

Embedded Media Marker: Linking Multimedia to Paper

Qiong Liu, Chunyuan Liao, Lynn Wilcox, Anthony Dunnigan, Bee Liew

FX Palo Alto Laboratory
3400 Hillview Avenue, Bldg. 4
Palo Alto, CA 94304
1-650-842-4854

{liu, liao, wilcox, tonyd, bee}@fxpal.com

ABSTRACT

An Embedded Media Marker (EMM) is a transparent mark printed on a paper document that signifies the availability of additional media associated with that part of the document. Users take a picture of the EMM using a camera phone, and the media associated with that part of the document is displayed on the phone. Unlike bar codes, EMMs are nearly transparent and thus do not interfere with the document appearance. Retrieval of media associated with an EMM is based on image features of the document within the EMM boundary. Unlike other feature-based retrieval methods, the EMM clearly indicates to the user the existence and type of media associated with the document location. A semi-automatic authoring tool is used to place an EMM at a location in a document, in such a way that it encompasses sufficient identification features with minimal disturbance to the original document. We will demonstrate how to create an EMM-enhanced document, and how the EMM enables access to the associated media on a cell phone.

Categories and Subject Descriptors

H.5.1 Multimedia Information Systems. H.5.2 User Interfaces. (D.2.2, H.1.2, I.3.6) H.5.4 Hypertext/Hypermedia. J.1 ADMINISTRATIVE DATA PROCESSING

General Terms

Algorithms, Design, Documentation, Human Factors.

Keywords

Augmented paper, barcode, camera phone, document recognition, marker on paper, vision-based paper interface.

1. INTRODUCTION

Although paper is one of the most widely used viewing devices, it cannot play dynamic media such as video and cannot be used to access the Web. Cell phones are commonly used to play video and access Web pages, but do not have the affordances of paper such as high resolution and readability. Current image recognition

Multimedia '10, October 25–29, 2010, Firenze, Italy.

Copyright 2010 ACM X-XXXXX-XXX-X/XX/XXXX...\$10.00.

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, or to publish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

technology can provide the best of both devices by linking digital media to paper documents. To access dynamic media associated with a paper document patch, a cell phone camera is used to capture an image of the document patch. The document patch is identified using features in the image, and digital media linked to that patch in the document is retrieved and then played on the cell phone.

Although 2D bar codes and other marks printed on documents can be easily recognized, they are visually obtrusive and may interfere with the document content layout if we want to associate them with specific content in a document. Thus many current systems rely on the document content to provide features for identification. Ricoh iCandy[1] links media to text using features based on word bounding box relations[2]. SnapTell[3] and Kooba[4] both use visual features to link to media content or the corresponding electronic document. A problem common to all these methods is in making the user aware that there is media associated with a part of the document. iCandy users pan a cell phone over the paper to find indications of linked media. SnapTell and Kooba rely on text explanation and give no specific information on the part of the paper to capture.

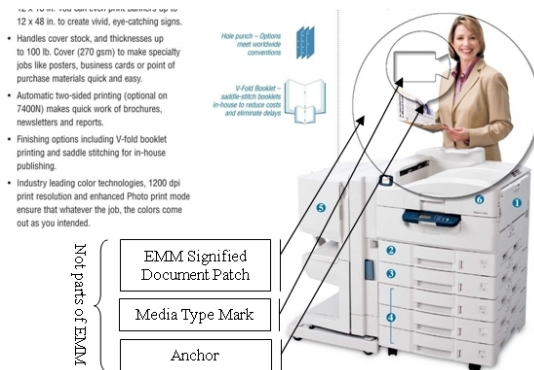


Figure 1. The circle shaped EMM printed on the page (top-right) indicates that there is associated video. (The bottom-left callouts are not parts of the EMM.)

Our approach is to augment paper with meaningful awareness-marks, called Embedded Media Markers (EMMs) that indicate the existence of media links. The design of EMMs also indicates what media type (e.g. video, web page) is linked and what part of the document should be captured. On seeing an EMM, the user knows to capture an image of the EMM-signified document patch with a cell phone in order to view associated digital media. This is analogous to Web pages that use underlines, font differences, or image tags to indicate the existence of links that users then click

for additional information. Unlike other approaches that use content features to link media, EMMs clearly indicate signified document patches and locations. Furthermore, by requiring the captured image to cover the whole mark, we can make the document retrieval more accurate and efficient.

2. EMM CONSTRUCTION

Figure 1 shows an EMM on a brochure describing a printer. The EMM signifies that there is a video associated with the brochure, and in particular with the portion of the brochure showing the magazine in the person's hand. The portion of the document inside the larger EMM feature boundary circle is the EMM-signified document patch (Figure 1). This patch should be completely included in a snapshot for successful retrieval. Within the EMM-signified document patch is the media type mark, a smaller boundary circle containing an icon that indicates the type of media associated with the EMM, in this case video. The arrow connected to the smaller circle points to the exact location in the document that is associated with the media, and is called the anchor (Figure 1).

Local features of the image within the EMM boundary circle are used to identify the document patch. We currently use FIT features[5], which are a variation on the standard SIFT features[6]. These features are derived from the local appearance of the document image and are distributed non-uniformly. If a boundary circle is positioned at a place with dense feature distribution, the circle radius R can be reduced without compromising the identification accuracy. On the other hand, a large circle is required at a location with sparse features.

We set the location and size of the EMM boundary circle so that it contains a minimum number of features for accurate patch identification. Using fewer features increases the speed of identification. A constraint we impose on placement is that the anchor point must be contained in the feature boundary circle. Also, to reduce the interference with the content by the icon-circle, our algorithm moves this circle to a place with minimum feature density within the outer feature boundary circle.

To get an optimal location and size for a feature boundary circle or an optimal location for an iconic callout circle, the system has to count the number of features inside these circles. Since this computation must be done for every possible circle center location and radius, it must be done efficiently.

We designed a fast algorithm for estimating the number of points in a circle. We approximate the number of features in a circle with the number of features in a square inside the circle. We then compute a cumulative feature distribution map, in which the value of each point is the number of features in its top-left region. Using this map, the number of features in a square can be computed by adding and subtracting the cumulative distribution at the corners of the square. Details of the algorithm are given in [7].

3. EMM SYSTEM

The authoring client is a PC application for authors to add EMMs to a document with the previously presented algorithm. With this tool, an EMM author can open a document file, specify an anchor point in a page via a mouse-click, type the URL of associated media, and get an EMM overlaid image. This tool also allows users to change the media type mark, adjust the EMM alpha-blending coefficient, adjust the document DPI, or view feature

distribution on the image. Currently, our system supports five types of media links: audio, video, text, webpage, and image. The generated EMMs and their associated multimedia URLs are uploaded to the EMM server, which indexes the marks in a

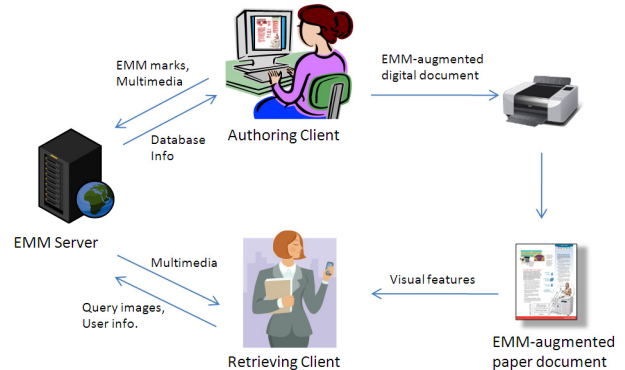


Figure 2. The architecture of the EMM system.

database and stores the link information. The resulting EMM-augmented digital document can then be printed and delivered to end users, who run a retrieving client application on a camera phone. With the application, an end user takes a snapshot of an EMM on paper to capture its visual features. The snapshot and the user information are then submitted to the EMM server to fetch the associated multimedia. Upon successful retrieval, the user can review and interact with the multimedia data.

4. EVALUATION

We are exploring a number of applications of the EMMs technology for linking media to paper. Promising directions include enhancing sales brochures by linking them to video with visual information, linking a product catalogue to the direct order web site, and linking detailed training video to manuals. Results show that EMMs provide an important capability for bridging the paper and digital world.

5. REFERENCES

1. iCandy. <http://icandy.riohinnovations.com/rocket2/>
2. Erol, B., Emilio Antunez, and J.J. Hull. HOTPAPER: multimedia interaction with paper using mobile phones. In Proceedings of ACM Multimedia'08, pp. 399-408.
3. SnapTell. <http://www.snaptell.com/>
4. Kooaba. <http://www.kooaba.com/>
5. Liu, Q., H. Yano, D. Kimber, C. Liao, and L. Wilcox. High Accuracy and Language Independent Document Retrieval With A Fast Invariant Transform. In Proceedings of IEEE ICME'09, pp. 386-389.
6. Lowe, D.G., Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vision*, 2004. 60(2): pp. 91-110.
7. Liu, Q., C. Liao, L. Wilcox, A. Dunnigan, B. Liew. Embedded Media Markers: Marks on Paper that Signify Associated Media. In Proceedings of ACM IUI'10, pp. 149-158.