development it can be observed that digital photography has not replaced printed photos. Rather physical photo products complemented digital photos. Many value added services evolved that allow customers to print their digital photos on postcards, calendars, mugs and many more. A very specific and successful product is the photobook. Supported by authoring software users can arrange their photos and annotations over the pages of a photobook and have it printed by a provider such as CeWe Color. Such photobooks, as illustrated in Figure 1 are a convenient way to preserve events. The photobook market rapidly increased in the last years. Sales figures of photobooks produced by CeWe Color, for example, exceeded one million photobooks in 2007 [7].

¹<http://www.cewecolor.de>



Figure 1: Digitally produced photobook.

Photo sharing online community sites such as Flickr² that hosted over two billion photos [1] in November 2007 are another rapidly increasing market. Exchanging photos with friends and colleagues with only minimal cost is a distinctive feature of digital photos compared to printed photos. To exchange a printed photo, its physical reproduction had to be produced for every recipient. By uploading digital photos to an online community site photos can be shared with a virtually unlimited number of people. However, once printed to paper exchanging and processing the now analogue photos easily is not longer possible. There is no connection between the printed photo and its digital counterpart. Users have to search their digital photo collection manually based on the analogue printout which might be a time consuming task especially for inexperienced users. What is missing is a link that connects the analogue printout with its digital representation.

We developed a system, Bookmarkr, which uses printed photobooks as anchors to their digital representation. The Bookmarkr enables users to retrieve a high resolution digital photo based on a snapshot from a printed photobook using a mobile camera phone. The user selects a photo in a photobook by shooting an image of the respective photo. The image is transmitted to a photobook server that analyses the image and compares it to the digital sources of the photobook. Using image analysis techniques the corresponding digital photo is retrieved and transmitted back to the user's mobile phone. Thus, users can snap a photo to use and share it.

In Section 2, we present related work in the field of photowork, interaction with photo collections, and computer supported interaction with the real world. The concept of the Bookmarkr system and how it supports the tasks of accessing digital photos is described in Section 3, followed by an overview of the system design and implementation in Section 4. We present a performance test of the technical solution and the user evaluation of the system in Section 5. We close this paper with a conclusion and outlook to future work in Section 6.

2. RELATED WORK

Personal photo collections in general have received a great share of attention in the research fields of multimedia and human computer interaction. Most attention has been given to the question of how people create and manage their photos (e.g. [31, 29, 22]). Since we aim to support (co-located) sharing of photos we focus on this aspect in the following.

An investigation of how people use printed and digital photos has been carried out by Frohlich et al. [14]. They derived requirements for "Photoware" (tools that support tasks performed with photos) from interviews with eleven families and emphasise the necessity of tools for remote and co-located photo sharing. Crabtree et al. [10] carried out another study of printed photo use and identified an aversion amongst many users towards collaborative sharing through the use of existing digital technologies. They report that people tend to print photos in order to share them with others co-presently. Both Crabtree and Frohlich highlight that co-located photo sharing is a highly communicative process. A survey from the Photo Marketing Association [30] shows that the main reasons for taking pictures with a digital cam-

era are to preserve memories and to share photos with others. Miller and Edwards conducted a study of the digital photo-sharing site Flickr [26]. They identified two user categories the "Kodak Culture" people that communicate primarily within their existing social group and the "Snaps" representing a hybrid of traditional amateur photography, Kodak Culture, and blogging.

Some systems have been developed that support the sharing of photos. Balabanovic et al. [2] present a handheld device that aims to offer some of the advantages of printed and digital photos by enabling recording and playback of stories. Hilliges et al. developed a system to interact with photo collection in an instrumented environment [17]. Counts and Fellheimer designed and evaluated a mobile photo sharing tool based on Pocket PCs [9]. Likewise, Van House et al. investigated peoples use of a camera phone application that automatically uploads taken photos to a website [18]. Nunes et al. mention that within the home digital photos are hard to access and lack the physical benefits that makes sharing easy and opportunistic [28]. Consequently they developed a system that links digital photo sets to physical memorabilia.

To not completely displace the physical artefacts, such as printed photos, by their digital counterpart several approaches to enhance the interaction with physical artefacts exists. These can be distinguished in direct interactions such as touching or pointing and indirect interactions such as using the mobile phone as a remote control (e.g. [3, 34]). Fitzmaurice was one of the first who predicted the use of mobile devices for pointing based interactions. He described for instance an application with which the user could point onto certain locations on a map in order to get additional information [13]. In recent years more systems emerged that provide information related to physical objects. Mobile devices are used more and more for this kind of interaction. A common way to implement such systems is to mark objects either with visual markers [32] or with electronic tags [36]. Barcodes can be seen as the first implementation of visual markers back in 1948. Since special barcode readers are needed to read barcodes traditional barcodes are not used by common consumers. Thus, 2D barcodes such as QR codes [20] have been developed that can be read using the latest mobile phones. Some of the newer mobile phones also have integrated RFID reader that can be used to access related digital information and services.

Another approach is using digital images of objects to find the corresponding digital information using content based image analysis. Liu et al. use the camera of a mobile phone to select documents displayed on computer screens [24]. Erol and Hull use mobile phones' cameras to enable the user to select presentation slides by taking images of the slides [11]. Fan et al. developed a multimodal search engine which takes a photo and a textual query to find related images and web pages [12]. Finally, Zhou et al. developed a system to acquire information related to sights by taking a photo [37] and Henze et al. developed a system to access services related to posters and public displays using image and context analysis [15].

3. DESIGN OF THE BOOKMARKR

In the following we present the design of Bookmarkr a system that links photos in printed photobooks to their digital counterparts. Users access digital photos by taking an image with a mobile phone. This image is compared to dig-

²<http://www.flickr.com>



Figure 2: User taking an image of a part of a photobook's page using a Nokia N95.

ital photos by a photobook server using local features. The system supports users in providing other people with digital copies of photos without losing the advantages of printed photos. It does not presume any changes of the design or cost of individual photobooks.

Photobook based co-located photo sharing is enhanced by providing easy access to the book's digital sources. We illustrate how Bookmarkr works through the following scenario. Afterwards, the requirements of the system are analyzed and the system's concept is outlined. Finally the necessary matching algorithm is discussed.

3.1 Scenario

Mary has invited her family to celebrate her birthday. After welcoming all guests and receiving birthday gifts Mary digs out a couple of photobooks she produced for recent events such as her trip to Mexico, her son's baptism, and her daughter's school enrolment. A lively discussion about the events emerges between the guests and the photobooks are passed from one to the other. While Mary is chatting about Mexico with her brother Mary's grandparents discuss the baptism in detail.

After a while some guests express the wish to have reprints of some of the photos. Mary's brother Oliver takes out his new mobile phone to access the digital version of the printed photos. He selects the photobook about Mexico by shooting an image of the book's cover page using his mobile phone. Afterwards he starts to discuss with Mary again. Periodically he selects a photo by shooting an image of the respective page (see Figure 2) and his phone receives the corresponding digital version of the selected photo.

Later that day Oliver uses the digital photos collected with his mobile phone to send them to his wife and order reprints from a local photo finisher.

3.2 Requirements analysis

When friends or family are visiting as in the scenario described above we often show photos possibly to tell about the last vacation, the good old times or more generally to share memories. Frohlich, for example, found that photos are mainly used to share memories and that "sharing photos in person was described as the most common and enjoyable"

[14]. In such situations it is best if people can freely explore the photos. They can split into smaller groups focussing on certain photos stories that are of particular interest to them.

Without printing the photos we rely on media players to show our digital photos. Common tools used for that are notebooks, TV sets with DVD player as well as digital photo cameras. Using these devices only one photo can be shown at a time. Someone – possibly the host – has to control the playback device. This hinders free discussions and leads to a situation where the host presents the photos. Guests can not explore the photos themselves. All attendees are forced to follow the photo presentation. The emergence of multiple storylines is not possible in those situations and it is also prevents the group from being able to split up to focus on photos and stories of their particular interest. In addition, it is often difficult to arrange the audience in a way that provides a good view on the playback device for everyone. Especially if notebooks or digital cameras are used the photos usually appear with colour and geometric distortions from most viewing angles. Printed photos, on the other hand, do not have those limitations. They can be viewed from more positions without colour or image distortions. Printed photos can be passed around and can be freely explored. Thus, a group of people can split where each group can focus on a different topic. It does not require everyone to participate in sharing photos.

Often people like to get a copy of a photo while looking at it. Whether digital or printed photos are used this is typically a difficult task. If presenting digital photos via a playback device this means that the presenter has to interrupt the presentation and mark or send the photo. However, using printed photos it is even more difficult since the digital version has to be found first. This involves switching to the computer to search for the photo. In both cases people rely on the owner of the photo to get a copy.

A system as the one used by Oliver in the previous scenario can support exchanging copies of photos while maintaining the advantages and tangibility of printed photos. A prerequisite of the scenario is a suitable interaction device. Personal mobile devices, like camera-equipped mobile phones, open up new ways of interaction with our environment. We are comfortable with using our own mobile devices and usually have them with us. They can connect to online services via diverse connection types such as UMTS or wireless LAN and integrated sensors, in particular, digital cameras provide a powerful input channel. Camera based interaction techniques have been developed for various use cases such as interaction with public displays [4]. An interaction technique that has proven its usefulness to access digital information related to physical objects is the point-and-shoot interaction technique [4]. To select an object, users aim the interaction device at a physical object and press a shoot button much like using a mouse cursor in a GUI to select a widget. The digital information related to the selected object is transferred to the mobile phone where the information can be shown to the user and be shared with others.

To use the point-and-shoot interaction technique for a concrete type of physical objects it is important to understand the nature and usage of the object. We focus on pictures in printed photobooks. Photobooks are used to record events or to assemble photos with some correlation. Even with the support of software tools it is a time consuming

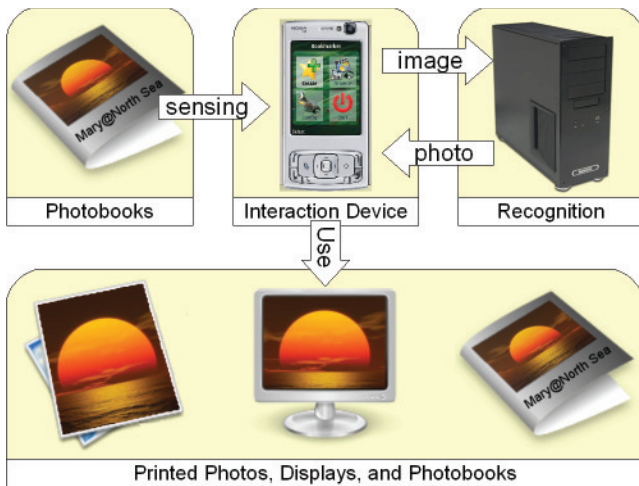


Figure 3: Architecture of the Bookmarkr.

task to produce a photobook and the outcome represents a person's individual impression of the respective topic. Combining photos and choosing an appealing design often involves using artistic abilities. As a result, photobooks are highly individual and the visual appeal is one of the main features.

To implement the point-and-shoot technique it is necessary to determine which photo the user intends to select. Equipping the photos inside the photobook with barcodes (see Section 2) that can be scanned by users' camera phones could provide a link from the physical photo to its digital counterpart. However, due to the visual nature of photobooks it is not reasonable to alter the visual appearance of the photobook in any way. Another solution is to integrate electronic markers such as RFID tags in the photobook for every photo. However, even the short range electronic signals of RFID penetrate a photobooks pages. Thus, it is difficult to clearly discriminate which photo a user selects. In addition, up to hundreds of RFID tags would be needed for each photobook resulting in a reasonable increase of costs.

3.3 Concept

We designed the Bookmarkr system to provide a link from the physical photo to the digital counterpart that does not presume major changes of the design or cost of individual photobooks. Our solution to create the link between digital and physical photo uses content based image analysis. We developed a process (see Figure 3) that enables the user to access printed photos' digital counterpart using a mobile phone application. The user creates an image of the printed photobook using the phone's camera. The image is sent to a photobook server which accesses a photo repository containing all digital photos. This server could, for example, be located at the user's photofinisher company that prints the photos to paper and thus has access to the digital photos. The photobook server retrieves the photographed picture by matching the image taken by the mobile phone with the digital photo. The retrieved digital photo is sent back to the mobile phone.

Designing the system required taking into account the enormous amount of digital photos on photo sharing systems today and the amount of photobooks produced each

year. It can be considered as impossible to retrieve the correct photo in a reasonable amount of time and with sufficient precision for every photo ever created. We therefore envisage a two stage process. The user first selects a photobook before he or she can select photos from the book. We envisage the same interaction technique for selecting a photobook as for selecting individual photos. Photobooks are selected by taking an image of its cover page³. Thus, images taken by the user are compared with all previously selected photobooks and all cover pages (or barcodes on the back side of each photobook).

Retrieving a digital photo based on an image taken by a mobile phone's camera is the core of the overall process. This process can be further divided in three parts. First the digital photos in the photobooks are analyzed and stored in a media repository. Afterwards the server receives images from the mobile device. These images are pre-processed and passed to the matching module. This module determines the digital photo and finally the digital photo is transmitted back to the requesting mobile phone as shown in Figure 3.

3.4 Matching Algorithm

Matching images taken by the mobile device with photos from the photobook requires the recognition of the respective photo. Since the images are taken by different users it can be assumed that the characteristics of such images vary. Even if the quality of the image might be high due to recent development of mobile phone's cameras (see [23]) photos will be distorted and might be affected by shadows and light sources. In addition, the pages of the photobook might be curled and there is no guarantee whether the image is taken in landscape or portrait format. The image might be rotated, partially occluded, and varied by light conditions. Furthermore, it can be expected that an image taken by the user, as shown in Figure 4, does not cover the extent of the printed photo exactly.

These potential variations require a robust algorithm.

³Considering the amount of printed photobooks it might be necessary to use other techniques to reduce the search-space. This can be achieved by taking an image of the individual barcode on the back side of each photobook instead of taking an image of the cover page.

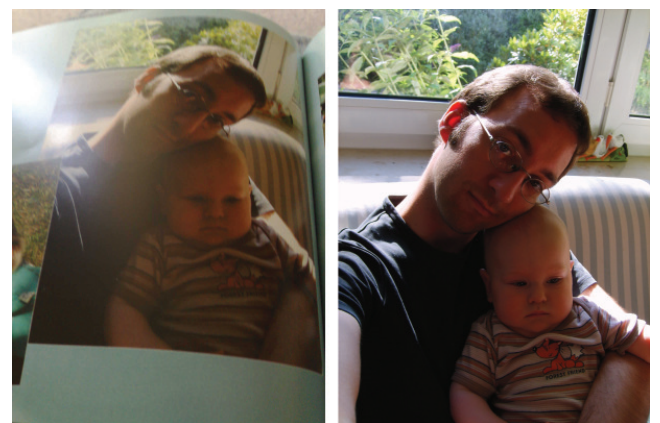


Figure 4: Image of a photo in a photobook taken with a mobile phone's camera (left) and the digital photo used to produce the photobook (right).

Biologic neuronal networks store features of objects independent from scale and rotation [35]. Inspired from this idea are algorithms that extract local features from images. A widely used rotation invariant algorithm is the Harris-Corner-Detector, however, it is not invariant against scale. One of the first local features extractor that is rotation and scale invariant is the Scale Invariant Feature Transform (SIFT) algorithm developed by Lowe [25]. Lowe claims that SIFT keypoints are invariant against scale and rotation. In addition, the SIFT algorithm is robust against lighting condition and small distortions. Based on SIFT, algorithms can be developed that are robust against partial occlusion. Beside SIFT a number of algorithms exist which extract local features that outperforms SIFT in terms of speed or robustness [5, 21]. However, since this is only true for specific domains we use a SIFT based approach.

4. IMPLEMENTATION

The main parts of the Bookmarkr system are the pre-processing of photo collections and the matching of received images created by a mobile phone's camera with pre-processed photos. These two steps are described in the following sections. Afterwards the implementation of the system is outlined.

4.1 Pre-processing photo collections

In the pre-processing phase digital photos that are part of a photobook are analyzed and SIFT keypoints are extracted. The digital photos are structured in photo collections. We define a photo collection as all photos that are part of one photobook. Digital photos have resolutions up to twelve million pixels and more. To make the process robust against noise caused by the camera and lossy compression the photo resolution is reduced. This is also necessary to reduce the number of keypoints extracted by the SIFT algorithm which increases the processing speed (see Section 5.1) and reduces the required amount of disk space. In addition the photos are converted to greyscale since the SIFT algorithm does not require colour images. Afterwards, the SIFT algorithm is used to extract keypoints from each photo. Figure 5 shows a photo with the SIFT keypoints detected in the photo. Every keypoint is annotated with an identifier that links to the respective high resolution photo.

To find the photo that was photographed by the mobile phone it is necessary to match the image taken by the mobile phone with the digital photos. In order to do that SIFT keypoints are extracted from the image. For each of the keypoints from the image taken by the mobile phone the keypoint which is the nearest neighbour from the digital photos must be found. To find the nearest neighbour in a reasonable amount of time we use a best-bin-first heuristic [6] based on a kD-tree to find the nearest neighbour.

The desired outcome of the matching is the most similar photo and not a collection of similar photos. It is therefore sufficient to determine the keypoint which is the nearest neighbour of the whole photo collection and it is not necessary to determine the nearest neighbours of each photo. Thus, we use a single kD-tree to store all keypoints from one photo collection. This decreases the matching time dramatically compared to using a kD-tree for each individual photo. It must be taken into account that using the best-bin-first heuristic decreases the robustness of the process. The robustness is further decreased by larger kD-trees resulting



Figure 5: SIFT keypoints (marked with yellow circles) detected in a photo.

from storing keypoints from multiple photos in a single tree. However, as we will show in Section 5 this is, at least for our application, not relevant. The outcome of the pre-processing are one kD-tree for each photo collection each containing the SIFT keypoints of the photo collection. The kD-trees are stored alongside the digital photos in the media repository.

After pre-processing all photo collections the process is ready to receive images taken by the user's mobile phone. Due to the same reasons as for the photos from the photo collections received images are scaled to a consistent resolution and converted to greyscale. Then the SIFT keypoints are extracted from the image and stored in a simple list. This list of keypoints is handed over to the matching process.

4.2 Image matching

The matching process compares the image taken by the mobile phone with the photo collections using the extracted SIFT features. The process iterates through the list of photo collections treating each photo collection individually. For each keypoint of the image the nearest neighbour from the photo collection is determined using the best-bin-first heuristic. The keypoint is considered as a match if the distance between the second nearest neighbour and the image's keypoint is twice as high as the distance between nearest neighbour and the image's keypoint. Using this relative threshold instead of a fixed threshold is more robust against noisy images. Since it is likely that photos contain similar content (e.g. an object photographed from different angles) it is important to determine the second nearest neighbour that belongs to the same photo as the nearest neighbour. Otherwise sections of photos that appear in more than one photo will not be identified as a match.

If a match is found the nearest neighbour keypoint from the kD-tree is marked as used. This keypoint will not be considered in further nearest neighbour searches until all keypoints of the image are processed. Discarding already used keypoints makes the process more robust against noise. Otherwise false positives are more likely to occur in blurry regions. In addition it slightly increases the process's performance.

For each found match an identifier of the respective photo is stored in a list. After all of the image's keypoints are processed and the total number of matches is below a certain



Figure 6: Photo sharing application running on a Nokia N95. (a) Application’s main menu, (b) view to select a photo by taking an image, (c) digital photo is displayed in conjunction with metadata

threshold (e.g. 8 matches) it is assumed that no matching photo has been found. Otherwise the identifier of the photo with the highest number of matches is returned. The process is repeated for each photo collection and the photo with the most matches of all photo collections is determined. The found high resolution digital photo is then transmitted back to the mobile phone application. This client application presents the photo to the user.

4.3 Client/Server platform

We implemented the server application including the matching process described above using a combination of Microsoft’s C sharp and C under Microsoft Windows XP using Rob Hess SIFT implementation [16] and Intel’s Computer Vision Library OpenCV [19]. All parts of the system that are not performance critical, for example the network communication, are implemented in C sharp. The performance critical parts, in particular the SIFT algorithm and the keypoint matching, are implemented with C using Microsoft’s C compiler provided with Visual Studio 2005. To benefit from multi-core systems the matching is implemented using a configurable number of threads.

To address a large spectrum of mobile devices the mobile client application is implemented using the Java ME (MIDP 2.0, CLDC 1.1), the most common platform for mobile phones. Communication between the mobile application and the photobook server can be established via any TCP/IP connection. The retrieved digital photos can be sent to other devices, stored in the devices file system, and deleted. Figure 6 shows three screenshot of the mobile application. Upon starting the application the main menu is displayed. By selecting the “Create” item the camera’s view is displayed on the screen. After shooting an image of a photo to select it the digital photo is displayed in conjunction with metadata. Using this view’s option menu the photo can be transferred to others. Users can also browse through a list of their collected photos.

5. PERFORMANCE EVALUATION AND USER STUDY

Using the Bookmarkr prototype described in the previous section we performed two evaluations. Synthetic experiments were conducted to determine the performance of the developed system. A user study was performed to evaluate the suitability of our system for the targeted audience.

5.1 Performance tests

We conducted a synthetic evaluation to determine the systems performance regarding speed and recognition accuracy. The time needed to retrieve a photo out of a collection of photos and the number of correctly determined photos was measured.

5.1.1 Evaluation method

The experiments focused on testing the matching process. Thus, we did not evaluate the two stage process described in Section 3 but instead evaluated how fast and accurate the system can determine the correct photo from a given photo repository.

To test the system 2000 photos grouped in 20 photo collections each containing 100 photos and representing one photobook were collected. The photos were taken by different people using various cameras, dates, and locations. We randomly selected and printed three photos from each photo collection resulting in 60 printouts. We took images from three different angles (see Figure 7) from each printout to acquire 180 test images in total. The task of the evaluation was to return the correct photo for each of the 180 test images.

The experiment was conducted using an Apple MacBook running Microsoft Windows XP equipped with a 2 GHz mobile Intel Core2Duo processor and 1 GB memory. The images of the printouts were taken using the developed system running on a Nokia N95 8GB [27] mobile phone equipped with a five megapixel digital camera mounted on a tripod.

To determine the necessary resolution to achieve a high precision the system was tested with different image sizes. The photo collections were pre-processed (see Section 4.1) in four different resolutions. The maximum widths and heights were 480 pixels, 640 pixels, 960 pixels, and 1440 pixels. The images of the printouts created with the mobile phone were resized to four different resolutions as well. The maximum widths and heights were 640 pixels, 800 pixels, 1024 pixels, and 1280 pixels. For each combination the required amount



Figure 7: Four of the 180 test images. The top-left and the bottom-right image are taken from above, the top-right image from front-above, and the bottom-left image is a close-up.

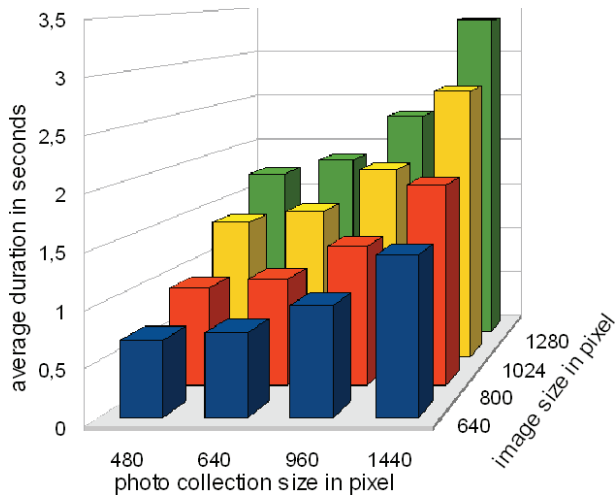


Figure 8: Average processing time per image to select a photo from a collection of 2000 photos.

of time to process all images, the number of unrecognized photos, and not recognized photos was measured.

5.1.2 Results

The average processing time to retrieve one photo out of 2000 was between 0.68 and 3.39 seconds. It took 0.68 seconds for a resolution of 480 pixels for the pre-processed photos and a resolution of 640 pixels for the image taken by the mobile phone. For a pre-processing resolution of 1440 pixels and an image size of 1280 pixels it took 3.39 seconds. The average processing time per image is shown in Figure 8.

In addition, the number of images without a match (because all photos' number of matches was below a threshold) and the number wrong matches was recorded. Figure 9 shows the number of errors for the four image sizes 640 pixels, 800 pixels, 1024 pixels, and 1280 pixels and a fixed photo collection size of 960 pixels on the left. On the right the number of errors is shown for the four photo collection sizes 480 pixels, 640 pixels, 960 pixels, and 1440 pixels with a fixed

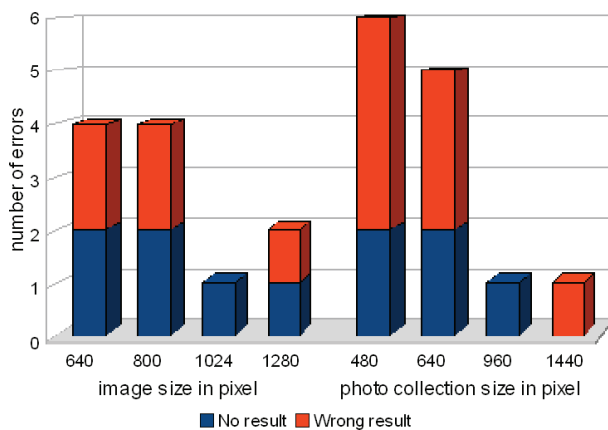


Figure 9: Total number of errors for four image sizes on the left and four photo sizes on the right if selecting 180 photos from a collection of 2000 photos



Figure 10: Two images with a small number of detected keypoints.

image size of 1024 pixels. The highest number of wrongly returned photos was four wrong photos for a resolution of 480 pixels used for pre-processing the photo collection. In addition, with this resolution the system returned no photo for two request images. With a resolution of 960 pixels used for pre-processing the photo collection and an image size of 1024 pixels 179 photos were determined correctly and for one image no photo could be determined.

5.1.3 Discussion

The system returned 179 correct photos out of 180 photos (i.e., more than 99 percent) with a pre-processing resolution of 960 pixels and a resolution of 1024 pixels for the images taken by the mobile phone. For these resolutions it took 1.9 seconds on average to retrieve a photo. Errors observed with lower resolution revealed which kinds of images are difficult to match. Typical images that lead to incorrect results are the two images shown in Figure 10. Common for these images was that the number of detected keypoints was low due to noise and large homogeneous or blurry areas. The small number of keypoints resulted in a small number of matches below the threshold used to consider a photo as a match.

In theory the time to retrieve a photo from a collection of photos is independent of the number of photos in this collection and needs constant time due to the used heuristic. The needed memory, however, increases linear with the number of photos. In practise the robustness of the photo recognition will degrade with the number of photos in one photo collection because of the nature of the best-bin-first heuristic. Scalability of the used techniques beyond 2000 photos is not addressed in the conducted performance test. However, as described in section 3.3 the search space can be reduced if the user must select the photobook first. Thus, robust recognition for 2000 photos is sufficient if the search space can be reduced to photobooks selected by the user.

Another source of error resulted from similar photos as the two shown in Figure 11. Such photos are very common if photographers do not remove similar photos from their photo collection. We, however, assume it is unlikely that real photobooks contain very similar photos or photos with large blurry areas. Thus, it can be concluded that we can expect even better results for user generated photobooks.

5.2 User study

After having shown that the system is able to reliably return correct photos in a synthetic performance test a user study was conducted to evaluate the system with average persons. In the user study the system, as described in Section 4, was used to determine the accuracy of digital photo retrieval when used by the ordinary person. The user study



Figure 11: Two very similar photos that are a common error source.

focuses on using a point-and-shoot interaction technique to retrieve photos from a photobook. Thus, participants did not have to select the photobook first. In addition to objective data we collected the feedback of potential users. We analyzed the percentage of correctly recognized photos and measured participants' impressions using questionnaires.

5.2.1 Evaluation method

Our group of ten participants was diverse with regard to gender, age, and mobile application experience. The participant population consisted of an equal split of males and females. Two participants were between 20 and 30 years old, four participants were between 30 and 40 years, two between 50 and 60 years, and two were between 60 and 70 years. Two participants had no experience with digital photography and did not own a mobile phone. All the other participants had experience with digital photography. Four of them stated that they share a mobile phone with their partner and rarely use it. The others use a mobile phone on a regular basis.

The experiment was conducted in a living room styled setting at different daytimes and without electric light. The room has one window. Figure 12 shows the setting of the evaluation. Each participant was welcomed and asked to take a seat on a sofa. The aim of the system and the evaluation procedure was explained to the participants. They were



Figure 12: Participant during the evaluation using a Nokia N95 to select photos from a photobook.

not asked to pay special attention while taking the images. Afterwards the participants were handed a photobook and a Nokia N95 mobile phone running the application described in Section 4. The assistant asked the participants to select at least ten photos using the application.

After selecting the photos participants were asked to complete a questionnaire. Besides demographic questions the questionnaire consisted of multiple choice questions, ratings on a five point Likert scale ranging from zero (never/difficult) to four (often/easy), and an open question for additional remarks.

An Apple MacBook running Microsoft Windows XP equipped with a 2 GHz mobile Intel Core2Duo processor and 1 GB memory was used to run the server application. The photos of the photobooks were pre-processed with a resolution of 1024 pixels. The images taken with the Nokia N95 8GB had a resolution of 960 pixels. The communication between the Nokia N95 and the server application was established using a local Wi-Fi connection.

5.2.2 Results

Each participants took twelve images on average resulting in 121 images. For 117 taken images the system returned the correct result (96.7%). For the remaining four images (3.3%) the system returned no result. Two participants signaled that they would like to stress test the system's capabilities (e.g. "Lets see if this also works") before taking an image that could not be recognized. Figure 13 shows four examples of images taken by the participants. On the top are two examples that were correctly recognized and on the bottom are two examples that could not be recognized. Another photo taken during the evaluation that was correctly recognized is shown in Figure 4. In general the vast majority of



Figure 13: Examples of images taken with a Nokia N95 during the user study.

the taken images are very blurred. It could be observed that most images were close-ups consisting of only a fraction of the respective printed photo.

Two participants mentioned that they are afraid of using the system after the system was explained. They stated that they are not used to handling mobile phones. For these two and an additional participant it was difficult to handle the mobile phone with one hand. They repeatedly held their fingers in front of the phone's camera while trying to take an image. Apart from holding the phone and pressing its keys no participant had difficulties in using the system. For one participant the introduction to the system was not sufficient. After taking four images she realized that the photos shown on the phone's display are not the images she took with the phone's camera.

The questionnaires revealed the following insights. Tools participants use to show their photos at home are diverse. Most often mentioned tools are printed photos (9 times), notebooks (7 times), digital cameras (7 times), and desktop computers (3 times). Participants stated that they would like to receive photos from others either occasionally or often ($M = 2.3$, $SD = 0.82$). They are occasionally promised to receive photos they never obtained ($M = 2$, $SD = 1.15$). All but two participants rated the system as easy or very easy to use ($M = 3.2$, $SD = 0.79$). Most participants could envision using the system ($M = 2.5$, $SD = 1.08$). They would like to use the photos received on the mobile phone to order printed photos (8 times), send them via e-mail (5 times) or to their desktop computer (2 times), and view them on the mobile phone (2 times). Most participants envision letting others use the system to annotate their photos ($M = 2.9$, $SD = 0.99$).

5.2.3 Discussion

The system returned the correct result for 96.7% out of 121 images. Subtracting the two images that were only shot to stress test the system's capabilities the system returned 98.3% out of 117 images. This can be considered a promising result in particular regarding the blurriness of the taken images. One cause for the blurriness is the auto-focus mechanism of the Nokia N95 that is not able to focus on very close objects. We assume that the auto-focus abilities of mobile phones will improve in the future.

It could not be observed that different light conditions affect the robustness of the system. However, the influence of light condition was not explicitly tested but user's need proper light condition to view the photobook itself. We assume that the light conditions sufficient for users are almost covered by the system. SIFT keypoints can be matched robustly under different light conditions and using the phone's integrated flash is possible for photobooks not printed on high-gloss paper.

A remarkable aspect is that all participants even the elderly, that do not own a mobile phone nor use a computer were able to understand and use the system. However, in particular older persons might not see the system's benefit because many of them believe to have no use for digital photos in general.

6. CONCLUSIONS AND FUTURE WORK

In this paper, we presented the Bookmarkr system that enables sharing of digital photos using printed photobooks. Using a point-and-shoot metaphor, digital photos in pho-

tobooks can be selected by taking an image with a mobile phone's camera. The taken image is compared with a photo repository and the user retrieves the photographed digital photo. Using a two-stage process the photobook is selected first to narrow the search-space. Users can experience the advantages and haptic sensation of printed photobooks while enjoying the benefits of digital photos. Our performance evaluation shows that digital photos can be retrieved with a high probability if the resolution of the photo collection and the mobile phone's camera is sufficient. The conducted user study shows that all participants can use the point-and-shoot technique to select photos without any problems.

The main aim of our future development is to further narrow the gap between physical and digital media. A promising approach is to bring together printed photos and online photo sharing communities that enables to share and annotate digital photos. However, we assume that a point-and-shoot interaction technique is not the best technique to access and create digital annotation. A magic lens as it has been implemented with markers for paper maps [33] seems to be the better approach. However, for objects that captivate with their visual appeal, such as printed photos, photobooks, and posters marker based approaches are clearly not acceptable. Thus we will investigate options to implement a markerless magic lens for photos. Furthermore integration with our ongoing work on Contextual Bookmarks [15] will enable the use of one system to access information related to diverse media.

7. ACKNOWLEDGMENTS

This paper is supported by the European Community within the InterMedia project (project No. 038419). We thank all partners for sharing their ideas with us.

8. REFERENCES

- [1] E. Auchard. Flickr to map the world's latest photo hotspots. <http://www.reuters.com/article/technologyNews/idUS/HO94233920071119>, November 2007.
- [2] M. Balabanovic, L. L. Chu, and G. J. Wolff. Storytelling with Digital Photographs. *Proceedings of the Conference on Human Factors in Computing Systems*, pages 564–571, 2000.
- [3] R. Ballagas, J. Borchers, M. Rohs, and J. G. Sheridan. The Smart Phone: A Ubiquitous Input Device. *Proceedings of the Conference on Pervasive Computing*, 5(1):70–77, 2006.
- [4] R. Ballagas, M. Rohs, and J. G. Sheridan. Sweep and point and shoot: phonecam-based interactions for large public displays. *Proceedings of the Conference on Human Factors in Computing Systems*, 2(07):1200–1203, 2005.
- [5] H. Bay, T. Tuytelaars, and L. V. Gool. Surf: Speeded up robust features. *Proceedings of the European Conference on Computer Vision*, 1:404–417, 2006.
- [6] J. S. Beis and D. G. Lowe. Shape indexing using approximate nearest-neighbour search in high-dimensional spaces. *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 1000–1006, 1997.
- [7] CeWe Color Holding AG. Cewe color holding ag factbook, November 2007.

- [8] CeWe Color Holding AG. Cewe color achieves projected earnings for 2007. Press release, April 2008.
- [9] S. Counts and E. Fellheimer. Supporting Social Presence through Lightweight Photo Sharing On and Off the Desktop. *Proceedings of the Conference on Human Factors in Computing Systems*, pages 599–606, 2004.
- [10] A. Crabtree, T. Rodden, and J. Mariani. Collaborating around Collections: Informing the Continued Development of Photoware. *Proceedings of the Conference on Computer Supported Cooperative Work*, pages 396–405, 2004.
- [11] B. Erol and J. J. Hull. Linking presentation documents using image analysis. *Proceedings of the Conference on Signals, Systems, and Computers*, 1, 2003.
- [12] X. Fan, X. Xie, Z. Li, M. Li, , and W.-Y. Ma. Photo-to-search: using multimodal queries to search the web from mobile devices. *Proceedings of the Workshop on Multimedia information retrieval*, pages 143–150, 2005.
- [13] G. W. Fitzmaurice. Situated information spaces and spatially aware palmtop computers. *Communications of the ACM*, 36(7):39–49, 1993.
- [14] D. Frohlich, A. Kuchinsky, C. Perring, A. Don, and S. Ariss. Requirements for photoware. *Proceedings of the Conference on Computer Supported Cooperative Work*, pages 166–175, 2002.
- [15] N. Henze, R. Reinert, X. Righetti, E. Rukzio, and S. Boll. Using camera-equipped mobile phones for interacting with real-world objects. *Proceedings of the Computer Graphics International Conference*, 2008.
- [16] R. Hess. Sift feature detector. <http://web.engr.oregonstate.edu/~hess>, 2008.
- [17] O. Hilliges, M. Wagner, L. Terrenghi, and A. Butz. The living-room: browsing, organizing and presenting digital image collections in interactive environments. *Proceedings of the Conference on Intelligent Environments*, pages 552–559, 2007.
- [18] N. V. House, M. Davis, M. Ames, M. Finn, and V. Viswanathan. The uses of personal networked digital imaging: An empirical study of cameraphone photos and sharing. *Proceedings of the Conference on Human Factors in Computing Systems (CHI 2005)*, pages 1853–1856, 2005.
- [19] Intel Corporation. Open source computer vision library website. <http://www.intel.com/technology/computing/opencv>, 2008.
- [20] International Organization for Standardization: Information Technology. Automatic identification and data capture techniques - bar code symbology - qr code. ISO/IEC 18004, 2000.
- [21] Y. Ke and R. Sukthankar. PCA-SIFT: A more distinctive representation for local image descriptors. *Proceedings of the Conference on Computer Vision and Pattern Recognition*, 2:506–513, 2004.
- [22] D. Kirk, A. Sellen, C. Rother, and K. Wood. Understanding Photowork. *Proceedings of the Conference on Human Factors in Computing Systems*, pages 761–770, 2006.
- [23] K. Leichtenstern, A. D. Luca, and E. Rukzio. Analysis of Built-in Mobile Phone Sensors for Supporting Interactions with the Real World. *Proceedings of the Pervasive Mobile Interaction Devices workshop*, pages 31–34, 2005.
- [24] Q. Liu, P. McEvoy, and C.-J. Lai. Mobile camera supported document redirection. *Proceedings of the ACM Conference on Multimedia*, pages 791–792, 2006.
- [25] D. G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
- [26] A. D. Miller and W. K. Edwards. Give and Take: A Study of Consumer Photo-Sharing Culture and Practice. *Proceedings of the Conference on Human Factors in Computing Systems*, pages 347–356, 2007.
- [27] Nokia Corporation. Nokia N95 8GB technical specifications. http://www.forum.nokia.com/devices/N95_8GB, 2008.
- [28] M. Nunes, S. Greenberg, and C. Neustaedter. Sharing digital photographs in the home through physical mementos, souvenirs, and keepsakes. Research Report 2007-875-27, Dept Computer Science, University of Calgary, 2007.
- [29] D. Okabe. Emergent Social Practices, Situations and Relations through Everyday Camera Phone Use. *Proceedings of the Conference on Mobile Communication*, 2004.
- [30] Photo Marketing Association International. Pma photo book report, 2006.
- [31] K. Rodden and K. Wood. How do people manage their digital photographs? *Proceedings of the Conference on Human Factors in Computing Systems*, pages 409–416, 2003.
- [32] M. Rohs and B. Gfeller. Using camera-equipped mobile phones for interacting with real-world objects. *Proceedings of the Conference on Pervasive Computing (PERVASIV 2004)*, pages 265–271, 2004.
- [33] M. Rohs, J. Schöning, A. Krüger, and B. Hecht. Towards Real-Time Markerless Tracking of Magic Lenses on Paper Maps. *Adjunct Proceedings of the 5th International Conference on Pervasive Computing*, pages 69–72, 2007.
- [34] E. Rukzio, G. Broll, K. Leichtenstern, and A. Schmidt. Mobile Interaction with the Real World: An Evaluation and Comparison of Physical Mobile Interaction Techniques. *Proceedings of the European Conference on Ambient Intelligence*, pages 1–18, 2007.
- [35] K. Tanaka. Mechanisms of visual object recognition: monkey and human studies. *Current Opinion in Neurobiology*, 7(4):523–529, 1997.
- [36] R. Want, K. P. Fishkin, A. Gujar, and B. L. Harrison. Bridging physical and virtual worlds with electronic tags. *Proceedings of the Conference on Human Factors in Computing Systems*, pages 370–377, 1999.
- [37] Y. Zhou, X. Fan, X. Xie, Y. Gong, and W.-Y. Ma. Inquiring of the Sights from the Web via Camera Mobiles. *Proceedings of the IEEE International Conference on Multimedia and Expo*, pages 661–664, 2006.