Telling a Story on a Tag: The Importance of Markers' Visual Design for Real World Applications

Enrico Costanza[1]
MIT Media Lab
77 Mass Ave
Cambridge, MA
USA
enrico@media.mit.edu

Mirja Leinss[2]
Harvard GSD
48 Quincy St
Cambridge, MA
USA
mirja@post.harvard.edu

ABSTRACT
Tag recognition can be a convenient and quick information access method for mobile applications. Tags act like URLs allowing users to leave and retrieve digital information associated to physical objects or locations. Visual tag can be recognized using standard digital cameras, now common on commercial mobile phones. This paper demonstrates an application of d-touch, a visual tag recognition system running on Symbian OS phones that leaves considerable freedom in the graphic design of the tags. The design process for visual tags to be used in a urban Ubicomp project is presented, highlighting the importance of the graphic design of tags for real-world applications.

Categories and Subject Descriptors
H.5 [**Information Systems**]: Information Interfaces and Presentation. J.5 [**Computer Applications**]: Fine Arts.

General Terms
Algorithms, Design, Experimentation, Human Factors.

Keywords
d-touch, Electronic Lens, Topological Recognition, Fiducial Recognition, Marker Visual Design.

1. INTRODUCTION
Mobile applications can provide ubiquitous information access through the increasing connectivity and multimedia capabilities of mobile phones and PDAs. While output capabilities of these devices are constantly improved (improved audio quality, high resolution and high contrast displays and even display peripherals embedded in eyeglasses and mini video projectors), input often acts as a bottleneck. Both the limited size of keyboards and the conditions of use ("on the move"), make text entry often problematic, especially for text not based on dictionary words such as URLs.

When it is possible to associate information, or more generally interaction, with a specific physical object or location, automatic *tag recognition* can be used as an efficient and fast input method. The idea is to have a set of tags – each with a unique identifier – that users can *scan* or *read* through their mobile devices; each tag identifier can then act like a generic URL, allowing users to retrieve or leave digital information related to the tagged object or location. For example, tags can be associated to artworks exhibited in a museum, items for sale in a shop, monuments or points of interest in a city [5]. Different technologies can be used for tagging, the most common are Radio-Frequency ID (RFID) and visual recognition based on computer vision.

RFID tags can be recognised by dedicated readers that emit a radio signal, generally in the MHz range, and "listen" to responses to this signal by nearby tags [4]. Readers can have range from few centimetres to metres. Because of the properties of radio propagation, RFID tags do not need to be at line-of-sight with the reader, so they can be hidden under a surface and made invisible to the eye. This property can be advantageous in some circumstances, such as for scanning all items getting in and out of a storage warehouse, or when the presence of a tag can be taken for granted (for example if *all* items in a shop are tagged). However, in more generic situations the presence of RFID tags still needs to be made evident to users through visual signs [6]. RFID readers are embedded on some phone prototypes, but are not yet common on commercial devices, probably due to the specificity of their use.

Visual tags, also referred to as *markers*, are graphic symbols that can be read by a standard digital camera, like the ones embedded in camera-phones, and decoded through a computer vision algorithm that can either run locally on the phone or on a remote server. Because the system uses the visible spectrum, to decode a visual tag the camera has to be pointed directly at the symbol and this has to be visible to the human eye. Several systems for decoding visual tags on commercial mobile phones are available, both commercially and as output of academic research [7, 9, 10].

In all of the existing systems, information is encoded in the geometry of the markers (i.e. their shape), as a consequence the markers visual design is determined solely by the system specific encoding algorithm, without taking into account any aesthetic criteria.

2. D-TOUCH: TOPOLOGY-BASED VISUAL TAG RECOGNITION
d-touch is a visual tag recognition system that encodes information in the topological structure of its markers, rather than

[1] The author is now with the Media and Design Lab, Ecole Polytechnique Fédérale de Lausanne, Switzerland.

[2] The author is now with the MIT Design Lab, Cambridge, USA

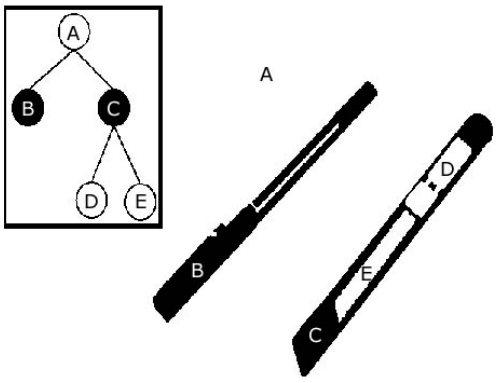

Figure 1. An example region adjacency tree. Regions in the scene (on the right) correspond to nodes of the tree (on the left).

their shape (i.e. their geometry) [1,2]. As a consequence, the visual design of the markers is less constrained than many other visual marker systems, and it can be governed by aesthetic principles: form and function, rather than just function.

The topology of d-touch markers is considered in terms of their adjacency structure. The markers are composed of black and white regions (*connected components*) and the adjacency structure of these regions (i.e. the way the regions are *nested*) can be used as an identifier. The adjacency information is stored in a region adjacency tree, as illustrated for a generic scene in Figure 1 and for one marker in Figure 2. Each node of the tree corresponds to a connected region, two nodes are connected by an arc if and only if the two corresponding regions are neighbouring. Each tree can be represented as a string of integers [1], conveniently usable as a marker ID. Figure 3 shows two markers that are topologically equivalent to that of Figure 2, demonstrating how the system leaves freedom in the visual design of the markers. Figure 4 illustrates another marker "hidden" in text, together with its topological structure.

The markers can be decoded purely through their topology, but it is optionally possible to also take into account their geometry. This option can produce a larger number of different identifiers. For example even though the two markers of Figure 3 are

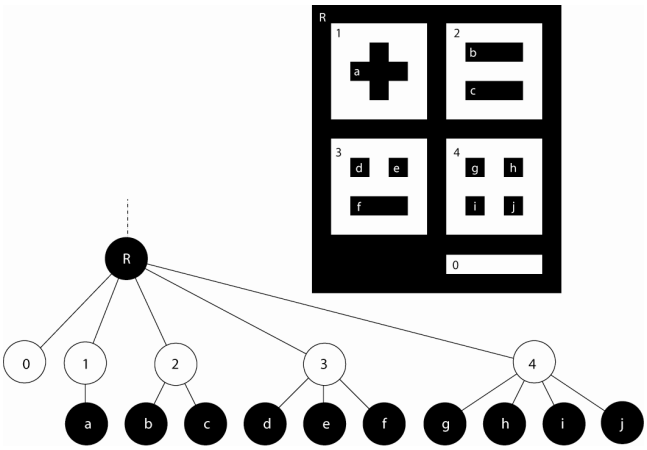

Figure 2. A d-touch marker with the corresponding region adjacency tree.

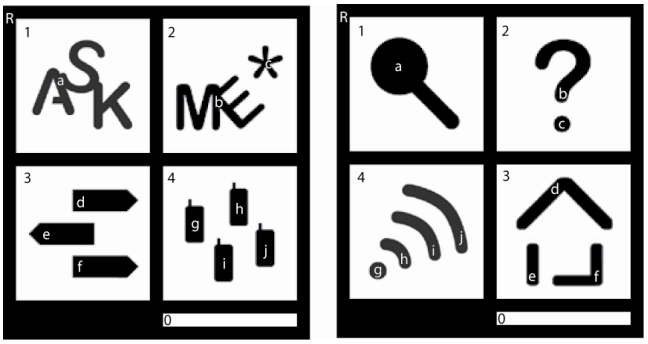

Figure 3. Two markers with the same topology as the one in figure 2.

topologically equivalent, they can be read as 1234 (left) and 1243 (right), where the numbers correspond to the number of black regions inside each white square (reading left to right, top to bottom). Using geometry it is possible to define 24 different markers of this type. It must be underlined that this is just an example and that larger number of different identifiers can be available with d-touch.

The system was originally developed in C++ for desktop applications under Linux and MS Windows. It was later ported to the Symbian OS mobile platform. d-touch was designed for real-time graphics and audio interactive applications, it uses little or no floating point operations (depending on whether or not the marker geometry is taken into account). This characteristic makes it ideal for embedded devices without FPU, such as most commercial mobile phones. d-touch is available under GNU Public Licence on http://sourceforge.net/projects/libdtouch/.

3. CONTEXT: THE ELECTRONIC LENS PROJECT

d-touch was used in the Electronic Lens, a social networking application developed at the MIT Media Lab [3]. The application enables users to create location based discussion and communities. Citizens can share information and opinions – which are related to a specific place in their city – and participate

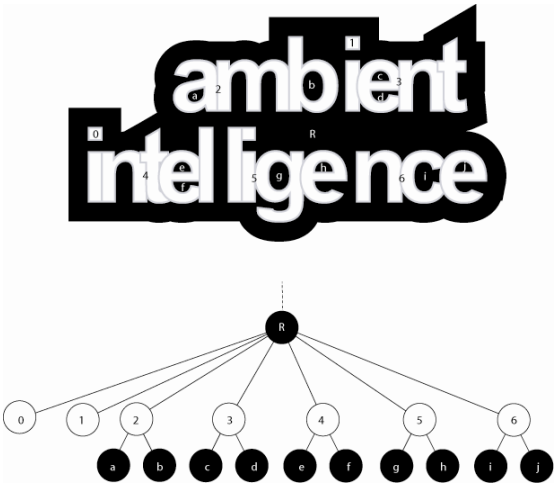

Figure 4. A marker "hidden" in type, and the corresponding region adjacency tree.

in discussions of public interest in a democratic way.

Private or more restricted discussions are also possible by setting up personalized groups. When leaving a message, a user can decide who will be able to access it. The Electronic Lens incorporates several components: the d-touch tag recognition system, the RadioActive voice chat system [11], the xLink context manager system [8] and other ones currently under development.

The Electronic Lens provides a tangible experience of the physical environment and local communities. It creates the opportunity to gain a new perception of the city, and encourages participation in civic decision making process. The application can function as a guide, as a local story-teller and as a way to network people with the same interests. Personal opinions as well as corporate information can be part of a tag and its related content. Within the project, visual tags function as a landmark, around which the interaction takes place.

The project in its initial phase was tested in April 2006 in Manresa, Spain with the support of the Generalitat de Catalunya (the Catalan regional government) and the Ajuntament de Manresa (city council). In a workshop setting, a group of 16 local high school students tagged building in their city and used the mobile device application during one week to explore new ways of communication.

4. DESIGN PROCESS

The Electronic Lens project was promoted and supported by the Generalitat de Catalunya, the Catalan regional government which somehow took the role of a client in terms of expectations and

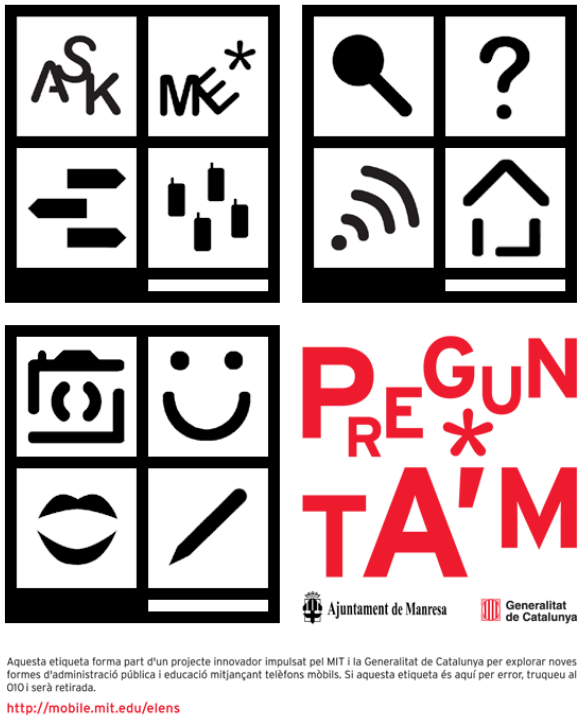

Figure 5. A tag for the Electronic Lens project. It includes 3 d-touch markers and other graphic elements.

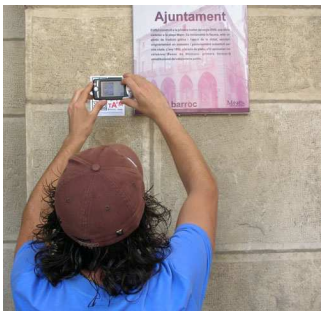

Figure 6. A tag in use during the trial in Manresa.

ideas of the project. The interest of the institution was to examine how mobile and location based technology can improve the mutual relationship between citizens and government. Because of the official endorsement, the tags needed to be perceived as formal, authorized items that add value to the objects they are placed on – rather than polluting the environment as ordinary stickers could do. Part of the project involved placing tags on the façades of private buildings, this caused special concern, as the owners should feel honored by having their walls tagged. Similar to a seal, the tag should stand for a place of interest and special distinction and attract pedestrians and users in a visually appealing way. With the institution logotype of the Generalitat on the final design, the tag got even an official signature.

When designing the physical tag for the Electronic Lens project, the aim was to tell a story by transforming the markers into sets of icons. Supplemental text, illustration and logotypes that had no technical functionality were also used to complete the design from an informational and aesthetic point of view. The design includes icons for expressing one's opinion, for leaving one's view of things through pictures, for a networked city with information accessible through mobile devices, and for the pleasure that comes up with exploring what potentially lies quiet behind this physical marker.

The design is visually coherent and the technical functionality could be completely *hidden* into it. All the technical constraints were fulfilled without any visible detriment as a result of an iterative optimization process. The final design included 3 sets, each of 4 icons, each set embedded in a d-touch marker. From the technical point of view, each "Electronic Lens tag" includes 3 d-touch markers of the type shown in Figures 2 and 3 – because each marker can assume 24 different values it is possible to define more than 12000 different tags (to simplify the processing only *sets* of markers are considered, not *permutations*), enough for the project that required only 500. The logotype of the government and the red pregunta'm ("ask me" in Catalan) did not have a technical function but are essential part to communicate the project and its official aspect (Figure 5).

Different types of tags were iteratively designed – evolving from a pure, abstract data code (Figure 2) to a carefully crafted design, with an "invisible" data code that leaves space for communication graphics.

The functional aspects of the tag constrain the design to the use of high contrast, positive and negative planes and modularity, as well as use of matte materials because of camera recognition issues with reflective surfaces. Still, d-touch is a flexible visual tag

recognition system that allows a wide range of personalization by encoding information in the topological structure of the markers.

The choice of the material for the tags was also object of research. The aim was to create a high-quality flexible label, that could be placed on curved surfaces (e.g. lamp-posts), water resistant, removable but definitely easy to stick on rough textured surfaces as sandstone or on shiny surfaces as plastic. The tags were printed with an ink-jet plotter on a water-resistant heavy-weight vinyl, with an ink-jet plotter. They were attached to buildings with an easily removable adhesive, applied by the user the moment he or she places the tag. The characteristics of the adhesive allowed to fix the tag on practically any type of surface.

5. FURTHER DEVELOPMENT AND FUTURE APPLICATIONS

As discussed in the previous section, the design process described in this paper was a collaborative effort and required technical understanding of the recognition algorithm.

Further development will aim at enabling the design of d-touch markers with minimal understanding of the underling technology, for example through the realization of a drawing software (or even better a plug-in for an existing application) that can interactively guide the design of functional markers.

Personalization and appropriation play a strong role in the mobile devices market: users customize their phones with screen backgrounds, phone shells and ring-tones, generating high revenue volumes. Based on this, it can be interesting to allow end-users to design or modify tags, for example through a web-application. The personalized tags can then be printed by users themselves with standard consumer-grade printers and be affixed to objects or buildings, in order to annotate them with digital information. This process also resonates with the phenomenon of *graffiti tagging*: spray-painting their initials or pseudonym on wall individuals or groups mark and appropriate the urban territory.

Because d-touch can be deployed on a large number of commercial phones without custom hardware, there is potential for large scale tagging trials. It would be possible to equip a a medium size community with camera phones and study how individuals take advantage of the ability of leaving messages in specific locations or attached to specific objects. The results could then be easily generalized to other tagging technologies. Potential is also foreseen for printed paper: adding a visual tag to a printed magazine or poster involves absolutely no additional cost and it can allow publishers to more easily link on-line resources (multimedia material, discussion forums, updates) to the printed content. The possibility to adapt the graphic aspect of the markers is very important in a market where aesthetic plays a major role.

6. CONCLUSION

This paper presented the design of a set of visual tags that can be read through a computer vision system but at the same time tell a story through their visual iconic aspect. The design was possible thanks to the use of d-touch a recognition system that leaves considerable freedom for the geometry of its tags.

Liberating the design of visual tags from being data codes, and making it a visual medium that allows to tell a story or give a glimpse of what is hidden behind the tag is a very powerful way to improve the user's experience. The personalization of visual tags offers an additional channel to communicate the identity of multimedia projects.

7. ACKNOWLEDGEMENTS

The authors would like to acknowledge the Electronic Lens team, lead by Bill Mitchell and Federico Casalegno, the Generalitat of Catalunya and the Ajuntament de Manresa.

d-touch was initially developed while Enrico Costanza was at the University of York, England, under the academic supervision of John Robinson, and at Media Lab Europe, Ireland, under the direction of Rebecca Allen.

8. REFERENCES

[1] Costanza, E., Robinson, J., "A Region Adjacency Tree Approach to the Detection and Design of Fiducials", in *Proc. of Vision, Video and Graphics*, 2003, Bath, UK, July 2003

[2] Costanza, E., Shelley, S. B., Robinson, J., "d-touch: a Consumer-Grade Tangible Interface Module and Musical Applications", in *Proc. of Conference on Human-Computer Interaction (HCI03)*, Bath, UK, September 2003.

[3] Electronic Lens website: http://mobile.mit.edu/elens.

[4] Feldman, A., Munguia Tapia, E., Sadi, S, Maes, P., Schmandt, C., "ReachMedia: On-the-move interaction with everyday objects," *in proc.* ISWC'05, Osaka, Japan, October 2005.

[5] Kindberg, T., Barton, J., Morgan, J., Becker, G., Caswell, D., Debaty, P., Gopal, G., Frid, M., Krishnan, V., Morris, H., Schettino, J., Serra, B., and Spasojevic, M. 2000. People, places, things: Web presence for the real world. In *Proc. 3rd IEEE Workshop on Mobile Comp. Sys. and App. (Wmcsa'00)* Washington, USA, December, 2000

[6] Riekki, J.; Salminen, T.; Alakarppa, I., "Requesting Pervasive Services by Touching RFID Tags," *Pervasive Computing, IEEE* , vol.5, no.1pp. 40- 46, Jan.-March 2006.

[7] Rohs, M., Gfeller, B. "Using Camera-Equipped Mobile Phones for Interacting with Real-World objects." In: Advances in Pervasive Computing, Austrian Computer Society (OCG), ISBN 3-85403-176-9, pp. 265-271, Vienna, Austria, April 2004.

[8] Sadi, S., Maes, P. "xLink: Context Management Solution for Commodity Ubiquitous Computing Environments." In UBICOMP Workshop ubiPCMM: Personalized Context Modeling and Management for Ubicomp App. Tokyo, Japan, September 2005.

[9] Scott, D., Sharp, R., Madhavapeddy, A., and Upton, E. "Using visual tags to bypass Bluetooth device discovery". *SIGMOBILE Mob. Comput. Commun. Rev.* 9, 1 (Jan. 2005)

[10] Semacode webite: http://semacode.org

[11] Zinman, A., Donath, J. Navigating Persistent Audio. *In Proc. CHI 2006*, Montreal, Canada, April 2006.