MOBILE WEB BROWSING INITIATED BY VISUAL SEARCH

Berna Erol, Jamey Graham, Jonathan J. Hull, and Daniel G. Van Olst

Ricoh California Research Center
2882 Sand Hill Rd, Suite 115, Menlo Park, CA, USA
{berna_erol, jamey, hull, vanolst}@rii.ricoh.com

ABSTRACT

Mobile applications are increasingly becoming part of our lives and our work. Moreover, visual capture capability on mobile devices is increasingly being used for linking the real world and the digital world. In this demonstration paper we present an application for accessing and browsing web sites by pointing a mobile phone at a text book or a printout of a web site, taking a picture of a small region, performing content recognition and automatically displaying a web site that is linked to the content.

1. INTRODUCTION

Mobile phones are increasingly being used for functionalities other than voice calls, such as texting, e-mailing, taking pictures, listening to music, watching movies, playing games, and browsing the Internet. Although the computation and memory capabilities of mobile phones are increasing at a rapid pace, there are still limitations on display size and data entry because of the lack of a full keyboard. These limitations make search and access to information on mobile phones more difficult than access from a desktop computer.

In recent years, companies and researchers started to address these issues by supporting search via images and audio captured by a mobile device [1]-[11]. Visual and audio content search has been a research area for the last few decades. Easy-to-use audio and visual capture capabilities of mobile phones, combined with their being always-on and always connected, makes them an ideal platform for performing content search using images and audio. Currently available systems/research using visual data for mobile retrieval include [1]-[7], and web search and music search using audio/music data include [8]-[11].

We developed a mobile phone based document retrieval system where a user can point their mobile phone at a paper document, capture a picture or video, and retrieve the electronic version of the document or image and audio clips associated with a document [3][4].

In this demonstration paper, we present a new application where users invoke web links from paper printouts and books. For example, users take a picture of a small section of a text book and they retrieve the web site that explains that book chapter, as illustrated in Figure 1. Currently our system supports web link retrieval and browsing from an iPhone and Windows Mobile phones. In the following section we explain how our application works.
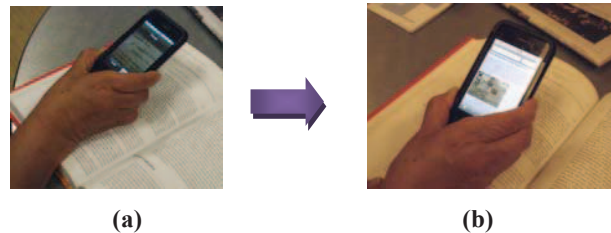

(a)   **(b)**

Figure 1. User points their camera phone at a text book, takes a picture, and retrieves a web page that gives more information about the particular section of text book.

2. SYSTEM DESCRIPTION

An overview of our visual retrieval and mobile web search system is given in Figure 2. A paper document or a book is input in either of two ways:

- **Scanning,** where our server receives a PNG file for each page image. This method is used when there is no electronic copy of the paper available.

- **Printing,** where the electronic document is printed (either through a printer or to a file) and our printer driver saves individual page images as PNG files.

Our image-based document retrieval engine, Mixed Media Reality (MMR), identifies the exact document id, page, and location on the page, given a document image captured by a mobile phone. MMR works based on extracting a unique signature from the layout of text boxes. A small text patch is sufficient to differentiate a document region from thousands of document pages. Retrieval accuracy depends on many parameters including the captured image size, lighting conditions, amount of focus and motion blur, and the amount of text in an image. Our experiments showed that for 176×144

resolution Palm Treo images, about 50% of document images can be correctly retrieved [3].
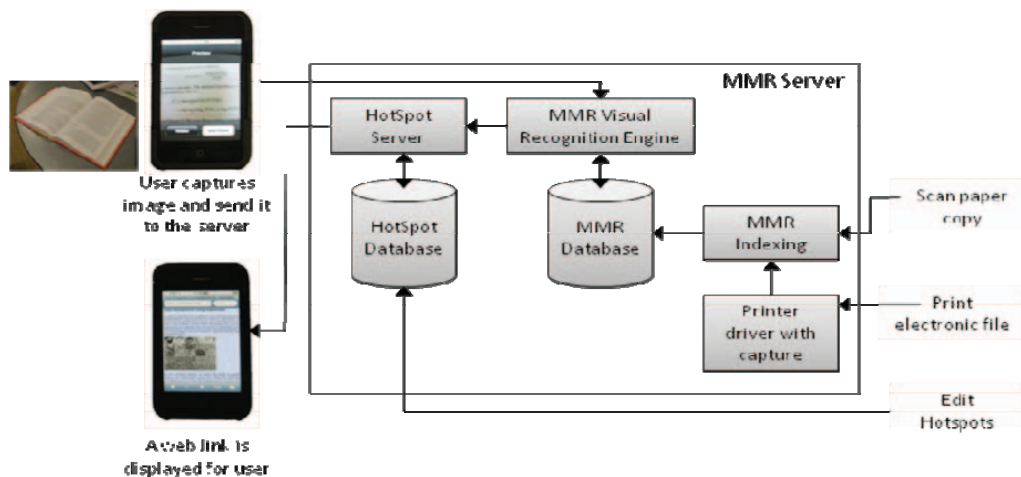

Figure 2. Mixed Media Reality visual retrieval system overview

A more advanced version of our algorithm was developed for targeting images captured by an iPhone. Our experiments showed that for 1200×1600 resolution iPhone images and image patches that contain many text lines, about 99% of the query images can be accurately linked to the electronic documents.

After PNG files are generated through printing or scanning, each page image is assigned a unique document id and indexed by the MMR indexer.

For associating a web link with a region on the document, we support two methods:

1. Users manually specify a region of the document by an x, y location and width and height values. Then the user can type or drag-and-drop a url to the region.

2. If a web page is printed from a web browser that uses our web browser plug-in, during printing we automatically parse the web page, extract links, and associate each link with a region defined by x,y and width and height.

Each document's links are stored as an XML file in a hotspot database and contains the document id, x and y locations, width and height, and the web link to follow.

On the mobile device, the user runs either our Windows Mobile or iPhone application to capture a text patch and to send it to our server. When the MMR visual recognition engine receives the patch image, it identifies the document id, and location on the page where the user took the picture. This information is then used to query the hotspot database and retrieve any web links associated with that part of the document. If a web link is found, then this is communicated back to the mobile client. At the mobile client, a mobile browser is launched and the retrieved web link is invoked.

3. CONCLUSIONS

In this paper we presented our demonstration system for invoking web links via visual content recognition on mobile phones. There are many extensions possible. For example, web links can point to student discussion groups for a particular chapter or a problem in a textbook.

4. REFERENCES

[1] Kooaba Image Recognizer, http://www.kooaba.com/

[2] SnapTell, http://www.Snaptell.com

[3] B. Erol, E. Antúnez, J. J. Hull, "HOTPAPER: Multimedia Interaction with Paper using Mobile Phones," ACM Multimedia Conf., pp. 399-408, 2008.

[4] Hotpaper Youtube video clip, http://www.youtube.com/watch?v=9Gv3wJNdvYA

[5] J. J. Hull, B. Erol, J. Graham, Q. Ke, H. Kishi, J. Moraleda, D. G. V. Olst, "Paper-Based Augmented Reality," 17th Int. Conf. on Artificial Reality and Telexistence, pp. 205-209, 2007.

[6] Q. Liu, P. McEvoy, and C.J. Lai, "Mobile Camera Supported Document Redirection," ACM Multimedia Conf., pp. 791-792, 2006.

[7] D. Schmalstieg, D. Wagner, "Experiences with Handheld Augmented Reality," IEEE/ACM ISMAR, pp. 3-15, 2007.

[8] Google Mobile App with Voice Search http://www.google.com/mobile/apple/app.html.

[9] Shazam Audio Search, http://www.shazam.com/.

[10] Yahoo OneSearch with Voice, http://mobile.yahoo.com/onesearch/voice.

[11] Lie Lu, Frank Seide. "Mobile Ringtone Search through query by Humming," Proceedings of ICASSP, 2008.