Image Recognition Techniques for in a Mobile Public Interactive Display

Graeme Smith[1], Professor Gary Marsden[2]
Department of Computer Science
University of Cape Town, Private Bag X3, Rondebosch, 7701
Tel: +27 21 650 2712, Fax: +27 21 650 4511
email: graeme0811@gmail.com[1]; gaz@cs.uct.ac.za[2]

Abstract - This paper describes a system which runs on a mobile phone that allows for the distribution of media packages to users with Bluetooth enabled camera phones. Users take photographss of specially designed posters and send them, using Bluetooth, to the system. An algorithm that enables the system to recognize the user images is developed, evaluated and modified using two rounds of user experiments. The final algorithm is found to correctly recognize the image photographed 87% of the time.

Index Terms—xxxx, xxxxx, xxxxxxxx

I. Introduction

A World Bank study in 2006 found that in “upper middle income” nations (a class which includes South Africa) 20.2% of the population has a fixed line telephone subscription, whereas 65.8% has a mobile phone. In countries classed as “lower middle income” the ratio was 30.4% mobile users against 10.9% fixed line and in “low income” nations the ratio was 7.5% against 0.9% [1]. Clearly, then, when designing technology solutions for the developing world, cellular handsetcellular handsets should feature prominently in our thought processes.
One project which utilizes the high prevalence of cellular handsetcellular handsets is the Snap ‘n Grab project (also known as Big Board) [2]. A Snap ‘n Grab system consists of a large screen display and a computer to drive it. The screen displays images representing content on the computer. This content can be pictures, videos, music files or text. It can represent job offers, AIDS information or simply (as was the case in a recent trial) recordings of the local gospel choir. Users can take a photo of an image representing a subject they wish to know more about (there are eight such options shown in Figure 1). Having done this, they can then send it, via Bluetooth, to the system. The system carries out image recognition on the image and replies by sending data, via Bluetooth, back to the user. All this is done at no cost to the user, as Bluetooth is free. Snap ‘n Grab can also allow users to create their own “media packages”: an account can be set up by sending the system a v-card, followed by the media for the package. This new package will then be displayed on the screen.

Figure 1 - The Snap 'n Grab system being displayed on a large LCD display.

Some drawbacks to the current Snap ‘n Grab system are *cost*, *security* and *mobility*. A large screen display of 40”(the average size of a Snap ‘n Grab screen) costs about $1000. The computer to drive the screen would add an extra $300. These costs are prohibitive in developing communities.
In places where there is a high crime rate, a high theft rate in particular, it is difficult to find a balance between system accessibility and security. Ideally the system should always be available for people to use; however, it is costly and difficult to ensure that the screens and computers are not easily stolen.The big screens are further constrained by the fact that they require a power source. This means that they could not be easily deployed in vehicles or places where electricity is not constantly available.We have developed a solution to the above three problems. Firstly, we replace the big screen with either printed posters, or individual stickers for each item. We can remove the need for a PC by porting the software to a cellular handsetcellular handset thereby dramatically cutting the cost. Using a cellular handset and paper posters greatly simplifies the security issue and bypasses the mobility problem. A high-end cellular handset costs about $550, approximately 30% of the total cost of the original system. Secondly, security will be reduced to finding a suitable place to store the phone, such as a locked cupboard, or in the glove compartment of a vehicle. The posters will not need to be secured, as they will be of relatively low cost to replace. Finally, since the system is based on a cellular handset, it is clearly mobile. It could be deployed in a vehicle, in the middle of a field or in a small village where there is no constant source of electricity. The cellular handset would need to be charged occasionally, but it would not need constant power.

II. Related work

A. Image Recognition

In recreating the Snap 'n Grab system on a cellular handset we encounter the problem of conducting image recognition, an inherently complex process, on reduced hardware. The Snap 'n Grab system utilized Scale Invariant Feature Transform (SIFT) [3] to carry out the image recognition step. It was, however, too slow to run on a cellular handset's limited processor, as it took about 5 seconds to run on a normal desktop computer.

Cellular handset image recognition research has focused primarily on barcode reading – both one and two dimensional – with many phones, such as the Nokia N95, being released with a built in barcode scanner. Cellular handsets have been shown to be capable of reading and decoding a barcode using the built-in camera [4]. Our system, however, required that the cellular handset be able to locate the barcode in a larger image. Rohs [5] has developed a method of orienting two-dimensional barcodes using a system of guide bars. The system works by locating a "guide" bar on the side, and then using that location to find the two blocks at top and bottom left of the image. The image can then be transformed to its original dimensions, and the two-dimensional barcode read accurately (Figure 2) (this is discussed in greater detail in the Algorithm Design chapter). Our system makes use of this concept of guide bars in order to locate the one-dimensional barcode along the top of the image. The barcode can then be decoded, and the image matched to an equivalent media pack.

Figure 2 - The 2D barcode used by Rohs, with the guide bars on the right and bottom of the image

B. Public Displays

The Hermes system, developed by Cheverst et al. [6] was one of the first publicly situated displays to utilize Bluetooth technology to allow people to download media. It did, however, rely on a touch screen to allow users to select which data they wanted. The Snap 'n Grab system [2] was the first system to allow users to Bluetooth their selection, by means of a photograph, to the system in order to choose which media they desired. Another important aspect of the Snap 'n Grab system is that it did not require any software to be downloaded onto the user's cellular handset: all they are required to do is take a photo, using the cellular handset's built-in camera and to send it, via Bluetooth, to the system. This is a task people have become used to performing and thus is not unknown or intimidating.

In terms of static posters, shot codes have been placed on posters such as movie posters to allow cellular handset users to easily access websites related to the poster. Software to read these the codes has been included on cellular handsets such as the Nokia N95 [7,8]. All of these systems require 3rd-party software to be downloaded onto the user's cellular handset in order to perform the decoding task.

Our system uses the Bluetooth technology in the same way as the Snap 'n Grab system, and couples it with a form of static poster-based shot codes, in the form of barcodes and guide bars. This system does not require any special software to be installed on the users' handsets.

C. Usability

Interactive displays are still fairly novel. To improve their usability Somervell et al. have suggested a list of heuristics associated with large, publicly situated displays [9]. Due to the fact that a static display was used, not all the heuristics apply to this project. Those that do are:

- Appropriate colour schemes to support information understanding;
- The layout should reflect the information according to its intended use;
- The display should show the presence of information, but not the details.

We will use these heuristics in creating our system. Finally, the Hermes system was evaluated by a user study where the user was asked to perform tasks such as taking a picture of items on the screen and sending these pictures to the display. Our evaluation will follow a similar form, wherein typical users will be asked to take photographs with their own handsets to create a realistic corpus of test data.

III. SNAP 'N GRAB LITE

A. System Overview

The new system, termed Snap 'n Grab Lite, was developed in two, iterative phases with user experiments taking place at the end of each phase.

The system was developed for Windows Mobile devices, and was deployed on an HTC Touch Pro running Windows Mobile 6.1. This platform was chosen due to the high processing power of the device and the fact that the original Snap 'n Grab system was written in C#, which simplified the transition from desktop computer to mobile device significantly.

The Snap 'n Lite program is a server program that runs on the Touch Pro and listens for incoming Bluetooth messages. When a message is detected, it is received and, if it is an image, it is analyzed and the requisite data is then sent back to the sender.

B. Algorithm Design

1) Guide Bars

As mentioned previously, image recognition is a complex process. Because speed of response is a significant factor in our application, it is not, at present, possible to utilize existing image recognition algorithms such as SIFT or SURF [3,10]. This is because the average response time for these algorithms, when implemented on a mobile phone, is greater than 5 seconds [11], which is far slower than the response time required for our application. Thus additional mechanisms must be utilized in order to simplify the recognition process. Barcode scanning is essentially a solved problem, with many applications now being shipped with standard mobile phones. A simple solution, from a technical viewpoint, would be to replace the images used in the original Snap 'n Grab system with barcodes. This, however, would not be of great use, considering the application of the system, since people need to know what they are taking a picture of, as that will allow them to make a decision as to what media they wish to receive. Therefore, a dual system must be implemented, utilizing both pictures, for the human users, and a barcode system, for ease of recognition. The next technical challenge is that, in most barcode reading applications implemented on mobile phones, the barcode is assumed to take up the majority of the image, and is thus simple to find and read. In our application there is no guarantee that this will be the case, and indeed it is extremely likely that it will not be. Thus we need a way to efficiently locate the barcode within the image, so that it can be read and the requisite media sent to the user.

Rohs et al developed an algorithm for locating and orienting a two-dimensional barcode within an image [12]. They introduced the concept of guide bars, which sit alongside the barcode and are designed in such a way as to make them easy to identify within an image. Their algorithm for finding the guide bar works in the following manner:

- The image is thresholded (converted to black and white).
- The resulting image is then divided into contiguous regions using a two-pass approach.
- The regions are analyzed and their second-order moments are determined. These signify their "eccentricity" in the x and y direction.
- The ratios of the moments for each region are calculated and compared the to expected ratio for the guide bar. The best matching region is thus accepted as the guide bar.

In initial tests it was found that it was simpler to adapt Rohs algorithm and use two guide bars, one on each side of the image. Once found, a scanline can be run from the top of the left guide bar to the top of the right guide bar.

Figure 3 - An example of a poster used in the first iteration of Snap 'n Grab Lite

2) Barcode System

In the first iteration of the Snap 'n Grab Lite system we decided to use an adaption of the UPC barcode system [13]. The UPC system uses a combination of seven bars of equal thickness to represent a single digit. Further, the bars are arranged in such a manner that, since there is no gap between them, there will always appear to be only four bars per digit (since adjacent bars of the same colour will appear to form one, thicker bar).

We originally used a seven digit code with the UPC barcode system and positioned it along the top and bottom of the image, so that the picture can be taken upside down if necessary – see Figure 3.

IV. Experimental Design

A. Aim of experiment

The aim of the experiment was two-fold:

Firstly, to ascertain ideal values for a set of controlled independent variables (namely image size, poster medium and lighting conditions) under which the algorithm performs best. This will assist us in future deployments of the system to ensure optimal results.

Secondly, we wish to test how well the algorithm performs under the best-case set controllable variables, as well as some uncontrollable variables such as differing camera types, resolutions and "user idiosyncrasies" in taking photographs. We wish to test whether users are able, without instruction, to take photos that the system can analyze and from which it can decode the correct barcode number.

B. Task

The experiment took place in two parts, one involving users and one conducted by the designer.

For the user-participation part: Five different posters were created. Each poster was printed in five different sizes (5cm, 7cm, 10cm, 15cm and 20cm) on glossy and matte paper. They were then grouped according to size and material, thereby forming ten groups of images (a group will contain all the posters of a certain size and a certain paper material). The users were required to take 10 photographs using their cellular handset cameras. They took one photo of a random poster in each group. The lighting conditions were changed after each photograph, alternating between directed lighting (simulated by a bright lamp), dappled light (simulated by lamp light through a diffusing material) and ambient light.

For the second part: The pictures taken by the users were retrieved and sent to the HTC Touch Pro. The recognition software was run on each image to determine whether or not it could locate and decode the barcode.

Figure 4 - The five posters used in the first round of experiments

C. Variables

Poster medium:	Glossy and matte posters were tested.
Poster size:	The size of the physical poster (which therefore affects the size of the barcodes)
Lighting:	Three types of lighting were tested: ambient, directed and dappled. Directed light was created by aiming a lamp directly at the poster. Dappled light was created by placing a diffusing material in front of the lamp.
Success ratio:	The ratio of correct media to incorrect.

D. Materials

1) Equipment and venue

The system was run on an HTC Touch Pro.
The users used their own mobile phones to take the pictures. The experiment was conducted in a closed room in the Computer Science Department at UCT.

2) Participants

Participants were sourced from the University of Cape Town. Ages ranged from 21 to 26. All were comfortable with taking images using a mobile phone.

E. Procedure

The users were asked to choose a random poster from each group and to take a photo of it using their mobile phone. Once they had taken all ten photographs the images were retrieved. They were not told how to take photos of the posters.

V. RESULTS OF EXPERIMENT

A. Results

A simple initial analysis was carried out on the data. This involved finding the success ratio (number of successful decodings compared to the total number of images). The results showed that the algorithm only worked on 44% of the images. This was not an acceptable level of success and, hence, a second iteration of development was undertaken.

B. Discussion

Two key problem areas where identified. The first was that users were taking photos that were not adequately focused on the poster they were photographing. For example, some users took photos where the poster in question occupied approximately 20% of the total image area. This makes it difficult for the algorithm to correctly locate the guidebars.
The second problem area is somewhat related to the first in that, even if the bars were successfully located, the barcodes were of such a fine grain (containing forty nine "bits" representing the seven digit number that was encoded) that they were often misread and returned incorrect values.

VI. REVISION OF SNAP 'N GRAB LITE SYSTEM

A. Changes to algorithm

To enable easier reading of barcodes, the barcode system was altered. The new system utilized binary encoding rather than the UPC system. A thirteen bit binary code was used, thereby reducing the number of bars from twenty eight to thirteen. The barcode along the bottom of the image was removed and the height of the top barcode was increased by 40% to further increase the chance of a successful decoding.
Further, the thresholding algorithm was modified to use an adaptive threshold algorithm as described by Gu [14].

B. Changes to sticker design

There was a definite need to get users to take photos that were composed, as far a possible, of only the poster in question. To do this we replaced the "target lines" from the first iteration with a full frame (as can be seen in Figure 5).

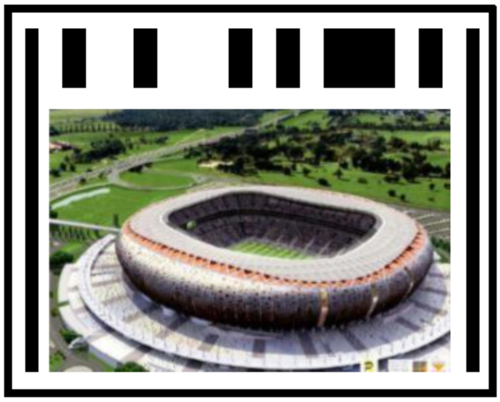

Figure 5 – An example of a poster used in the second iteration of Snap 'n Grab Lite

C. Second round of experiments

New posters were then created, using the new barcode system and framing lines. The experiments were then repeated. The second round was run as similarly as possible to the first round, with the same images (albeit with the new frames) being used for the posters, the participants being sourced from the same demographic and the same instructions being given.

Figure 6 - The five posters used in the second round of experiments

VII. RESULTS OF SECOND EXPERIMENT

A. Results

The same initial analysis as that for the first round was carried out, this time showing a 73% success rate.
The next step was to determine the best set of controllable parameters under which to deploy the system.
Since we have multiple independent explanatory variables and a binary response (the system either outputs a correct value or it does not), the chosen modeling approach was logistic regression within the framework of the generalized linear model (i.e. GLM with a logistic link function) [15].
This enables us to determine to what extent each variable influences the chance of getting the correct output from the algorithm.
The results showed that the majority of variables were not found to be significant factors on the final result. The size of the posters, however, was a significant predictor of success, with images 10cm and larger performing significantly better than those of 5cm and 7cm (with p-values of 0.0045, 0.00595 and 0.03390 for 10cm, 15cm and 20cm respectively).
In order to determine the accuracy of the algorithm under "best case" conditions the dataset was trimmed to exclude all images under 10cm in size. The resulting success rate was 87%.
The final step was to combine the trimmed datasets from the first and second rounds of experiments and run the GLM again to determine if the second round results were significantly better than the first round.

The resulting table showed the second round to be significantly more successful than the first round (p-value=0.031).

B. Discussion

Thus we can conclude that the second version of the algorithm, in conjunction with the new poster design, performed significantly better than the original algorithm and poster design. A success ratio of 87% was considered acceptable since the remaining 13% of images that did not work were, in general, too poorly photographed for the current algorithm ever to work on (see Figure 7 for an example)

Figure 7 - An example of a poorly taken image. Notice how the left guidebar is not in the frame, thereby causing the algorithm to fail.

VIII. Conclusion and Future Work

A. Conclusion

This paper has showed how the original Snap ‘n Grab media distribution system has been adapted to run on a mobile phone, thereby opening it up to many potential further uses. The challenges faced were presented, as well as how the system was adapted to overcome these challenges. Experiments were carried out to evaluate how the algorithm performed. It was found to perform poorly in the first round of experiments, thus it was revised and a second round of experiments followed. The results from the second round of experiments were significantly better than those of the first round, with a success rate of 87% for posters of 10cm and larger.

B. Future Work

Further study will be conducted by deploying the system in environments where it is likely to be used in the future. We will use the information gathered from the experiments as to what poster sizes to use to ensure that the algorithm performs optimally. Theses further studies will enable us to better understand how users go about acquiring media as well what media they are most interested in.

IX. References

[1] Donner J. Research approaches to mobile use in the developing world: A review of the literature. The Information Society 2008;24(3):140-159.

[2] Creating and sharing multi-media packages using large situated public displays and mobile phones. Proceedings of the 9 th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI 07).

[3] Object recognition from local scale-invariant features. Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on; 1999.

[4] Barcode readers using the camera device in mobile phones. Cyberworlds, 2004 International Conference on; 2004.

[5] Rohs M. Real-world interaction with camera phones. Lecture notes in computer science 2005;3598:74.

[6] Exploring bluetooth based mobile phone interaction with the hermes photo display. Proceedings of the 7th international conference on Human computer interaction with mobile devices & services: ACM New York, NY, USA; 2005.

[7] Bluetooth content distribution stations on public transit. Proceedings of the 1st international workshop on Decentralized resource sharing in mobile computing and networking: ACM New York, NY, USA; 2006.

[8] Physical posters as gateways to context-aware services for mobile devices. Sixth IEEE Workshop on Mobile Computing Systems and Applications, 2004. WMCSA 2004; 2004.

[9] Usability heuristics for large screen information exhibits. Proceedings of the Ninth IFIP TC13 International Conference on Human Computer Interaction; 2003.

[10] Bay H, Tuytelaars T, Van Gool L. Surf: Speeded up robust features. Computer Vision–ECCV 2006 2006:404-417.

[11] Interactive museum guide: Fast and robust recognition of museum objects. Proceedings of the first international workshop on mobile vision: Citeseer; 2006.

[12] Rohs M, Gfeller B. Using camera-equipped mobile phones for interacting with real-world objects. Advances in Pervasive Computing 2004:265–271.

[13] Adams R. UPC and EAN Bar Code Page. 2009; Available at: http://www.adams1.com/upccode.html. Accessed 30/4/2010, 2010.

[14] J. Gu. Development of computer vision algorithms using J2ME for mobile phone applications.University of Canterbury. Computer Science and Software Engineering; 2009.

[15] Cook D, Dixon P, Duckworth WM, Kaiser MS, Koehler K, Meeker WQ, et al. Binary response and logistic regression analysis. Part of the Iowa State University NSF/ILI project Beyond Traditional Statistical Methods 2001.

Graeme Smith received his undergraduate degree in 2007 from the University of Cape Town and is presently studying towards his Master of Science degree at the same institution. His research interests include usability, mobile computing and ICT4D.