Shoot & Copy: Phonecam-Based Information Transfer from Public Displays onto Mobile Phones

Sebastian Boring[1], Manuela Altendorfer[2], Gregor Broll[1], Otmar Hilliges[1], Andreas Butz[1]

University of Munich, Media Informatics
Amalienstrasse 17, 80333 Munich, Germany
+49-89-2180-4684

[1] {sebastian.boring, gregor.broll, otmar.hilliges, andreas.butz}@ifi.lmu.de
[2] altendor@cip.ifi.lmu.de

ABSTRACT

Large public displays have become pervasive in our everyday lives, but up to now, they are mostly information screens without any interaction possibilities. Users tend to forget what they saw relatively fast after leaving such a display. In this paper, we present a new interaction technique for transferring information from a public display onto a personal mobile phone with its built-in camera. Instead of having to rely on their memory, users simply take a picture of the information of interest. Instead of just storing the image, our system then retrieves the actual data represented on the screen, such as a stock quote, news text, or piece of music. The *Shoot & Copy* technique does not require visual codes that interfere with shown content or reduce screen real estate. Our prototype allows users to capture an arbitrary region of a standard desktop screen, containing icons, which represent pieces of data. The captured image is then analyzed and a reference to the corresponding data is sent back to the mobile phone. Once the user has time to view the information in more detail, our system allows retrieving the actual data from this reference. We present our prototype and the methods it uses for image processing, as well as an evaluation of our interaction technique illustrating its potential use and applications.

Categories and Subject Descriptors

H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems – *evaluation/methodology*; I.5.4 [**Pattern Recognition**]: Applications – *Computer vision*; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *evaluation/methodology, interaction styles*;

General Terms

Algorithms, Experimentation, Human Factors

Keywords

Mobile camera phones, large public displays, image processing

1. INTRODUCTION

Large electronic displays become increasingly affordable and can thus be found in many public places. They display high-resolution images, commercials, late-breaking news, weather forecasts, or stock exchange prices (see Figure 1). In spite of this rich information, people are not intended to actively interact with public displays but to passively view their displayed content. Usually public displays are also too big for personal interaction or even out of reach in order to protect them from damage. Hence, if data displayed on a public screen is of interest for an individual, s/he needs to remember the content in order to retrieve additional information at home. In most cases, the information has been forgotten by the time users reach their home/office. This decreases the intended communication efficiency.

Figure 1. Left: an unreachable public display in a subway station. Right: a user capturing information of interest.

As camera-equipped phones become more and more available and accompany their users most of their time, people use them as their external visual memory when taking pictures of interesting things like advertisements. In addition, today's cameras are capable of taking high-quality images, which allows the retrieval of highly detailed information. But while people can conserve this information, they can not directly interact with it. They have to extract the displayed information themselves and transfer it to other applications (e.g. a browser) – in most cases even to a personal computer – in order to observe the corresponding data and take the desired action (e.g. open a URL or browse the web for certain words contained in the photograph).

So far, direct interaction with pictures taken with mobile phone cameras is only possible through the augmentation with visual markers and the recognition of their visual patterns. Semapedia [18] for example uses visual markers for linking objects from the real world to corresponding articles on Wikipedia. The existing

approaches in this area use little two-dimensional barcodes that can be attached to different objects. These visual markers usually only encrypt a small amount of information (e.g. a number identifier or a URL) which can nevertheless be used as input for the interaction with different services and applications.

In this light, the combination of powerful mobile devices with high-resolution cameras and large displays bears interesting new opportunities. Besides taking a picture of the information of interest, additional data can be transferred to the mobile phone. For that purpose, the photograph needs to be analyzed using computer vision. In the process, whole screens or displays can be used as rich input into the image recognition process. The technology for communicating information via visual patterns blends in with the media that display them – the screen becomes the marker.

In this paper, we present *Shoot & Copy* as a new interaction technique for retrieving information from displays by simply taking a picture of the desired content. We describe its prototypical implementation and the algorithms we used for image comparison, and finally discuss the results of an initial user study.

2. RELATED WORK

Ever since the work of Kato and Billinghurst [11], recognizing visual markers and their encrypted codes through image processing is an established technology for the identification of augmented objects and the interaction with them. With the advancement and dissemination of mobile technology, the step from desktop systems like the AR Toolkit [1] to mobile devices has been made. During the last years, the mobile recognition of visual markers has become one of the most established and widespread technologies for mobile interaction with objects from the real world. The information gathered through mobile image processing serves as physical hyperlinks to access actual object-related information thus bridging between entities in the real world and associated digital information.

Rohs and Gfeller [14] present one of the most elaborate visual code systems for mobile phones that uses their cameras as sensors for the recognition of two-dimensional markers. The system implements a lightweight visual code recognition algorithm that is adapted to the technical constraints of mobile phones (e.g., low resolution and image quality, limited processing power). It supports the simultaneous detection of multiple codes and determines the mobile phone's orientation.

This system is advanced and allows using the rotating angle of mobile phones, the amount of tilting and phone movements as additional input parameters next to the actual code value [15]. The system also enables the application of a code coordinate system that extends the inherent coordinate system of the visual markers. This feature allows the precise selection of unmarked items e.g. regions on a map or rows in a table in relation to a single visual marker. The combination of these features enables new interaction patterns which can also be applied for the interaction with public displays.

Similar to the visual code system, other visual marker recognition systems have emerged in the last years. Among them are Semacode [17] that is based on the Data Matrix standard [8], QR Code [9], ShotCode [1] or BeeTagg [4]. Most of these standards simply use encrypted code values as input for mobile applications. Semapedia [18] for example uses Semacodes [17].

In contrast to the previously presented systems, d-touch [5] encodes information in the topological structure – the nesting of adjacent black and white regions - of its markers. That way the visual design of markers is not dependent on the function – the encryption of identifiers – and allows greater artistic freedom. Visual markers can not only become more aesthetically pleasing, but they can also blend in better with their environment.

An application that uses visual codes for the identification of service parameters is presented by Broll et al. [6]. It explores the application of mobile interaction with physical objects for invoking Semantic Web Services. For that purpose, this approach uses mobile phones to extract information from posters that are augmented with visual markers and NFC-tags to use it as input for a corresponding Web Service, e.g. in order to buy movie or transportation tickets.

The transition from the interaction with static markers to the interaction with large (public) displays and their dynamic content is the next logical step in the evolution of the interaction with the real world using mobile phone cameras.

Two mobile interaction techniques for large displays are presented by Ballagas et al. [2]. For *point & shoot* users aim at objects on large displays using the mobile phone screen with a cross-hair cursor. Pushing the joystick of the phone, a grid of visual code markers (see [14]) is temporally superimposed over the contents shown on the large display and shot by the mobile phone camera. This approach relies on the local coordinate systems of the visual codes in order to identify the absolute position of the targeted object on the large screen where the selection is finally highlighted. The *sweep* interaction technique allows moving the cursor on the large display as it detects movements of the mobile phone by sampling successive images from the camera instead of using visual markers. While *point & shot* has users focus on the mobile phone screen, users of the *sweep* technique can focus on the large display.

Ballagas et al. [3] examine different issues concerning the interaction with large public displays. They define this kind of interaction to happen in three domains (personal, semi-public and public) and identify different constraints that should be considered for the design of interaction techniques: serendipity, portability, sanitation, dexterity, multi-user, physical security, information security and privacy, social acceptability, intentional versus unintentional interaction and maintenance. Taking these aspects into account, the authors approve the use of personal mobile devices for the interaction with large displays, compared to direct touch-based interaction as well as speech- and gesture input.

Pick-and-Drop [13] is a pen-based interaction technique for the transfer of information between multiple displays. This direct manipulation technique is based on the common drag-and-drop interaction-paradigm, but extends it to the interaction between multiple computers and takes it beyond the constraints of input devices and user interfaces for the interaction with single computers. With Pick-and-Drop, users pick up objects from a screen by tapping and lifting it with a pen. The pen virtually holds the object and can drop it at arbitrary positions on the same or

different displays. This technique virtually binds pen-ids to object-ids in order to implement the transfer of information between multiple displays and computers that are connected via a network.

In a similar approach, InfoPoint [10] uses a more elaborate pointing device in order to implement a drag-and-drop like interaction technique for transferring information between network-connected computers and appliances. The system relies on visual markers as a uniform interface for the identification and selection of target objects. In addition, a pointing gesture is needed to select them, which is deemed to be a very natural selection metaphor. Similar to Pick-and-Drop, InfoPoint translates the drag-and-drop interaction-paradigm from desktop computing into a Get-and-Put operation in the real world. The InfoPoint handheld-device gets and puts objects, respectively their ids through the recognition of visual markers. A small display on the device provides visual feedback about captured objects, their selection and their dropping.

The C-Blink system [12] demonstrates an enabling technology for the mobile interaction with large public displays. Instead of using mobile phone cameras to track objects on the large display, a camera mounted above the display tracks the mobile phone which runs an application that rapidly changes the hue of the phone's screen in order to send light signals. Users can control a cursor on the large screen by simply waving the screen of their mobile phones in front of the camera.

Greenberg et al. [7] investigate the interaction between public displays and PDAs with a focus on the distinction between personal and public information and the interaction with concerned objects. The results of implementing and evaluating a prototype for creating and sharing notes using both PDAs and large displays encourages system designs that allow users to easily shift objects from private to public without a rigid separation of both domains.

3. SHOOT & COPY PROTOTYPE

In order to evaluate the *Shoot & Copy* technique, we built a fully functional prototype using a state of the art mobile phone (the Sony Ericsson K800i). Its camera provides high-quality pictures with a resolution of 3.2 mega pixels. In addition, several camera controls, such as focus, zoom, flashlight, white balance and image quality can be controlled. The pictures taken with the camera-phone are used to find a region on a large display captured by the user. The wireless communication between the phone and a server for the image calculations can be realized using either Bluetooth or GPRS. For our evaluation, we used a direct Bluetooth connection to a laptop computer nearby.

3.1 Basic Concept

The prototype can be divided into two major phases, the capture phase and the access phase. In the capture phase, the user takes a picture of the desired region on a large display (see Figure 2). The image is then sent to the display's host computer, which compares the image to its own screen content and identifies the originally captured region. Once the region has been identified, the application takes the icon closest to the region's center (the reference icon). The information represented by this icon (e.g., a text document, image, music or video file) is then placed on a

web server while the corresponding URL is sent back to the user's mobile phone. Alternatively, the original file could be sent directly to the mobile phone where it can then be viewed or played back by the user at any time.

Figure 2. A user interacting with a large display using the *Shoot & Copy* prototype.

In the access phase, the user places the mobile phone next to a Bluetooth enabled computer. The list of references stored on the mobile phone is then sent to an application running on the computer, where it is shown as an interactive list of icons and corresponding URLs. The user can now select single or multiple items to retrieve the actual data from the specified url. Images, text documents, music files and videos are shown (or played back) within the browser, while complex files, such as ZIP archives or other binary files, need to be downloaded by the user. If the file has been sent directly to the mobile phone, the user can access it using the phone's built in media browser. Thus, the content can be viewed or played back on demand.

Our prototype is designed to make technology invisible. Thus, it does not use standardized visual markers or barcodes. Instead, the entire screen acts as a single marker without being recognized as such by the user. Hence users are only exposed to familiar technology, such as a desktop screen and a camera-equipped mobile phone.

3.2 Icon Identification

In this section we will illustrate how sub-regions of pictures can be identified by a computer. Before designing our algorithm, we studied several desktop configurations in order to determine icon relations that could be used in further image processing. Given the view-angle of the built-in camera and assuming a minimum distance of 50 centimeters between the camera phone and the display surface, we found that a captured image will almost always show several icons at a time. Hence, if a captured image contains a group of icons, the image processing algorithm can search the entire screen for the same constellation of icons. Spatial relations are measured regarding the reference icon, which is the one closest to the image's center. All other icons are then

sorted according to their angle and their distance to the reference icon.

In the rare case that the captured image only contains a single icon, the constellation hence does not have any further angles and distances stored. Considering the width of the captured icon, it is compared to the set of desktop icons whose neighboring icons all have large distances.

In order to allow a wide range of desktop settings in the general case (i.e. several icons in the image), we have implemented an algorithm, which is independent of zoomed or rotated images. This will be described in further detail in the following sections.

3.3 Image Processing

To identify the spatial relations mentioned above, the captured image needs to be preprocessed. We asked a small group of people to capture a sub-region of a large display using the camera-phone and found that in most cases the images were twisted, because they did not hold the camera in parallel to the display borders. To address this issue, our system first uses a Hough transformation algorithm to detect characteristic lines that correspond to the edges of icons. It then rotates the image such that the most prominent line is horizontal. This line does not necessarily represent the bottom border of an icon. Hence, the system will consider four possible rotations (0, 90, 180 and 270 degrees) after subsequent, rotation-invariant steps (see Figure 4).

Once the image has been rotated, the system uses the Canny edge detection algorithm to detect the icons' borders. This also identifies unwanted structures, such as the textual labels of icons. This can be addressed by assuming a nearly quadratic shape of an icon. If this step results in a large number of identified icons, the image will be further processed using a region growing algorithm for verification. Using both results, the captured icons are cut out of the originally captured image for later color comparison. At this time, the application splits up the results into four possible rotated candidates. Thus, the following steps are being performed four times.

The icon closest to the image center is chosen as the reference icon. For all other icons, their distance and relative angle to the reference icon are stored and they are sorted according to the

distance in ascending order. The same is then done once for all icons on the desktop. Their positions can easily be obtained from the operating system. We now have four candidate vectors consisting of a number of angles and distances, and the same lists for each desktop icon.
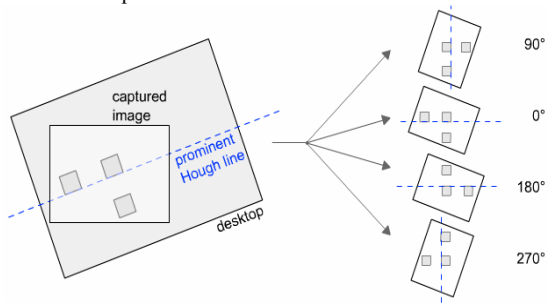

Figure 4. Left shows four possible prominent lines in the captured image. Right shows the four rotations (0, 90, 180 and 270 degrees) of the captured image.

During the next phase the system matches the captured icon relations with existing constellations on the desktop by looking for close matches in the combinations of distances and angles. While the lists for all desktop icons contain the distances and angles to all other icons, the match is only done using as many entries, as the candidate icon provides. Since the lists are sorted by distance, these will automatically be the closest neighboring icons.

While distances have linear coherences (scale factor), angles only have constant relations (summation of angles). The system is now searching for a desktop icon that has the same constellation using polar coordinates for distance and angle (see Figure 5). Each possible candidate is then stored in a list for further use. In several cases, this phase results in only one match. If more relations match, our prototype will use the color information of icons to determine the correct constellation.

Figure 3. Left shows the prototype with one possible match. Right shows the interaction with two identical icon constellations on the desktop. The magnifications denote the capturing process (top) and the result shown to the user (bottom).
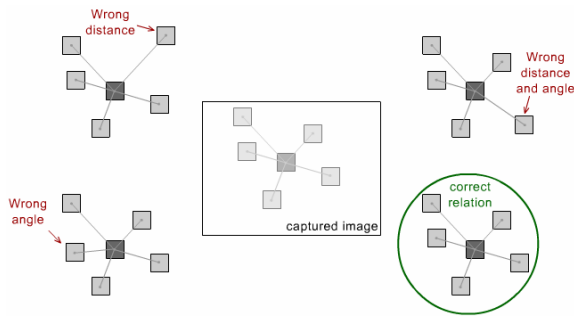
Figure 5. Center shows the originally captured desktop constellation. Lower left shows the correct match whereas all others indicate incorrect constellations.

Before the system is able to compare the color information of two potential matches it needs to preprocess the captured icon in order to correct the color shift of the phone's camera. Our prototype first automatically adjusts the color levels of the captured image followed by a slight gamma (value = 0.98) correction. Subsequently, the system will compare both the captured and the original icon by applying the following sequence of filters:

- *Image difference and image inversion*: This filter calculates the difference of both icons followed by a color inversion of the resulting image. Thus matching pixels will gain a bright color whereas highly different pixel will turn dark.

- *Image grayscale*: The inverted image will be transformed into a grayscale image as we are only interested in the luminance values of each pixel. Our prototype uses the BT709 algorithm (21.25% of the red, 71.54% of the green and 7.21% of the blue value), since this will also yield perceptually correct color differences.

- *Image pixelation*: Bicubic subsampling of the difference image allows error reduction in the case that a single pixel is nearly identical to the original one while the surrounding pixels are highly different. Using this filter a faulty pixel will get an average color and will then be left out by applying the following filter.

- *Color threshold*: The processed picture will now be transformed into a binary image in order to count black and white pixels. This is done using a range according to the luminance value to decide whether a pixel will turn black or white. Our prototype uses a threshold filter with an acceptance range from 205 to 255.

The number of white pixels in the resulting binary image indicates how similar these icons are. Larger numbers of white pixels suggest a higher correlation between both icons (see Figure 6). As mentioned before, the system considers four rotations at the same time as it does not know the camera's rotation at capture time. The color comparison algorithm only considers the best equal matches of each rotation. Subsequently, the algorithm merges the four results and evaluates the matches considering their color distances. Our prototype is robust against a matching failure due to the different rotation of the image. In nearly all cases the best match is the correctly rotated one. To disambiguate

remaining images with almost identical color distance the system presents a list of possible candidates to the user.
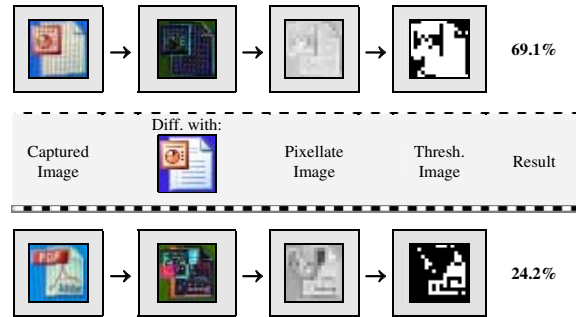

Figure 6. Demonstration of the color matching algorithm: Top shows the comparison of two identical icons. Bottom denotes the relation of two similar icons in terms of color with different content.

3.4 Presentation of Results

Once the system has calculated the matching icons it sends the matches to the user's mobile phone. In the case that several potential matches have been identified, a list of these matches is sent. Depending on the length of this list, the mobile phone decides whether to show the exact match as it appears on the desktop (one result) or several possible candidates that can be selected as in the picture browser on the phone (multiple matches). Figure 3 illustrates these visualizations in different situations.

The user is now able to select the relevant match. In order to obtain the information associated with the icon, s/he has to accept one of the system's suggestions. In addition, s/he can also reject the suggestion and either chose another one or take a new picture. After accepting a match, the user's mobile phone will either receive a URL of the corresponding file or the file associated with the icon.

3.5 Accessing Stored Content

Accessing the files can be done in two ways. In the case that a reference has been stored on the mobile phone, the user needs to place it next to a Bluetooth enabled device (mostly a personal computer). The list of stored references is then being sent to the application running on the device and will be displayed for further selection. The content of each reference can be accessed by selecting the corresponding icon. The file will then be shown in a browser window or the associated application respectively.

After downloading a file, the reference will not be deleted from the mobile phone's memory. This can be done be explicitly delete the references using the user interface on the phone.

If the entire file has been sent to the mobile phone, the user can select it at any time using the built-in media browser. In most cases such files are image, music or video files that can be rendered directly on a mobile phone. The user can select them in the same way as if they would have been transferred using standard information transfer mechanisms (e.g., USB).

4. ARCHITECTURE

As mentioned before, our system is divided into two separate phases. The first phase is called capture phase whereas the second one is called access phase. Viewing the previously captured content can be further split up. The references can either be downloaded using an additional (desktop) computer or viewed using the mobile phone.

4.1 Capture Phase

During the first phase, the computer which serves a large display acts as the server accepting Bluetooth connections from mobile phones. The phone itself represents the client. Whenever a new client connects to the host, a unique identifier (e.g. its MAC-address) will be used for later association. Thus, multiple users are able to use the system simultaneously.
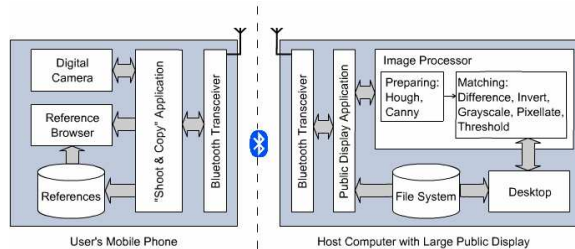

Figure 7. Architecture of *Shoot & Copy* in the capture phase.

Figure 7 illustrates the system's architecture during the first phase. The left side shows the mobile phone (client), whose main components are: a digital camera functionality, a browser window which displays suggested and/or stored references similar to the phone's media browser, the main application that processes received Bluetooth data or prepares data, e.g. captured images, to be sent wirelessly to the connected server.

The right part of the figure shows the server that is hosting a (semi-)public display. The Bluetooth transceiver receives data from connected mobile devices or sends references back to one of the clients. The image processor module is responsible for the matching of captured icons to displayed desktop icons. The computer's file system is connected to its desktop. Hence the server application is able to determine the associated file represented by an icon.

4.2 Access Phase

In the case that references will be downloaded to a secondary device, such as a computer, the mobile phone acts as the server. If the application is running on the client machine, it scans for mobile servers until one has been found. As soon as the mobile phone has been detected as a *reference server*, the application requests a list of stored references. Each item of this list contains the visual representation (i.e., the icon) as well as the URL to the corresponding file. After the list has been received, it can be presented to the user in order to allow a selection of a previously captured reference.

If the files are viewed directly on the mobile phone, a secondary device is of course not necessary. If the user wants to view the files on a nearby computer this is, of course, still possible.
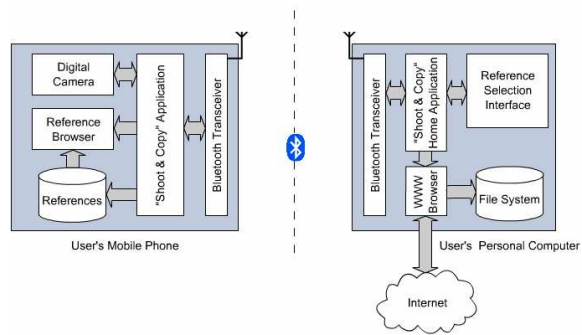

Figure 8. Architecture of *Shoot & Copy* in the access phase. The user's personal computer is optional.

Figure 8 illustrates the architecture during the second phase using a secondary device. The software on the mobile phone remains the same. Thus, users do not need to switch between different applications. The computer's part is entirely different. Its reference browser is responsible for showing the received list of references of the mobile phone. Using an internet connection, selected files can be downloaded using the URL associated to each reference item.

5. EVALUATION

We have conducted an initial user study to observe how people would use our *Shoot & Copy* prototype. Our main goal was to find out if the system is easily useable even by inexperienced users. In addition, we were interested in the acceptance of users according to our scenarios. Furthermore, we also wanted to see what kind of applications users envision when using the *Shoot & Copy* prototype.

5.1 Participants and Environment

The study we conducted consisted of 28 participants (14 subjects were female). 25 attendees own a mobile phone whereas only 15 of them use their built in photo camera. The average rated expertise in using a mobile phone was 3.32 on a scale from 1 to 5 (1 means inexperienced, 5 denotes an expert user). The operating knowledge of a photo camera was 2.53 on the same scale. The participants that use their phone's camera take about 4.4 pictures per month. Among the subjects a variety of professions was present, such as computer scientists, animal health students, secretaries, lawyers, researchers, bankers, consultants and school teachers. Their average age was 25.2 years whereas the youngest participant was 18 and the oldest was 47.

The system was deployed on a standard personal computer with a digital projector attached to it. A laptop has been used to simulate the user's home computer. In addition, the mobile part of the software was deployed on a Sony Ericsson K800i. Both computers have been placed in different rooms to indicate the spatial separation of both the capture (public) and the access phase (private).

5.2 Tasks and Procedure

In the beginning participants had to fill out a questionnaire. In this they were asked about their technical expertise, their knowledge

regarding mobile phones and cameras as well as some demographical statements. Subsequently we introduced our idea and some scenarios the system could be used with.

In the next phase, users were asked to take several pictures of icons placed on the large displays desktop. Between subsequent shots, we changed the arrangement to enforce multiple matches for some captured images. The desktop icons we used represented well-known file types, such as Word documents, PDF documents, JPEG images, standard media files (audio / video), file folders and archives (ZIP and RAR). We asked the participants to take snapshots of different kinds of icons in order to test the system robustness as well. After completing the first task, users were asked to move to another room representing their home. In this phase, they were able to download the references they have captured using our *Shoot & Copy* prototype onto a laptop. During this study users were not able to view the files directly on the mobile phone because of bandwidth limitations caused by our Bluetooth connection.

Finally, we asked the participants about their subjective judgments regarding our prototype. We assessed advantages and disadvantages of the system as well as their opinion whether they would use such a system or not. We asked them about the maximum time the system should need until a potential match is returned to the user's mobile phone. Furthermore, we wanted to know if they felt uncomfortable by copying data onto the mobile phone by taking a picture. We also wanted to figure out in which situations a system like *Shoot & Copy* could make sense in public spaces such as train stations or airports.

5.3 Results of the Study

Using the questionnaires we were able to answer several important issues while interacting with mobile devices and large displays. One interesting question was the time period users would accept until the reference is stored on the mobile phone. We did not expect that ten seconds would be acceptable by most users. Because estimating a time period is hard for people, we also observed their reaction during the study whereas most of the participants stated that the reference was stored quite fast. Their subjective feeling was that it took less than five seconds. In addition, some of them were surprised when the mobile phone displayed the potential matches.

We also asked the users about their subjective judgment of *Shoot & Copy*. Here we found that the average rating of our system was 4.1 on a scale from 1 to 5 (5 is best). 45% of the participants that rated the system at least with 4 were experts in using their mobile phone. In addition, 82% of the attendees stated that they would use the system if it is available (see Figure 9). 55% of the subjects that would use the system again in the future rated their expertise at least with 4. They also mentioned that scenarios especially in subway stations are interesting due to short periods of time until the next train arrives. This also has been found by Rukzio et al. [16]. By waiting for a train, scanning and retrieving information displayed on public screens would be of interest to almost all participants. One common statement was that people forget information regarding concerts, movies or book releases until they arrive at home. Taking pictures of such displayed items in combination with additional data can reduce the loss of information. In addition, some participants mentioned that this technique could offer an alternative to information transfer

technologies (e.g. phone explorers) that come off-the-shelf with a new mobile phone. They stated that they would like to use *Shoot & Copy* on their desktop as well.
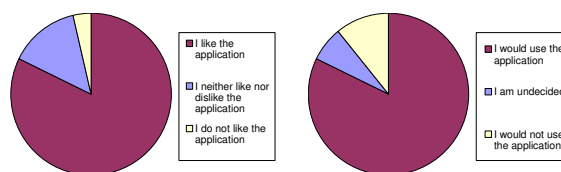

Figure 9. Left side illustrates the user's opinion whether they like the application or not. Right depicts how many users would use the application if it is available.

During the test we also observed the time users need to feel comfortable with the copying technique. In all cases users were able to use it after a short explanation of the camera's capabilities. In addition, even the first icon they tried to copy did work in all cases indicating an easy-to-use technique.

The most interesting result we have observed is that nearly all participants felt comfortable while using our prototype. In the post questionnaire they stated that everybody in the environment could see that they are capturing information, however, nobody could see *which* one has been photographed. Thus, interacting with semi-sensitive data (i.e. public data to which an individual does not want to be linked) using a camera-phone at some distance does not affect privacy, neither perceived nor objective.

5.4 Performance and Error Rates

Besides the questionnaires we also observed potential performance issues and error rates. In two cases the system did not work in the beginning due to a broken Bluetooth connection. In all other cases, the connection worked without any problems. Regarding the error rates we have found that extremely bad lighting conditions (e.g. spot light hitting the display) can cause failures in the color comparison of two icons.

The connection problem will become less problematic in the near future as more and more mobile phones are equipped with a wireless LAN module. This allows a reliable as well as fast connection and therefore is a technology that will improve the *Shoot & Copy* prototype. In addition, the lighting issue could be solved using an adaptive gamma correction. Overall, the prototype worked well throughout the entire study. The average time between capturing and receiving the information was about 9.2 seconds and was not negatively mentioned by any participant. The icon comparison led to the right reference in all cases.

6. CONCLUSIONS AND FUTURE WORK

In this paper we have presented a new technique of gathering information units from large displays without using standardized visual markers. The basic idea is that a user can take pictures of the information of interest using a mobile phone camera. The corresponding information will then be sent to the user for later reference on a PC. Thus, users do not have to remember the information displayed on a large public display. In addition, the prototype can be modified to send the corresponding files directly

onto the mobile phone. In this case, users can view or edit the information immediately if they wish to.

We have built a prototype to test our algorithms and to further evaluate the use of such systems. The *Shoot & Copy* prototype employs a large front-projected screen, a laptop representing the user's home computer and a mobile phone with a built-in camera for high quality images. The total time between taking an image and receiving the reference is still close to ten seconds but will decrease due to increasing computational power and higher transmission bandwidths. In addition, some of the processing code can be implemented on the mobile phone once their processing power increases.

We have further conducted a user study to observe how users interact with our system. The study gave interesting insights on average times and potential use of such applications. Further, almost all participants would use *Shoot & Copy* to receive additional information about currently displayed commercials, news or the like. The most interesting result nevertheless is the fact that users did not think their privacy was being affected although transferring information in public.

In the near future we plan to deploy the prototype to a large public display. In addition, we want to implement several scenarios we have identified during the user study. This includes that we want to use images, posters or CD covers displayed on public screens instead of standard desktop icons. Some of the targeted scenarios are: displaying several headlines of interesting news at the same time; showing a number of CD covers in order to retrieve audio samples of an album; displaying advertisements such as book releases which allow obtaining some sample pages of the captured book cover. We will extend our image processing algorithm to handle these types of information. Subsequently, we will formally evaluate the system in order to learn about the way in which users interact in real-world public scenarios with content other than desktop icons. We also want to observe how users feel about their privacy in public. This might also be of interest with regard to commercial applications of our technique.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] AR Toolkit. http://www.hitl.washington.edu/artoolkit/

[2] Ballagas, R., Rohs, M., Sheridan, J., Borchers, J., *Sweep and Point & Shoot: Phonecam-Based Interactions for Large Public Displays*, In CHI ' 05 extended abstracts,New York, NY, USA, ACM, 2005, 1200-1203

[3] Ballagas, R., Rohs, M., Sheridan, J., Borchers, J., *BYOD: Bring your own Device*, In Proceedings of UbiComp 2004

[4] BeeTagg. http://www.beetagg.com/

[5] Costanza, E., Leinss, M., *Telling a Story on a Tag: The Importance of Markers's Visual Design for Real World Applications*, MIRW 2006, workshop in conjunction with MobileHCI 2006, Espoo, Finland, ACM, 2006, 6-9

[6] Broll, G., Siorpaes, S., Rukzio, E., Paolucci, M., Hamard, J., Wagner, M., Schmidt, A., *Supporting Mobile Service Usage through Physical Mobile Interaction*, In Proceedings of PerCom 2007, White Plains, NY, USA, IEEE, 2007

[7] Greenberg, S., Boyle, M., Laberge, J., *PDAs and Shared Public Displays: Making Personal Information Public, and Public Information Personal*, In Personal and Ubiquitous Computing 3(1/2), London, UK, Springer, 1999, 54-64

[8] International Organization for Standardization: Information Technology – International Symbology Specification – Data Matrix. ISO/IEC 16022, 2000

[9] International Organization for Standardization: Information Technology – Automatic Identification and Data Capture Techniques – Bar Code Symbology – QR Code. ISO/IEC 18004, 2000

[10] Kohtake, N., Rekimoto, J., Anzai, Y., *InfoPoint: A Device that Provides a Uniform User Interface to Allow Appliances to Work Together over a Network*, Personal and Ubiquitous Computing 5(4), London, UK, Springer, 2001, 264-274

[11] Kato, H., Billinghurst, M., *Marker Tracking and HMD Calibration for a Video-based Augmented Reality Conferencing System*, In Proceedings of IWAR 99, San Francisco, CA, USA, IEEE, 1999, 85-94

[12] Miyaoku, K., Higashino, S., Tonomura, Y., *C-Blink: A Hue-Difference-Based Light Signal Marker for Large Screen Interaction via any Mobile Terminal*, In Proceedings Of UIST'04, Santa Fee, NM, USA, ACM, 2004, 147–156

[13] Rekimoto, J., *Pick-and-drop: A Direct Manipulation Technique for Multiple Computer Environments*, In Proceedings of UIST'97, Banff, Canada, ACM, 1997, 31-39

[14] Rohs, M., Gfeller, B., *Using Camera-Equipped Mobile Phones for Interacting with Real-World Objects*, In Advances in Pervasive Computing, Austrian Computer Society (OCG), Vienna, Austria, 2004, 265-271

[15] Rohs, M., *Real-World Interaction with Camera Phones*, In Proceedings of UCS 2004, Tokyo, Japan, Revised Selected Papers, LNCS 3598, 2005, 74-89

[16] Rukzio, E., Schmidt, A., Hussmann, H., *Physical Posters as Gateways to Context-aware Services for Mobile Devices*, In Proceedings of WMCSA 2004,

[17] Semacode. http://semacode.org/

[18] Semapedia. http://www.semapedia.org/

[19] ShotCode. http://www.shotcode.com/

Workshop on Ubiquitous Display Environments, Nottingham, UK, 2004