Mobile Lenses: a Hybrid Approach to Direct Interaction with Maps and Kiosks

Derek Reilly and Huiqiong Chen

Dalhousie University, Faculty of Computer Science, 6050 University Avenue,
Halifax, Nova Scotia, Canada, B3H 1W5
`{reilly, huiqiong}@cs.dal.ca`
`http://www.edgelab.ca/`

Abstract. In this paper we discuss ongoing work considering ways of interacting with paper maps, posters and static kiosks using mobile devices. A technical implementation permitting lens-like interaction with maps is presented, employing a combination of short-range RFID and a smartphone camera. The approach is promising for truly mobile, ad-hoc interaction with paper maps and similar artefacts.

1 Introduction

A key technique in "pervasive mobile interaction device" research is to link the virtual and real worlds by hovering over, pointing at, or gesturing toward physical artefacts with a handheld device. Often the handheld device acts as a window on the virtual world [6], allowing the user to constrain their real-world interactions and gather feedback. Additionally (or alternately) the physical artefact itself may provide feedback and a means of constraining interaction, as when the artefact is a large interactive display. Examples of this paradigm exist already in our daily lives: in increasingly sophisticated remote control devices, and in vision-based technologies such as the standard barcode reader, and smartphone capture of '2-d barcode' glyphs, which is quite prevalent in Japan.

Our work considers how mobile interaction devices can be used to interact with large, static information artefacts such as paper maps and information kiosks. In this paper, we present a hybrid Radio Frequency Identification (RFID) and vision-based approach that permits interaction with maps using the handheld as a mobile lens [1,5]. Importantly, the technique permits direct interaction with the visual features of the resource. For example, the user can select a map icon by centering it in the visual field of a smartphone camera. Direct interaction can rely primarily on attributes of a map's visual presentation. Previous work has shown high user receptivity and consistency when using various interaction techniques for selecting map features such as swiping streets, multi-clicking icons and circling map regions, with basic feedback provided on the mobile device [11] (see figure 1). The hybrid technique presented here extends the previous work by increasing selection precision and permitting lens-like interaction with the underlying resource.

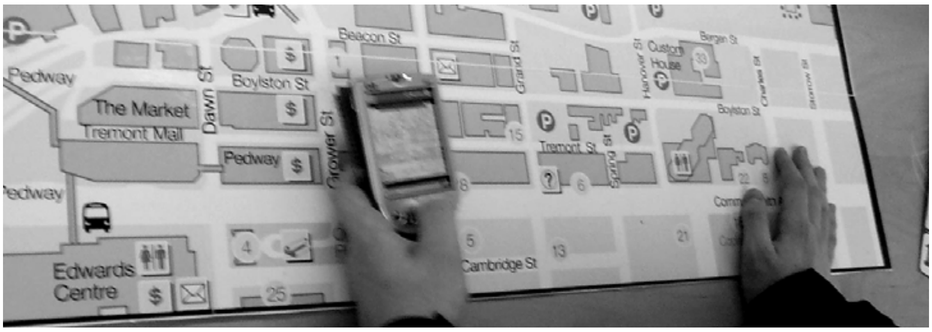

Fig. 1. RFID-only prototype permitting direct interaction with the paper map by swiping, click-selecting, and circling areas. The prototype was used to express queries such as “what restaurants are along this street?” by first constraining a query by selecting a paper menu item, and then identifying the region of interest.

1.1 Background and Motivation

The use of glyphs or ‘2D barcodes’ is a common approach for linking the digital and physical worlds. Visual tags are low cost, require only a standard camera, and can be used to determine the rotation and distance of the handheld from the tag. Unlike RFID, however, visual tags require line-of-sight and are normally clearly visible to the user. Near-field RFID systems that can detect distance and rotation have been demonstrated [8], however commercial RFID readers report read events only. For augmented reality, glyphs permit compelling integrations of the physical and digital. HMDs can superimpose interactive visuals on and relative to the glyph’s position [5]. When using mobile interaction devices, however, the utility of glyphs is limited: to avoid cluttering the visual field, glyphs are used sparsely, usually as physical URLs. The opportunity for interacting directly with the related artefact is limited in this case.

Digital ink technologies permit local positioning using a printed pattern barely visible to the naked eye. Extreme near-field or high-resolution capture and/or careful use of inks is required in order to achieve this, however [4]. Anchored relative positioning technologies and tracking technologies can give precise positioning of a node in 3D space, which can be used to permit interaction with stationary artefacts. However, such systems are usually constrained to a specific area, or to particular environmental conditions, or require physical installation that precludes truly everyday, mobile computing. Photosensing RFID tags can be used with handheld projectors to dynamically generate augmented displays [9], however this technique does not in itself support interaction with a surface. Combining projection and optical tracking permit interesting interactive map interfaces [7,12,13], but again this is not a mobile technology.

In previous work, Reilly et al. have explored the use of RFID tags embedded within paper maps as position indicators using off-the-shelf technology [10,11]. Tags have been placed in a regular grid, or beneath large landmarks and other points of interest. This approach provides direct interaction with map features and icons, without imposing constraints on cartographic design. The size of a passive RFID tag is a strong determinant of its read range; for map interaction, tags >=20mm in diameter have provided a suitable range, and needed to be carefully spaced to avoid read collisions. Therefore, the granularity afforded by an RFID-only solution is limited. This limitation is prohibitive even for prototype evaluations [11].

For an interaction modality to be truly mobile and pervasive, it should support direct interaction with physical artefacts without the need for intrusive stationary hardware configurations. Our work considers how to combine computer vision and RFID to support interaction, and overcome some of the limitations of either technology when considered for this purpose in isolation. The technique requires only that a map be outfitted with a coarse grid of RFID tags, and that the mobile device be equipped with an RFID reader, a camera and image recognition software.

2 Integrating RFID and Vision

The hybrid RFID/vision implementation relies on regularity of map icons and features. Image recognition algorithms are written to recognize a set of icons and/or features present on the map. An RFID grid is used to provide anchor points to constrain the search space. Ideally, the last tag read by the mobile device indicates which ‘grid square’ the device’s camera is centered on. The vision algorithm processes the camera image, focusing on identifying those map icons and/or features known to be perceptible in the region and useful for providing a visual fix. When the mobile device’s position, orientation and distance relative to the map is determined, this information is then used to update the display on the device. By focusing on repeated icons or features during image recognition, the code required to describe specific icons is reduced for any given map, and for related maps if the same icon set is used, as we often see with public maps in large areas.

The image recognition process involves 3 steps: (1) edge detection and GET extraction, (2) icon detection, and (3) icon recognition by its shape. Edge information in the raw image is used first instead of colour information, as the colour of icons will vary under different lighting conditions. Edges with distinct perceptual edge features called Generic Edge Tokens (GETs) are extracted from the raw images by the edge tracker [3]. GETs are perceptually significant image primitives which represent classes of qualitatively equivalent structure elements of edges. A complete set of GETs includes both Generic Segments (GS) and curve partition points (CPP). Each GS is a perceptually distinguishable edge segment falling into one of 8 categories (Figure 2). Each CPP is junction of GS’s.

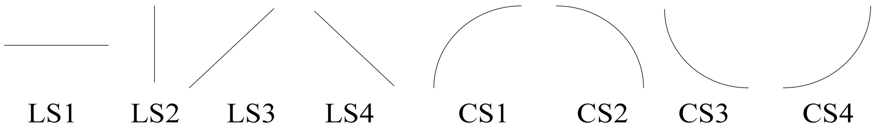

Fig. 2. The categories of Generic Segments (GS).

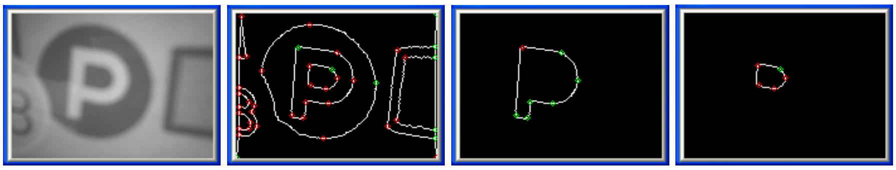

Fig. 3. Icon extraction example. From left to right: a) camera image; b) GETs extracted from the image; c) icon contour outline; d) icon contour inside.

The GETs carrying perceptual features are the main characteristics used in icon detection and recognition. Figure 3(b) shows the GETs extracted from Figure 3(a). In the next step, icon detection is achieved by perceptually grouping the appropriate GETs into icon contours (Figure 3c,d). Each icon consists of one or more homogeneous regions, therefore icon contours can be grouped by applying the perceptual region contour detection algorithm to the GET map [2]. Because of the descriptive nature of the GET representation, and the structural information inherent in the GET features, icon shapes/structures can be estimated easily based on their contour GET types. Shape descriptors are used to describe structure features of the extracted icons. An unknown icon can be matched to known icons according to human knowledge of the visual features of typical icon types in the map. Camera rotation can be estimated simultaneously during the recognition, for the features used in shape identification are invariant to rotation. The shape similarity between an unknown icon and a standard icon can be measured by the match confidence as follows:

$$confidence = \sqrt{\sum_i w_i (f_i' - f_i)^2}$$

where w_i is the weight of the i^{th} perceptual feature, f_i’ and f_i are the values of the i^{th} feature of the icon to be classified and the standard icon respectively.

RFID readers may or may not indicate when tag collisions are encountered; this can happen when the handheld device is situated between two tags, for example. In our prototype, tag collisions are not reported: this increases the area that must be searched to include the borders between tags in all directions. Furthermore, it is often not possible to rely solely on the presence or absence of map icons in a specific region to determine position and orientation. Some knowledge of the placement of icons relative to others in the region is normally necessary to gain a fix on location. This is true whenever a given icon appears more than once in a region, or when an icon’s symmetry (consider a triangle for example) makes determining its orientation difficult in isolation. However, by using RFID to constrain the size of the region to match, recognizing just one additional icon will almost always yield a fix on position and orientation. Therefore, for each RFID region a table is created, showing the relative distance(s) and rotation(s) of each pair of icons in the region. When an RFID region is entered, the algorithm retrieves the corresponding table. After recognition, if an icon that can be used in isolation to get a fix is recognized, the process completes. Otherwise the recognized icons are paired, and the table is searched to determine whether a fix can be achieved using each icon pair. The process continues until a pair yields a fix on position (and rotation), or fails if no pair can do so. If a map region exhibits high regularity in icon placement (to represent regularly spaced landmarks like oil wells, for example), icon triplets may need to be considered to compare the geometric relations in their relative placement. No such regularity exists in the map used in our prototype.

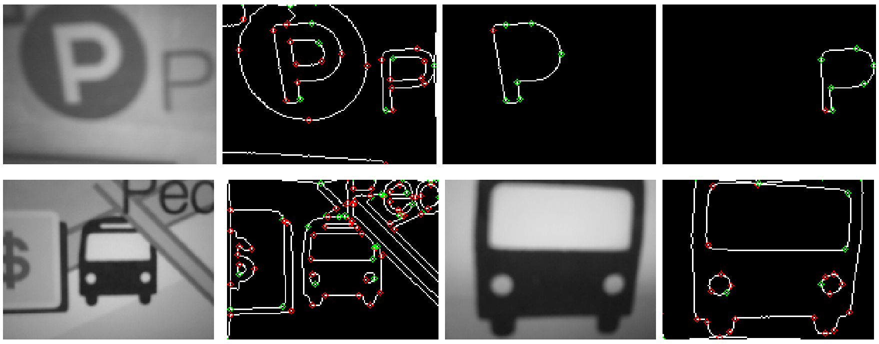

Fig. 4. Failed recognition examples. 1st row: similar icon shapes (resolved in second pass by comparing greyscale value with reference image); 2nd row: incomplete icon contour, first example due to occlusion, second example due to the camera being too close to the map surface.

3 Testing and Evaluation

Image recognition support was developed for a set of icons used in a commonly used map of Halifax, Canada (Figure 1). The icon detection and recognition steps were tested with three sets of camera images. The first set contains images taken by the handheld's camera at different rotations relative to the map. The second set contains images taken from various distances from the map. The third set contains a variety of images taken under two different lighting conditions (low light and high reflection). The testing shows that our methods are robust under different lighting conditions, camera focus, camera rotation, and distance from the map (see table 1). As illustrated in figure 4, false recognition may occur in certain situations: if several types of icons share similar shapes in the map (which is resolved in our prototype using basic greyscale value analysis), or if the icon contour cannot be detected completely from the image, due either to icon occlusion by other map features, or to the camera being too close to the map.

The recognition accuracy is estimated for efficiency evaluation by precision and recall. For all images in the testing set, let N_{all} be total number of icons needed to be recognized in the images; let N_{recog} be the total number of icons recognised in this image; let $N_{correct}$ be the number of icons correctly extracted and recognized in the test; recall can be defined as: Recall = $N_{correct}$ / N_{all}, and precision can be defined as: Precision = $N_{correct}$ / N_{recog}. Experiment results are listed in Table 1.

Table 1. Recognition accuracy test results.

Image Set	$N_{correct}$	N_{recog}	N_{all}	*recall*	*precision*
Base image	19	20	20	95%	95%
1 (rotation)	9	9	10	90%	100%
2 (distance)	8	9	10	80%	89.9%
3 (lighting)	38	39	42	90.5%	97.2%

4 Future Work

4.1 Mobile Device Prototype

Our prototype implementation employs a Windows Mobile 5.0 smartphone as the interaction device. Because the image recognition libraries used were written for the Windows platform, we perform image recognition and analysis on a separate computer. To achieve fast communication between the handheld and the PC we use a USB 2.0 connector and the Remote API (RAPI). Video is captured by the smartphone camera and sent directly to the PC for real-time analysis. We also physically attach an RFID reader to the back of the smartphone, however the reader is connected directly to the PC via a USB 2.0 connection. On the PC the algorithm as described in the previous section is carried out, using video from the camera and RFID reads from the reader. Once a visual fix is achieved, whenever the RFID reader reads a different tag, or if the algorithm cannot determine the location, the PC sends information to the smartphone. The smartphone interface, currently under development, will use this information to update its display appropriately. For the mobile lens prototype, the display will simply present the image being captured by the camera. Fix data is then used to superimpose additional information on the camera image.

4.1 RFID-based Heuristics

Some heuristics can be applied that consider the pattern of RFID tags read to reduce the region to consider first for a visual fix. Such rules could be applied if and when the space of possibilities is too large for the vision algorithm to identify the correct map feature, or to reduce average processing time. A timestamp associated with the last RFID tag read may be used as an indicator of likelihood that the handheld is centered on a grid square vs. being in a ‘border region’ between RFID tags or out of range. The set of reads within a set amount of time may also help determine a likelihood of position. For example, a recent sweep across a series of adjacent tags suggests that the device is over a tag or has continued along the same trajectory. A set of reads that lie predominantly between two adjacent tags indicates that the device is likely somewhere between these tags. We will conduct testing to determine the reliability and utility of such heuristics.

The RFID grid also provides a reference for movement. If image recognition lags behind movement of the mobile interaction device, the RFID tag reads can be used as approximate motion indicators, and the interface can display a smooth transition between RFID grid squares, maintaining the last distance and rotation of the mobile device as determined by image recognition. When the next position is fixed by the recognition algorithm, the rotation and distance can be smoothly adjusted to reflect this.

5 Conclusion

We have presented a hybrid RFID and vision-based approach to interaction with static maps and other similar visual resources. Preliminary testing indicates that this approach can offer richer position information compared to standard RFID alone. By giving rotation, distance, and precise position, the technique facilitates the use of a mobile device for lens-like interaction with the underlying map. Importantly, the technique requires only a grid of passive tags on the physical artefact, some regularity in its visual presentation, and a mobile device equipped with a camera and RFID reader, permitting truly mobile, ubiquitous interaction with such resources.

6 References

1 Bier, E.A., et al., *Toolglass and Magic Lenses: The See Through Interface.* in *SIGGRAPH 1993*. 1993.

2 Chen, H. and Gao, Q., *Image region and shape detection by perceptual contour grouping.* in *IEEE International Conference on Mechatronics and Automation 2005*. 2005. Niagara Falls, Canada.

3 Gao, Q. and Wong, A. *Curve Detection based on perceptual organization.* Pattern Recognition, 1993. **26**(1): p. 1039 – 1046.

4 Hecht, D. L. *Printed Embedded Data Graphical User Interfaces.* IEEE Computer, 2001. **34**(3): p. 47 – 55.

5 Looser, J. et al., *Through the looking glass: the use of lenses as an interface tool for Augmented Reality interfaces.* in *APCHI 2004*. 2004.

6 Fitzmaurice, G. W. *Situated information spaces and spatially aware palmtop computers.* CACM, 1993. **36**(7): p. 38 – 49.

7 McGee, D. R., and Cohen, P. R., *Creating Tangible Interfaces by Augmenting Physical Objects with Multimodal Language.* in *IUI 2001*. 2001.

8 Paradiso, J. A., et al., *Sensor systems for interactive surfaces.* IBM Sys. J. **39**(3): p. 892 – 914.

9 Raskar, R., et al., *RFIG Lamps: Interacting with a Self-Describing World via Photosensing Wireless Tags and Projectors.* in *SIGGRAPH 2004.* 2004.

10 Reilly, D., et al., *Marked-up Maps: Combining Paper Maps and Electronic Information Resources.* Journal of Personal and Ubiquitous Computing, 2006. **10** (to appear).

11 Reilly, D., et al., *Just Point and Click? Using Handhelds to Interact with Paper Maps.* in *Mobile HCI 2005*. 2005. Saltzburg, Austria.

12 Reitmayr, G., et al., *Localisation and Interaction for Augmented Maps.* in *IEEE ISMAR 2005.* 2005. Vienna, Austria.

13 Ullmer, B., and Ishii, H. *The metaDESK: models and prototypes for tangible user interfaces.* in *UIST 1997*. 1997. Banff, Canada.