Linking Digital Media to Physical Documents:

Comparing Content- and Marker-Based Tags

Tags can bridge the gap between paper and digital devices. This article explores trade-offs between two tagging technologies and issues to consider when developing systems that transition between the two media.

Scott Carter, Chunyuan Liao, Laurent Denoue, Gene Golovchinsky, and Qiong Liu
FX Palo Alto Laboratory

Tom is creating a flyer announcing his bluegrass band's performance next weekend. He puts all the significant information on the flyer—the band name, date, and time—as well as a few photos. Still, he feels the flyer doesn't do a good job of conveying the flavor of their music. He decides to upload some audio clips recorded during their last gig. He creates 2D bar codes that link to the music and places the bar codes with some descriptive labels in the flyer's margins. Satisfied, he prints his flyers and distributes them around campus.

Kathy, seeing a flyer, uses a bar code scanner application on her Droid phone to access the audio clips. In one of the flyer's photos, she spots a mandolin she's unfamiliar with. She opens another application on her Droid, snaps a photo of the mandolin, and types in a comment: "Specs?" In the background, the application scours the Web, matches the photo she took with one on a MySpace page, and automatically tags the mandolin in the online photo with her comment. Within an hour, an answer returns from MySpace: "Kentucky KM-855."

Tags can be seen as physical embodiments of their associated digital media, offering interaction points through which users can retrieve and manipulate digital content. There are generally two types of tags for linking digital content and paper documents. *Marker-based* tags, such as Tom's bar codes[1] and RFIDs,[2] employ a modification of the printed document. *Content-based* solutions, such as the application on Kathy's mobile phone, remove the physical tag entirely and link using features of the existing printed matter.[3,4]

Our goal is to explore the two tag types' use and design trade-offs by comparing our experiences developing and evaluating two systems that use marker-based tagging—DynamInk and PapierCraft[5]—with two systems that utilize content-based tagging—Pacer and ReBoard.[6] We situate these four systems in the design space of interactive paper systems and discuss lessons we learned from creating and deploying each technology.

Taking Apart Tagging Technologies

As Tom's and Kathy's experiences in the scenario show, tags can differ with respect to their visual characteristics, how they're created and inserted into a document, and the digital materials to which they link (see Table 1). Tags themselves can take a variety of forms. Marked-up Maps use RFID tags embedded in paper for document segment identification,[2] Eamonn O'Neill and his colleagues' system uses Near Field Communication (NFC) tags,[7] and Paper++ adopts conductive ink printed on paper,[8] although this

Published by the IEEE CS ■ 1536-1268/10/$26.00 © 2010 IEEE

TABLE 1
Device-based interactive paper systems' design space.*

Creation	Target	Marker based: Electronic marker	Marker based: Visual marker	Marker based: Fiduciary marker	Content based: Word geometry feature	Content based: Generic image feature
User defined	Fine			PapierCraft		Pacer
	Coarse				HotPaper	BookMarkr, ReBoard
Auto predefined	Coarse		DynamInk			
Manual predefined	Coarse	Marked-up Maps, O'Neill and colleagues' system, Paper++	Rohs' system, designable markers	iPaper, Anoto		Embedded Media Markers (EMM)

* Gray cells represent unlikely scenarios. The target granularity reflects the smallest document unit the system supports rather than its specific tagging technique; for example, Pacer allows document patches, and ReBoard could work at the pixel level.

approach can suffer from low spatial resolution and high production costs. Michael Rohs uses manually predefined optical markers, such as 2D bar codes, to indicate specific geographic regions on a paper map.[1] Through these markers, users can retrieve the associated weather forecast with a camera phone. In fiduciary-marker technologies, such as the Anoto pen, a camera recognizes small preprinted marks in the paper's background, leaving the content unaffected.

Other systems, such as HotPaper,[4] adopt a content-based approach, leveraging local word geometric features—for example, the spatial layout of word-bounding boxes in HotPaper's brick wall coding (BWC)—to identify text patches on paper. However, these systems rely heavily on text characteristics and don't work on document patches with graphic content or in languages with no clear spaces between words, such as Japanese and Chinese. Recent research makes it possible to use nearly any content in the document as a tag.[9]

On the surface, content-based tagging might appear superior because it's visually unobtrusive. However, marker-based tags provide a visual cue for retrieving tagged content and thus are often more reliable.

A tag can be created by a tag author, by a user, or automatically. For instance, 2D bar codes often require a tag author to semiautomatically insert a bar code into an electronic document or manually attach a physical bar code to a paper document or an object. Some new markers are "designable" in that creators can adjust the data cells in certain ways.[10] Moira Norrie and her colleagues proposed a framework based on the Anoto technique (www.anoto.com) wherein designers can define interactive areas, such as hot spots, within paper documents.[11] These tagged regions and supported interactions are predefined, and users can usually perform only hyperlink-clicking interaction to retrieve and play the linked multimedia. Recent advances in tagging techniques let users select an arbitrary chunk of a paper document's text or image region and apply a digital command to the associated digital media—operations equivalent to using a mouse.[9] In this new interaction paradigm, tags are created on-the-fly during user interaction.

Finally, tags can link to content at different granularity levels, ranging from the document to an individual word. The choice of tag target granularity will often depend on the supported activity. For example, Rohs' system links predefined geographic regions in maps to weather information, Bookmarkr[12] links printed pictures to their original digital content, and HotPaper[4] links text patches to digital annotations. Although tags differ on a range of features, their visual salience—whether they are marker- or content-based—impacts their design most directly.

Marker-Based Systems

DynamInk (http://fxpal.com/?p=dynamink) discovers and archives dynamic content embedded in a Web page and links to that content with bar codes. PapierCraft uses fiduciary marks recognized by Anoto pens to support digital operations applied to content in physical documents.

DynamInk

The goal of DynamInk is to enable people to print Web pages and read them away from the computer, without losing the ability to access the animations embedded in the original Web page. To achieve this goal, DynamInk detects embedded animations and automatically replaces them with a printable static image keyframe that includes a unique human-readable numeric or Quick Response (QR) code linking back to the animation's URL.

To use DynamInk, users simply drag the DynamInk bookmarklet onto their browser and click it whenever they want to print a Web page. The bookmarklet accesses the document object model to find animations including Flash and animated GIFs, discarding objects with an aspect ratio frequently

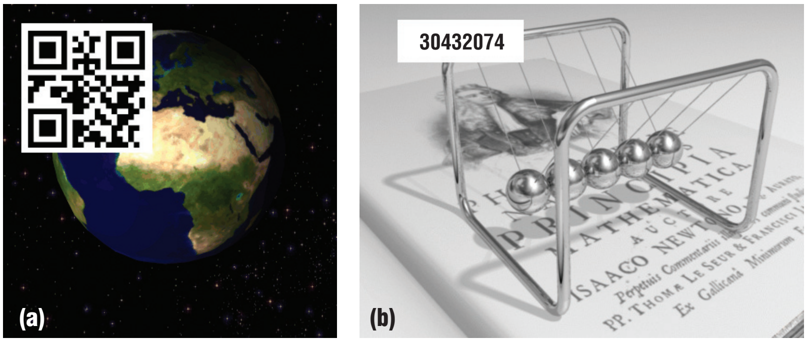


Figure 1. DynamInk discovers animations on a Web page and tags them with a code representing the animation file's location on a remote server. Images can be tagged with a (a) 2D bar code or (b) numeric tag. Participants found that the 2D bar codes were easier to use but that the numeric codes were less intrusive visually.

used by ad banners. Unlike other systems,[13] the code is overlaid on the keyframe to prevent changes to the document layout.

DynamInk includes both bar codes and human-readable codes to support different user preferences (see Figure 1). In a pilot study, we explored the tradeoffs between the two approaches, anticipating that participants would find the 2D bar codes easier to use but that numeric codes would be less intrusive.

Our findings supported these expectations, but participants found keying in codes difficult. Even though all had experience with mobile devices, they unanimously disliked typing in keys, and some refused to complete the task. However, participants had little difficulty understanding and using the bar codes. Furthermore, whereas participants did feel that the 2D bar codes were more intrusive, in most cases, even if the bar code covered most of the image, a partial representation was enough to signal the content. One participant said, "you can guess what the rest of the image is. … There's probably not a rocket ship behind that bar code."

Participants suggested other ways DynamInk could augment the static image—for example, showing the tagged animation's duration, size, and format as well as the number of views and user ratings. A few participants wanted to be able to quickly follow a printed URL by scanning it with a phone. We can extend DynamInk for that purpose by replacing URLs with a 1D bar code so that it doesn't interfere with the document layout.

Overall, the participants clearly indicated that DynamInk was useful. All but one said that they'd prefer to read documents on paper and that accessing dynamic or high-resolution content would be an important complement to their reading habits. Furthermore, participants wanted access to be immediate and unobtrusive.

PapierCraft

PapierCraft is a maker-based approach that, in some cases, could be immediate and unobtrusive. Paper offers key affordances well adapted to active-reading tasks, in which people read, understand, annotate, write, organize, and collaborate through documents. However, many powerful computer tools such as copy and paste, hyperlink, and Web search aren't usually available on paper. Systems such as Paper++[8] offer pen-based interaction with predefined "active" regions in paper documents for functions such as media browsing and form filling. However, these systems don't let users select arbitrary document segments and thus don't sufficiently support active reading tasks.

Chunyuan Liao and his colleagues built PapierCraft,[5] a pen-gesture-based command system for paper. PapierCraft lets people use a digital pen to select specific paper document content at a fine granularity—including individual words and arbitrary document regions—for various digital operations (see Figure 2a).

PapierCraft's interface consists of an Anoto digital pen and printouts augmented with special dot patterns in the background (see Figure 2b). By recognizing the patterns with a built-in infrared camera, the digital pen can identify each piece of paper and trace strokes in real time. To use the interface, the user first prints a digital document onto the augmented paper. The system automatically registers each hard-copy page with its original digital version. The user can then mark the printouts with a digital pen to manipulate the document content. For example, the user can draw a pair of cropping marks to select a figure in the printout, followed by a pigtail toward the right to specify the copy command. PapierCraft captures the strokes and the page ID and executes the command within the corresponding digital document context. Subsequently, the user can paste the figure into a separate paper note with a paste mark. PapierCraft applies all operations specified on paper to the associated digital documents, which can then be printed out for the next round of interaction.

Beyond the copy-and-paste operation, PapierCraft supports gestures for creating semantic associations between pages—the paper Web. The user can draw a hyperlink mark to create a hot spot on top of a document segment (see Figure 2c), then link it to another page specified with a hyperlink target mark (see Figure 2d). Pen tapping in the hot spot will bring up the digital version of

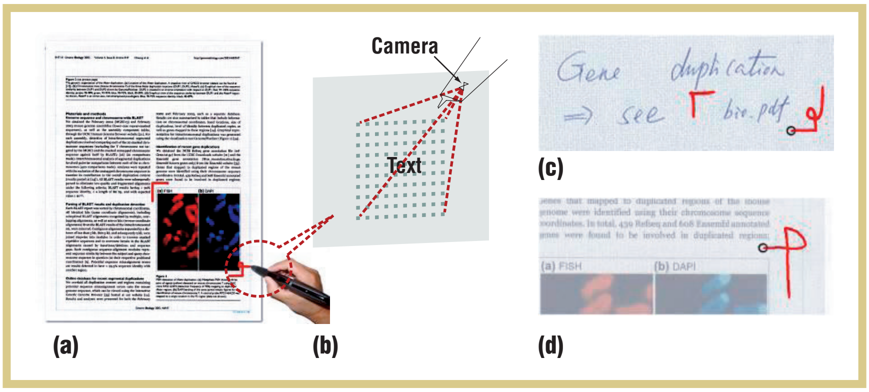


Figure 2. PapierCraft. (a) A user performs a copy command by drawing a pair of cropping marks, followed by a pigtail toward the right. (b) The Anoto digital pen recognizes inconspicuous marks on the page to determine the location of the marked area. (c) A user links two documents by first drawing a hyperlink mark for bio.pdf, then (d) drawing another mark on the hyperlink target. In this way, PapierCraft lets users establish a link that connects the physical version of one document to the digital version of another.

the linked page on a nearby connected computer, if any. PapierCraft also supports interaction between paper and the Web. Users can underline a word on a printout, followed by a Google command. PapierCraft forwards the selected word to an associated computer for Web search.

Field-biology research and e-learning have applied PapierCraft techniques successfully. Field biologists used PapierCraft marks to link multimedia data such as photos and sensor readings to paper notes.[14] College students used PapierCraft in classrooms to send their written answers from printed handouts to the instructor's computer.[15] Liao and his colleagues also conducted laboratory experiments comparing PapierCraft with Tablet PCs and normal printouts. They found that users can learn the PapierCraft gestures in a reasonable time frame (20 to 30 minutes) with a satisfactory success rate (approximately 92 percent). Users enthusiastically embraced the capability of issuing digital commands directly on paper documents. They preferred the PapierCraft interface, which provides Tablet PC-like functions without sacrificing paper-like flexibility.

On the other hand, the approach has some limitations. Because PapierCraft requires digital pens and augmented paper, it doesn't work with preexisting printouts. Also, some users worry about losing the digital pen. PapierCraft requires a carefully tuned printer to ensure that the Anoto dot patterns aren't too visually obtrusive yet are clear enough to be recognized by the pen. Furthermore, PapierCraft assumes the IDs of the Anoto sheets in the printer paper tray are consecutive, registering each printed page to the corresponding digital page. This makes PapierCraft susceptible to hardware problems such as paper jams. PapierCraft might be inconvenient for mass printing (for instance, handouts for a 60-student class) and prevents users from making last-minute modifications to the document. Although PapierCraft enables users to work across the digital/physical boundary, the special hardware requirements and a complex printing procedure might limit its usefulness.

Content-Based Systems

Pacer uses features derived from images captured by a mobile device to support fine-grained interaction with physical documents. ReBoard uses a similar technology to let users add digital tags to content written on physical whiteboards.

Pacer

Camera cell phones are becoming increasingly popular and capable but suffer from limited screen real estate and constrained input methods. To explore how paper's high display quality, relatively large display area, and spatial arrangement flexibility can complement cell phones, we designed and built Pacer (*p*aper- *a*nd *c*ell-phone-supported document *e*diting and *r*eading), a camera-phone-based interactive paper system.

Similar to many mobile interactive paper systems,[4] Pacer lets users take a snapshot of a paper document patch to retrieve and interact with the digital media associated with that patch. Unlike HotPaper and Rohs' system, Pacer features fine-grained document manipulation and rich phone gesture interactions beyond point-and-click.

As Figure 3 demonstrates, the user points a camera phone at a figure in a printout and captures a picture. When the system recognizes the picture, the user sweeps the phone over the paper document, using the screen crosshair as a cursor to sweep out a rectangular region. The gesture selects a rectangular region in the figure, which is then copied and emailed together with two automatically generated hyperlinks pointing to both the original page and the document.

Pacer's core enabling technique is the recognition algorithm for patch images captured by the camera phone (called *camera images* in this article). Pacer identifies document segments through their visual features extracted with the fast invariant transform (FIT) algorithm.[9] It performs FIT on every printed document page and stores the resulting features in a database. During user interaction, a camera image's visual features are computed and compared to those in the database. Pacer selects a stored page with the most matched features above a threshold as the digital version of the

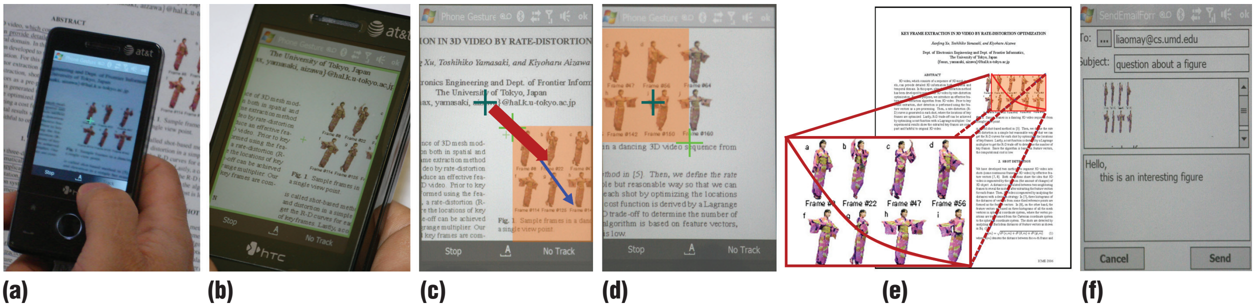


Figure 3. Pacer copy and email. (a) The user points the phone crosshair at an interesting area on paper and takes a snapshot. (b) Once the system recognizes the snapshot, it displays the corresponding high-quality version. (c, d) The user moves the phone over the paper to select (e) a region with a marquee gesture (highlighted in orange). (f) Pacer sends the selected region via email, together with hyperlinks pointing to the original page and document.

paper document. Through a coordinate transform,[3] Pacer can precisely locate a patch within the whole page and figure out the content the phone's crosshair cursor is pointing at (possibly at pixel-level precision).

Conceptually similar to the way a mouse can specify any point on a screen, Pacer lets the phone specify any point on a printed page. The system interprets continuous phone movement relative to printouts, called *phone gestures*, as mouse gestures within the corresponding digital pages.

Using the pixel-level pointing technique we described, Pacer lets users specify arbitrary document content for digital operations, including text document manipulation to augmented maps.[3] For instance, users can search for a keyword in a paper document. Upon recognizing the patch, the cell phone retrieves the document's full text and page images and highlights all keyword occurrences. By using the phone's visual guidance, users can easily flip to the pages containing the keywords. Users can also retrieve the street view of a location on a paper map and find driving directions to or from that location. Similarly, they could use a phone gesture to underline a section of a printed musical score and play it on the phone.

Pacer interaction benefits from the recognition algorithm in four ways. First, it requires no alteration of the original documents, reducing user effort and production costs. It can work on already printed documents—for example, a user can attach a new video advertisement to a poster. Second, it doesn't have the visual obtrusiveness and occlusion issues of bar codes. Third, it works on a broad range of document content types—for instance, text, graphics, and pictures—and is language independent. Finally, the approach facilitates implementing fine-grained content operations on arbitrary document contents, which are awkward when implemented with visual or electronic markers.

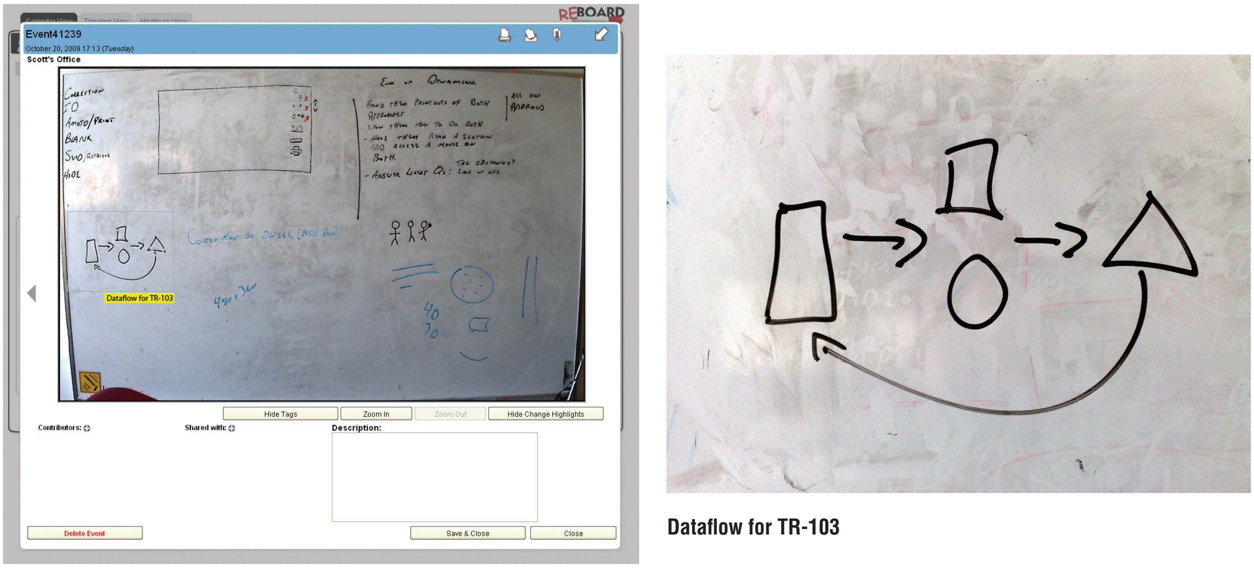
Pacer's early deployment and preliminary user studies also revealed some limitations. First, users might have difficulty locating predefined hot spots if no explicit cues are visible on the paper. (Developers have discussed similar issues with Paper++.[8]) Furthermore, recognition accuracy depends on the document patch content and camera image resolution. Pacer might fail to recognize a diagram with a large blank area. For example, like many other optical-recognition algorithms, Pacer's recognition accuracy might suffer in inadequate lighting conditions such as shadow and washout, motion blurring, curved paper surfaces, and perspective distortion. Moreover, the same content in multiple documents might lead to duplicate matching, requiring extra context information to distinguish the results.

ReBoard

Whiteboards provide an intuitive, flexible interface that's useful for a wide array of tasks. However, because whiteboards don't support archiving and reuse, people can easily lose important information. To address this issue, we built ReBoard.[6]

ReBoard couples automated capture and access with content-based tagging to archive and link ephemeral physical media (strokes on the board) to digital content. ReBoard gathers whiteboard data via a networked camera near the board. The camera takes pictures of the board when content changes. A Web UI lets users access captured images, manually snap images, and share content. The following scenario illustrates how users might interact with ReBoard.

After meeting with Ann to plan a user study, Mark returns to his office and sketches some rough ideas for the study design. He uses the Web UI to capture each new idea manually. The sketches remain on this board for a few days, but Mark eventually erases them to make room for a different project. A week later, Ann wants to hold another meeting in a public conference room. Mark uses the Web UI to find and print his study design sketches. During the meeting, Ann uses these printouts as a reference while they sketch some new thoughts on the whiteboard. Afterward, Ann uses the Web UI to email Mark im-



(a) (b)

Figure 4. ReBoard. Users can tag any content on any whiteboard. Here, a user has taken a picture with a camera phone of a diagram on the board and added the tag "Dataflow for TR-103." (a) After the user sends the image and tag to the system, ReBoard discovers the matching whiteboard image and associates the tag with the matching region in that image. (b) Clicking on the region launches a popup window showing the higher-resolution cell phone image.

ages of the new sketches that ReBoard captured automatically.

As this scenario illustrates, ReBoard not only supports manual and automatic capture but also lets users browse and search captured content and export media to other applications. Whereas automated-capture systems such as ReBoard can record changes to a whiteboard in a conference room, they don't augment the whiteboard itself. However, with content-based tagging technologies, users can annotate physical whiteboard content with digital media.

We built an extension to ReBoard that uses image-matching algorithms based on scale-invariant feature transform (SIFT) to locate snapshots in their matching overviews (see Figure 4). Users can snap a detailed picture of a sketch on the board and use a simple interface to upload the image and an associated comment. The system searches its database for an appropriate match using a procedure similar to Pacer's. When it finds the matching overview image, it links the comment to the sketch's location in the overview image and sends an email notification to the user. Users can access the image and comment at any time via the Web UI.

In our early experiments deploying 16 cameras, we could accurately locate multiple different sketches ranging from roughly 3 to 12 inches in diameter. We currently identify the best match using only the quality of match, but to scale the approach, we could use other metadata as well, such as the overview's time of capture and detail images. Similar to other work that relies on SIFT-based features, the system's main limitation is that it's time intensive. Still, it's a promising approach for asynchronous capture and access tasks.

Design Considerations for Tagging-Based Applications

On the basis of our experiences designing and using different tagging technologies, we found a set of issues that designers should keep in mind when developing interactive physical-digital linking systems. Table 2 summarizes the characteristics of typical tagging techniques. Designers should first prioritize their system requirements and then refer to the table to choose an appropriate technique.

Choosing a Tagging Technique

The technique you choose depends partly on the following 11 considerations.

Document type. Nontext documents, including graphics, pictures, and maps, don't work with word-geometry-based tagging techniques such as brick wall coding.[4] Similarly, users can't use a SIFT- or FIT-based approach to tag blank page margins because there are too few visual features. In contrast, marker-based approaches are usually immune to these issues.

Spatial density of tags. If the spatial density is high—for instance, one tag for every word in a report—laying out

TABLE 2
Tagging technologies compared.*

	Electronic marker	Visual marker	Fiduciary marker	Word geometry feature	Generic image feature
Typical tool or algorithm	RFID	Quick Response (QR) codes	Anoto	Brick wall coding (BWC)	Fast invariant transform (FIT), scale-invariant feature transform (SIFT)
Tagged document types	Generic	Generic	Generic	Text documents	Generic
Effort and cost to create and insert tags	High	Middle	Middle	Low	Low
Computational complexity	Low	Low	Low	Middle	High
Typical user hardware	RFID cell phone	Camera cell phone	Digital pen and cell phone or PC	Camera cell phone	Camera cell phone
Support for visual guidance	No	Yes	No	No	No
Support for arbitrary user-defined tags	No	No	Yes	Yes	Yes
Spatial density of tags on paper	Low	Low	High	High	High
Encoded data capacity	Low	High	High	None	None
Maximum interaction distance	Close	Far	Very close	Far	Far
Robustness for commercial products	High	High	High	Limited	Limited
Scalability for commercial products	High	High	High	Limited	Limited

* All ratings are relative. The gray cells indicate limited theoretically proven results, although some commercial products are emerging.

electronic or visual markers might be difficult or impossible because of their physical size and the between-tag spacing necessary for reliable recognition. Fiduciary marks and content-based tagging can prevent this problem because they don't interfere with layout.[3]

Tag definition flexibility. Unlike marker-based tagging, content-based tagging lets users define tags on-the-fly—for example, copying an arbitrary image region or adding content to any user-specified geographic region on a map. However, as we experienced with Pacer, content-based tagging can be prone to usability issues caused by imprecise image registration, hand jitter, and undesirable lighting conditions. Integrating other input sources such as touch screens and accelerometers can alleviate these issues.

Tag appearance. As we found with DynamInk, marker-based tags can sometimes occlude key areas of the tagged document. Higher camera resolutions can reduce the size of marker-based tags. Furthermore, tags with less visual obtrusiveness, such as embedded electronic markers, fiduciary markers, or content-based tagging techniques, can mitigate occlusion. This consequently improves the aesthetic experience of some tasks—for example, tagging a region in the middle of a map.

However, when a tag is unobtrusive (especially a content-based one), users might have difficulty finding it. Such cases require appropriate guidance for tag discovery, such as special translucent overlays on paper or dynamic information rendered on cell phones. New tagging techniques could mash up markers with content-based approaches to better balance visual guidance and document interference. Our recent research with Embedded Media Markers (EMM) overlays a semitransparent marker to signify the existence of a content-based tag.[16]

Amount of target data. If the target content amount is under a few hundred bytes—for instance, as with package shipping information—you can encode the data directly into a 2D visual marker. In contrast, for target data such as video and audio, you can encode only the URL pointing to the data in the tags; the actual data must be stored on a separate content server.

Interaction distance. If the tagged content is beyond the user's reach, such as on a public poster on a wall, visual markers and content-based tags are more appropriate because users can take pictures of the tags at a distance. On the other hand, the Anoto digital pen starts tracking only when its tip firmly touches a hard surface. Similarly, electronic tags' operating range is usually a few inches.

Tag creation cost. Electronic markers require special devices embedded in the tagged document and cost more than other tagging techniques. Visual and fiduciary markers need no special devices; you can produce them with off-the-shelf printers. However, tag authors often must manually attach visual markers to the tagged document. Even when the visual markers are automatic, such as in DynamInk, tag authors might still need to install tag-creation software. Content-based tagging avoids all these costs.

Tag recognition cost. Normal cell phones can recognize and decode normal electronic markers, such as RFID and visual markers, with no extra hardware cost to users, and some content-based tagging techniques (for example, word block coding) can be performed on a mobile device. However, more complex algorithms such as FIT and SIFT require a separate computer for fast computation, increasing the infrastructure cost. The Anoto technique requires a special digital pen.

Infrastructure. An interactive paper system usually includes tags, a user interface, a decoding service, and an associated digital content service. For applications that require self-contained tags, such as package tracking, visual markers are good candidates because content—such as sender and receiver information—can be encoded directly in the tag, and the decoding service can run locally on the device. To support more data as well as high-complexity decoding such as FIT and SIFT, the content and decoding services must run on a dedicated computer.

Tag scalability. Because marker-based tags are independent of one another, they can scale easily. On the other hand, content-based recognition accuracy might decrease with an increasing number of tags having similar visual features.

Research on this topic is still in its early stage. EMM achieved 99 percent accuracy with 2,188 indexed images and 110 testing snapshots.[16] Beat Fasel and Luc Van Gool exploited Speeded-Up Robust Features (SURF) for art object recognition and achieved approximately 90 percent accuracy with 205 indexed images and 116 testing images.[17]

One solution might be to partition the tag feature database and ask users to specify a subdatabase for query images. Future systems could also encode some unobtrusive binary information in the document to redirect query images to different servers managed by a scalable computing platform.

Robustness. Typical tagging techniques such as RFID, bar codes, and QR codes are mature and commercially available. In contrast, current content-based tagging techniques are susceptible to bad lighting conditions, shadows, and hand jitter, and they rely heavily on the original document content for recognition. This leads to issues such as duplicate matching or insufficient features.

Better tools could ameliorate some of these concerns by capturing and exploiting contextual information. Examples include using a video sequence rather than only one frame and capturing an additional overview shot with a separate wide-angle camera. At this point, though, content-based tagging's product-level robustness is unclear.

> Cell phones can recognize and decode normal electronic markers, such as RFID and visual markers, with no extra hardware cost to users.

Other Considerations

Of course, designing a tagging-based application goes beyond tag choice. Creating a potentially multidevice tagging-based system that's easy to understand and control requires dealing with the following three concerns.

An interaction paradigm. In most situations, interacting with tags is straightforward: users point a device at a tag and follow the link. However, for more complex tagging-based applications, supporting only point-and-click might be insufficient. Thus, Pacer and PapierCraft support a more complete range of GUI operations, such as fine-grained and user-specified document content selection and user-determined action selection. These operations aim to bring traditional screen-based interaction to physical paper interfaces. However, some applications might require more tailored approaches.

Interaction feedback. As the number of operations supported by a tagging-based application increases, so does the importance of user feedback. For point-and-click operations, a simple audio notification might suffice to indicate a recognized tag. To convey more advanced features, a more complex system will need to use the affordances of individual devices and modalities including visual, auditory, and tactile feedback. For instance, an overhead projector or a display on a mobile device could support full-text word search on paper. However, introducing more complex hardware might compromise paper's inherent flexibility. Designers must make a trade-off for their specific application.

Cross-device interaction. Once the target content is retrieved, it must be rendered. In some cases, rendering the content on the same device that recognized the tag is possible. In other situations, it's not—for example, using an Anoto pen to follow a link to a video. Designing interactions that span devices requires careful coordination of feedback, control, and display. For instance, a system spanning a cell phone and paper can briefly vibrate a digital pen to indicate that a video clip has finished downloading and is ready to watch on the cell phone. In this way, the system exploits different modalities to limit task interruption.

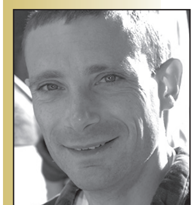
Marker- and content-based tags are each useful in different situations. Recall our earlier scenario in which Kathy read a flyer that Tom had created for his band. For Tom, creating a marker-based tag was preferable because it's an immediately recognizable link to additional content. He also had control over the document's layout. Kathy, on the other hand, interacted with the document as a user and needed to use a content-based tag to add her question about the mandolin.

There are many fields of tagging techniques to explore. Our future research will focus on combining the best of both techniques to create tagging technologies and applications that are robust, flexible, and scalable. Specifically, we are optimistic that EMMs can offer much of the flexibility of content-based tags while matching marker-based tags' robustness. We're currently investigating authoring tools to make EMMs easier to deploy and reuse and will investigate new infrastructures, such as cloud computing platforms, to support Web-scale image search for content-based tagging as well as digital watermarking techniques that encode information in EMMs to improve load balancing and personalization or, in some cases, eliminate the need for image search altogether. We're also exploring new image feature descriptors with high accuracy and relatively low computation complexity for mobile applications. At the application level, we plan to incorporate other mobile devices, such as portable projectors, into Pacer as well as research mobile augmented reality with more complex input and feedback mechanisms.

the AUTHORS

Scott Carter is a research scientist at FX Palo Alto Laboratory. His research interests include human-computer interaction, ubiquitous computing, and mobile computing. Carter has a PhD in computer science with an emphasis in human-computer interaction from the University of California, Berkeley. He's a member of the ACM. Contact him at carter@fxpal.com.

Chunyuan Liao is a research scientist at FX Palo Alto Laboratory. His research interests include human-computer interaction, ubiquitous computing, and multimedia processing. Liao has a PhD in computer science from the University of Maryland. He's a member of the ACM and IEEE. Contact him at liao@fxpal.com.



Laurent Denoue is a senior research scientist at FX Palo Alto Laboratory. His research interests include human-computer interaction, mobile computing, and multimedia processing. Denoue has a PhD in computer science from the University of Savoie. He's a member of the ACM. Contact him at denoue@fxpal.com.

Gene Golovchinsky is a senior research scientist at FX Palo Alto Laboratory. His research interests include interaction design, information-seeking interfaces, pen-based computing, and hypertext. Golovchinsky has a PhD in human factors from the University of Toronto's Department of Mechanical and Industrial Engineering. He's a senior member of the ACM. Contact him at gene@fxpal.com.

Qiong Liu is a senior research scientist at FX Palo Alto Laboratory. His research interests include paper user interfaces, infrared-camera-based vital-sign detection, camera control, teleconferencing, telecontrol, and autonomous learning robots. Liu has a PhD in computer science from the University of Illinois at Urbana-Champaign. He's a member of the ACM and a senior member of IEEE. Contact him at liu@fxpal.com.

REFERENCES

1. M. Rohs, "Real-World Interaction with Camera Phones," *Ubiquitous Computing Systems*, LNCS 3598, Springer, 2005, pp. 74–89.
2. D. Reilly et al., "Marked-Up Maps: Combining Paper Maps and Electronic Information Resources," *Personal Ubiquitous Computing*, vol. 10, no. 4, 2006, pp. 215–226.
3. C. Liao et al., "Pacer: Fine-Grained Interactive Paper via Camera-Touch Hybrid Gestures on a Cell Phone," to be pub-

lished in *Proc. 28th Int'l Conf. Human Factors in Computing Systems* (CHI 10), ACM Press, 2010.

4. B. Erol, E. Antunez, and J.J. Hull, "HotPaper: Multimedia Interaction with Paper Using Mobile Phones," *Proc. 16th ACM Int'l Conf. Multimedia*, ACM Press, 2008, pp. 399–408.

5. C. Liao et al., "PapierCraft: A Gesture-Based Command System for Interactive Paper," *ACM Trans. Computer-Human Interaction*, vol. 14, no. 4, 2008, pp. 1–27.

6. G. Golovchinsky, S. Carter, and J. Biehl, "Beyond the Drawing Board: Toward More Effective Use of Whiteboard Content," 2009; http://arxiv.org/ftp/arxiv/papers/0911/0911.0039.pdf.

7. E. O'Neill et al., "Reach Out and Touch: Using NFC and 2D Barcodes for Service Discovery and Interaction with Mobile Devices," *Pervasive Computing*, LNCS 4480, Springer, 2007, pp. 19–36.

8. P. Luff et al., "Augmented Paper: Developing Relationships between Digital Content and Paper," *The Disappearing Computer*, LNCS 4500, Springer, 2007, pp. 275–297.

9. Q. Liu et al., "High Accuracy and Language Independent Document Retrieval with a Fast Invariant Transform," *Proc. Int'l Conf. Multimedia and Expo*, IEEE Press, 2009, pp. 386–389.

10. E. Costanza and J. Huang, "Designable Visual Markers," *Proc. 27th Int'l Conf. Human Factors in Computing Systems* (CHI 09), ACM Press, 2009, pp. 1879–1888.

11. M.C. Norrie, A. Palinginis, and B. Signer, "Content Publishing Framework for Interactive Paper Documents," *Proc. ACM Symp. Document Eng.*, ACM Press, 2005, pp. 187–196.

12. N. Henze and S. Boll, "Snap and Share Your Photobooks," *Proc. Int'l Conf. Multimedia and Expo Multimedia*, IEEE Press, 2008, pp. 409–418.

13. S.R. Klemmer et al., "Books with Voices: Paper Transcripts as a Physical Interface to Oral Histories," *Proc. Int'l Conf. Human Factors in Computing Systems* (CHI 03), ACM Press, 2003, pp. 89–96.

14. R.B. Yeh et al., "ButterflyNet: A Mobile Capture and Access System for Field Biology Research," *Proc. Int'l Conf. Human Factors in Computing Systems* (CHI 06), ACM Press, 2006, pp. 571–580.

15. C. Liao et al., "PaperCP: Exploring the Integration of Physical and Digital Affordances for Active Learning," *Human-Computer Interaction—Interact 07*, LNCS 4663, Springer, 2007, pp. 15–28.

16. Q. Liu et al., "Embedded Media Markers: Marks on Paper that Signify Associated Media," to be published in *Proc. 2010 Int'l Conf. Intelligent User Interfaces* (IUI 10), ACM Press, 2010.

17. B. Fasel and L.V. Gool, "Interactive Museum Guide: Accurate Retrieval of Object Descriptions," *Adaptive Media Retrieval: User, Context, and Feedback*, LNCS 4398, Springer, 2007, pp. 179–191.