Visual Search Applications for Connecting Published Works to Digital Material

Jamey Graham, Jorge Moraleda, Jonathan J. Hull, Timothee Bailloeul, Xu Liu, Andrea Mariotti
California Research Center
Ricoh Innovations, Inc.
Menlo Park, California, USA
visualsearch@rii.ricoh.com

ABSTRACT

Visual search connects physical (*offline*) objects with (*online*) digital media. Using objects from the environment, like newspapers, magazines, books and posters, we can retrieve supplemental information from the online world. In this demonstration, we show a framework for delivering visual search services to users of mobile devices. We show how users can point a mobile device at any location in a document, magazine or book to view related, online material on the device. We describe client applications now being deployed for the iPhone and the server architecture used for recognition of scanned images.

Categories and Subject Descriptors

[**I.4 Image Processing and Computer Vision**]: Digitization and Image Capture; [**H.5 Information Interfaces and Presentation**]: Multimedia Information Systems; [**H.3 Information Systems**]: Content Analysis and Indexing

General Terms

Algorithms, Design, Experimentation, Human Factors

Keywords

Visual search, mobile application, augmented reality, publishing, tangible interface.

1. INTRODUCTION

Over the past few years the publishing industry has slowly begun the shift from the traditional model of selling paper books, magazines and newspapers to a model that includes online editions of the same material. Devices like the Kindle, iPhone and iPad have made the consumption of digital material not only convenient but also very attractive and easy to use [1]. In the digital form, articles in magazines and newspaper and pages from documents are now being connected to online material like web pages, videos, music, etc. This trend, no doubt, increases the popularity of eBook-like material while reducing the need for printed material. At present, the number of eBook-like publications in North America is somewhere around 2.5% and is expected to double in the next three years. This leaves the vast majority of published works disconnected from the supplemental digital material that eBooks benefit from.

This is where visual search comes in. Visual search leverages the natural characteristics of an object, e.g. a page or a building or even a human face, to form a unique pattern. Snapping a picture

of an indexed object can connect users with digital information related to that object.

Figure 1. DriveTube: take picture of page; watch related video

Traditionally, applications that connect the physical world to the digital world used bar codes. We designed such a system called iCandy [2] that provided users with both an authoring and scanning environment for Mac, Windows and iPhone. Users could create cards, books and posters representing music, web sites, YouTube videos, slide shows, etc. The drawback to this method is that one must add bar codes to the pages or paste them onto objects to make them recognizable. Visual search techniques completely replace the need for bar codes since the objects themselves act, in a way, as a bar code.

2. SYSTEM DESCRIPTION

In this demonstration we will describe several visual search iPhone applications. In Figure 1 we display our *DriveTube* iPhone application and a 2-step process of retrieving online information using a book with only text and images. In step one, a user hovers over the target area on the page and takes a picture. In step 2, the image has been recognized and a video has started playing. Our applications are based on two visual search architectures: client/server and stand-alone. The client/server architecture has the advantage of simultaneously supporting many large databases in the cloud. When any information in a database changes, the changes are automatically propagated to all mobile clients since clients submit images to the cloud to be recognized. Communication with the cloud service can sometimes be a disadvantage, however, since it takes extra time to communicate with the server. Such latency delays the delivery of digital material after an image has been taken of an object.

Stand-alone systems do not have this problem since all recognition is performed on the mobile device; there is almost no delay in delivering the supplemental material. The downside to this method is in the size of the database which also affects the number of databases that can be cached on a device. In addition, when changes are made to the database, this information must be propagated to each client. We believe both methods are valuable and we are, therefore, currently experimenting with a hybrid approach to address these issues. The remainder of this paper, however, focuses on the client/server model

2.1 iPhone Apps

We implemented several iPhone applications that use our server-based visual search , two of which are shown in Figure 2: *DriveTube* and *French Rev*.
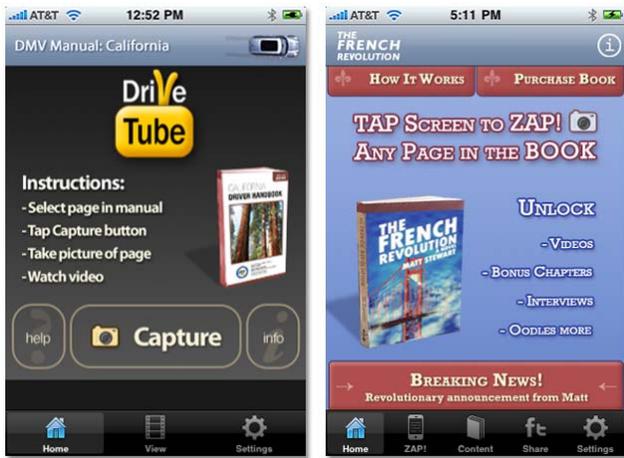

Figure 2: iPhone applications DriveTube and French

DriveTube connects the California Department of Motor Vehicles (DMV) Driver Handbook to video clips. This manual is freely available and contains relevant information for students learning to drive. The California DMV also has their own YouTube channel where they supply videos related to the material in the manual. DriveTube bridges the gap between the two resources by connecting the driver's manual to the videos. For example, while reading the section of the manual on *U-Turns*, a user can "**ZAP**" a section of the page (i.e. capture a portion of a page with the iPhone camera), as shown in Figure 1. The result is the delivery of a video to the device with information about *U-Turns*.

In a second application called "*French Rev*" we provide similar features but in the context of a novel. We have created a companion iPhone app for the novel "*The French Revolution*" by Matt Stewart (Soft Skull Press). The novel, which takes place in San Francisco and is loosely tied to the original French Revolution in 1789, contains descriptions of places, music, recipes and events. Using this companion app while reading the novel, a reader can ZAP any page to pull up online information related to the page. For instance, recipes related to the text on the page, videos showing locations discussed in the book, or map locations in San Francisco, California or Paris, France. Additional features provided in the app include timely news events related to the author that can be pushed to the client and the ability to share relevant links via Facebook and Twitter.

2.2 Server System

Our server is designed as a multi-layered system: http communication, visual search protocol, image analysis, indexing and recognition. The interface between layers is well defined enabling the system to scale and incorporate the latest technologies [3]. Using a RESTful visual search protocol built on top of http allows developers to add visual search functionality easily to existing or new applications in any language. Developers also control all aspects of the content stored in the server, including the naming of indexed images.

2.3 Authoring System

We created an authoring system that enables users (e.g. authors or editors) to add what we call "*hotspots*" to documents. Hotspots are regions of interest on a page that can have URL's and map locations assigned to them. When a user ZAPs a portion of the page connected to a hotspot, the online content associated with the hotspot is delivered to their mobile device. Because all information is stored in the cloud, changes made to the content are instantly available.

3. CONCLUSION

In this demo we presented several iPhone applications along with a server architecture that we believe will benefit both the consumer of printed material and the publishing industry. In addition to the two applications described here we are also working with several publishers to create complementary iPhone & Android applications for their printed material. We also presented an authoring environment that allows users to create hotspots.

4. ACKNOWLEDGMENTS

Our thanks to Matt Stewart and Soft Skull Press for their excellent collaboration on the French Rev project.

5. REFERENCES

[1] E-Publish or perish: The iPad and its kind are both the boon and bane for book publishers. *The Economist* (March 12, 2010), DOI= http://www.economist.com/business-finance/displaystory.cfm?story_id=15819008

[2] Graham, J. and Hull, J. J. 2008. iCandy: A tangible user interface for iTunes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy, April 07, 2008). CHI '08. ACM, New York, NY.

[3] Moraleda, J. and Hull, J. J., Toward Massive Scalability in Image Matching. *Proceedings of the 20th International Conference on Pattern Recognition*, 2010. (to appear)