(Project Page: http://www.merl.com/people/raskar/Sig04/)
Appeared in SIGGRAPH 2004, Los Angeles

RFIG Lamps: Interacting with a Self-Describing World via Photosensing Wireless Tags and Projectors

Ramesh Raskar* Paul Beardsley Jeroen van Baar Yao Wang Paul Dietz Johnny Lee
Darren Leigh Thomas Willwacher

Mitsubishi Electric Research Labs, Cambridge MA†

Abstract

This paper describes how to instrument the physical world so that objects become self-describing, communicating their identity, geometry, and other information such as history or user annotation. The enabling technology is a wireless tag which acts as a radio frequency identity and geometry (RFIG) transponder. We show how addition of a photo-sensor to a wireless tag significantly extends its functionality to allow *geometric* operations - such as finding the 3D position of a tag, or detecting change in the shape of a tagged object. Tag data is presented to the user by direct projection using a handheld locale-aware mobile projector. We introduce a novel technique that we call interactive projection to allow a user to interact with projected information e.g. to navigate or update the projected information.

The ideas are demonstrated using objects with active radio frequency (RF) tags. But the work was motivated by the advent of unpowered passive-RFID, a technology that promises to have significant impact in real-world applications. We discuss how our current prototypes could evolve to passive-RFID in the future.

CR Categories: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities; H.5.2 [Information Interfaces and Presentation]: User Interfaces—Interaction styles; I.4.1 [Image Processing and Computer Vision]: Digitization and Image Capture—Imaging geometry;

Keywords: human-machine communication, augmented reality, projector, radio frequency identification, stucture from motion, image stabilization

1 Introduction

As computational devices shrink and their power requirements and cost decline, it becomes feasible to envisage an everyday world whose surfaces are embedded with computational power and locally-held data. This paper describes the tagging of physical objects with radio frequency (RF) tags. Our experimental work is based on active, battery-powered radio frequency tags. However, our goal has been to develop methods that can be used with passive, unpowered radio frequency identification (RFID) tags. It is the cost of custom-design that prevented us from actually doing experiments with passive-RFID. Passive tags are just beginning to enjoy widespread deployment in real-world applications and the ability to implement our techniques on passive tags is fundamental to the usefulness of the ideas.

We augment each tag with a photo-sensor to significantly extend the current functionality and support radio frequency identity and geometry (RFIG) discovery. The ability to address and wirelessly access distributed photosensors creates a unique opportunity. We recover geometric information, such as 3D location of tags or shape history of tagged objects, and exploit the associated geometric operations to bring the RF tags into the realm of computer vision and computer graphics.

We present a complete solution for retrieving, viewing, navigating, and updating tag data. We build on our work with iLAMPS in [Raskar et al. 2003] using handheld locale-aware mobile projectors (LAMPs) to do direct projection onto tagged objects. Further our work goes beyond passive projection to show a truly interactive system in which projected information can be navigated and updated.

Figure 1: Warehouse scenario. A user directs a handheld projector at tagged inventory, with communication mediated by two channels – RF and photo-sensing on the tags. The user sees a projection of the retrieved tag information collocated with the physical objects, and performs a desktop-like interaction with the projection. A second user performs similar operations, without conflict in the interaction because the projector beams do not overlap.

1.1 Contributions

At the conceptual level, we show how to extend current tag-readers, which operate in broadcast mode with no concept of a 3D coordinate frame, to **allow selection of individual tags,** and create **a 3D coordinate frame** for the tags. This new functionality is achieved

*email:[raskar,pab,jeroen,dietz,leigh]@merl.com
†web:http://www.merl.com/projects/Projector/

by combining the tag-reader with a projector, and by augmenting each tag with a photo-sensor. In fact, we show that the system of projector and photo-sensing tag offers a new medium for many of the results from the area of structure-from-motion in computer vision, with projector and tags replacing camera and image interest points.

We also introduce **interactive projection,** allowing a user with a handheld projector to do mouse-style interaction with projected information. This is achieved by treating a projection as conceptually having two parts – a stabilized component that is static in the display surface, and a cursor that follows any user pointing motion of the projector – effectively allowing the user to track a cursor across a projection. Accompanying mouse buttons are used to do selection. Interactive projection is not specific to a tagged environment, but the technologies meld well.

1.2 A Self-Describing World

Our work in this paper involves experiments on the individual functions of active, photo-sensing RFID tags. However we motivate the work by suggesting a real-world application that could utilize the ideas. The first large-scale use of passive-RFID tags is expected to be for inventory control as part of logistics (a US$900 billion industry), so we turn to a scenario in a warehouse - locating objects with a required property and annotating them. For example, a warehouse employee identifies food products that are close to expiry date and annotates an instruction to trash them. If these products were items on a computer desktop, this could be done with a few clicks. Our goal is to craft a scheme in the physical world that maintains the simplicity of the computer environment. In the computer, the items are files, and the interface is via keyboard, mouse, and display. In the physical world, the items are tagged objects, and the interface uses a handheld projector and user interaction directly through the projected information. Figure 1 illustrates a handheld system being used in the warehouse scenario - selecting tags, collecting data, projecting imagery onto the objects to present the information to the user, and interacting with the projection. The term 'handheld' is used in this paper to cover both continuous handheld motion of the projector, and motion interspersed with static but opportunistic placement, as might occur during say augmented reality task-guidance.

We describe a number of innovations at the technical level. We show how to address the two hard requirements of augmented reality – object recognition, and recovery of 3D pose – without employing visually obtrusive markers. We extend passive display of augmented reality with an easy way to navigate through and update augmentation information. We show geometry-based functionality like finding the geometric configuration or geometric deformation of a set of tags. We show image-based functionality such as capturing imagery of the surface texture of a tagged object, with correction of perspective effects - this image data provides a suitable basis for creating the desired augmentation that is to be projected onto the surface. In the specific area of projection, we show a way to deal with the problem of projecting augmentation onto dark or cluttered surfaces, by capturing an image of the surface, then projecting the distortion-free image plus its augmentation onto a clean display surface.

The primary reason for using a projector in our work is that it underpins the technique for selecting tags and for setting up a 3D coordinate frame. This technique involves light communication that is not for human viewing. But it is convenient to make dual-use of the projector - using it also to project augmented reality data onto tagged surfaces, for human viewing. Projection has some advantages over other augmented reality approaches - augmented reality on a handheld display requires a context-switch as the user visually matches what is on the display with the physical object. And eye- or head-mounted displays need a real-time, accurate computation of the relationship between the user's coordinate frame and the world.

The limitations of our system are that a projector must have line-of-sight when interacting with a photo-sensing tag. The viewing quality of projected augmentation is affected by ambient lighting, low albedo surfaces, and shadows. Range for tag communication with our current setup is 2-3m. Within these constraints, the work still suggests many possibilities.

For applications, we focus on augmented reality, not just passive display, but demonstrating user interaction with the augmentation. Augmented reality faces challenges in doing reliable object recognition and computation of object position. Current solutions that use visual markers or user trackers are unwieldy. In contrast, a tag-based system can provide object identification and (as we show) object position in an unobtrusive way. Tags also allow direct storage of augmentation data with tagged objects and this sometimes has advantages over keeping an object's data in a remote database - say for mislaid objects, or for transient annotations.

1.3 Relationship of our Work to Passive RFID

This section describes why we assign importance to being able to support passive RFID tags, and what would be involved in evolving from the current work to a passive RFID-based system.

Powered radio-frequency tags currently use a battery that is about the size of a watch-battery, have a lifetime of a few years, and have a cost of a few dollars. In contrast, passive RFID tags are unpowered, can be as small as a grain of rice, and cost tens of cents [Want 2003]. Prices of both are dropping but the price differential will remain. The size and cost properties are such that RFID is showing signs of being adopted as a mass-deployment technology. Current commercial applications including embedding of RFID tags in packaging for inventory control, non-contact access control, and ear tags for livestock. Despite the interest in RFID, the available functionality is very limited – an RF-reader broadcasts a request, and in-range tags (collect energy from the RF pulse and) reply. Our work is motivated by the observation that RFID is showing the potential to be a ground-breaking, pervasive technology, yet current functionality is limited.

The key issue in evolving our active tag system to passive tags would be power. In the work in this paper, we only allowed ourselves computation and sensing consistent with the size and power levels we felt were achievable on a passive RFID system. For example, (a) tags are not photo-sensing or computing until woken up by the RF reader and (b) we do not have a light emitting diode (LED) on the tag as a visual beacon to a human or camera-based system because it would be power-hungry. Also note that our tags incorporate a photo-sensor, so a passive version could draw power not just from the RF channel, but also from the incident light. Of course, there would be significant engineering challenges in moving from active to passive RFID. Still, this has been the motivating vision for the work, and this has determined our design choices for the current prototype.

1.4 Related Work

Smart Objects: Multiple groups have looked at adding intelligence to objects and, in some cases, building human interactions around them. Smart-Its provides interconnected smart devices that can be attached to everyday items [Holmquist et al. 2001]. Visual feedback is not considered. SenseTable tracks and augments the area around sensing tablets on a tabletop using a projector [Patten et al. 2001]. Intelligent furniture has also been explored [Omojola et al. 2000]. Some systems use active RF tags that respond to laser pointers. The FindIT flashlight uses a one-way interaction and an indicator light on the tag to signal that the desired object has been found

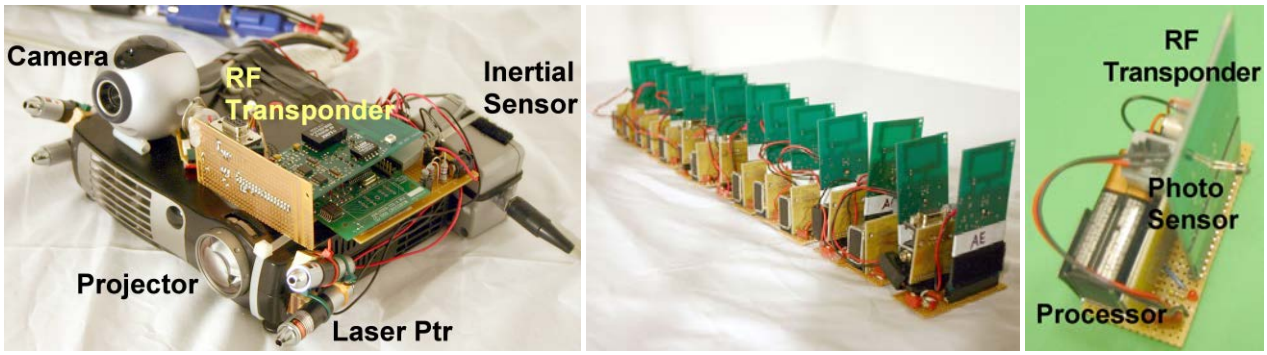

Figure 2: (a) Handheld projector system, (b) Fourteen of the tags and (c) Details of a tag.

[Ma and Paradiso 2002]. Other systems use a two-way interaction, where the tag responds back to the PDA using a power-hungry protocol like 802.11 or X10 [Patel and Abowd 2003] and [Ringwald 2002]. CoolTown [The CoolTown Project 2001] uses beacons that actively transmit devices references but without the ability to point and without visual feedback.

Interaction and Augmentation: Interaction with laser pointers for large display screens is common. A collection of sensing and interaction techniques for mobile devices is described in [Hinckley et al. 2000]. Augmentation of physical world objects has been primarily achieved via eye-worn or head-mounted displays [Azuma et al. 2001] or handheld screens. Screen-based augmentation using PDA, camera, and RFID tags is described in [Want et al. 1999; Rekimoto et al. 2001]. Projector-based augmentation has been explored by [Raskar et al. 1998; Underkoffler et al. 1999; Pinhanez 2001; Raskar et al. 2001; Bimber et al. 2001; Verlinden et al. 2003]. Image warping for a handheld projector to support shape and object adaptive projection is described in [Raskar et al. 2003]. The Cricket project [Teller et al. 2003] recovers pose of a handheld device using installed RF- and ultrasound- beacons, and does projected augmentation. A handheld projector as virtual flashlight, creating real shadows of virtual objects on real surfaces and supported by camera and optical tracking of fiducials on the wall, is described in [Foxlin and Naimark 2002]. Adding distinctive visual markers for tracking is not always practical, and unlike previous real-time AR applications, we do not add obtrusive environmental infrastructure to obtain a 3D context.

Other: Location sensing systems such as the Olivetti Active Badge [Want et al. 1992] and Xerox PARCtab [Want et al. 1995] recover location, but have typically been used for passive tracking, not an interactive system. Our work is also influenced by efforts in wireless sensor networks, pervasive computing and amorphous computing [Abelson et al. 2000], systems that use autonomous units to propagate information. Our passive tags lack the power to communicate with neighbors, but we use the handheld projector to provide an information conduit between tags.

A distinguishing factor between our work and related systems is in the parallel nature of the interaction. Laser-pointer systems require a user to identify a target object and direct the pointer at it to initiate interaction. But accurate pointing is difficult when the tags become visually imperceptible. And multiple tags can only be dealt with serially. In contrast, we use a casually-directed projector to interact with all tagged objects within the projector beam, and can select all objects or a subset for further interaction. Further, our tags are rich data sources, with the interactive projection technique providing a flexible user interface to the data.

1.5 Paper Organization

Section 2 describes the hardware. Section 3 describes tag selection and creation of a 3D coordinate frame. Section 4 describes interactive projection. Section 5 puts the pieces together in several examples of use. Section 6 describes results. Important issues that are not discussed because they are outside the scope of the paper are hardware design (miniaturization, cost, industry standards), RF communication (power, range, interference) and social aspects (privacy, security).

2 Physical Components

2.1 Handheld Projector System

The components of the prototype handheld projector system in Figure 2a are:

- a Plus V-1080 projector with 1024x768 resolution, 60Hz frame-rate, dimensions $7\times5\times1.5$ inches and weight 1kg,
- a rigidly attached 640x480 pixel webcam,
- a rigidly attached AC&C VS123 6DOF inertial sensor (accelerometer and gyroscope),
- four rigidly attached laser pens,
- a Parallax board for RF communication at 433.92 MHz with data transfer rate of 2400 baud,
- two click buttons for simple IO,
- a computer.

The camera is used to compute pose (position and orientation) of the device, as well as supporting user interactions that capture surface texture. The inertial sensor supports the computation of pose by providing high-frequency estimates (160Hz) of change in pose. The four laser pens, which could be invisible IR, are used to project distinctive points on the projection surface to support computation of pose, rather than constraining the form of the projector's own image data for this purpose. Euclidean calibration between all elements is done offline, and in the remaining discussion we assume projector and camera intrinsics are known and have been factored out. Much of the calibration uses established camera calibration techniques (modified for a projector) [Hartley and Zisserman 2000].

2.2 Tags

The prototype active tag, Figure 2b, has a microprocessor, memory, Parallax board for RF communication, and a Panasonic PN121PS-ND photo-sensor for sensing projector illumination.

2.3 Projector-Tag Communication

The handheld device has two ways to communicate with a tag. RF communication is omni-directional with a typical range in meters. Projected light provides directional communication via the photo-sensor. Figure 3 is a schematic of the devices and interaction.

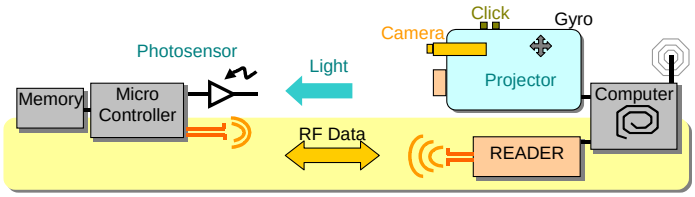

Figure 3: Communication between the photosensing tag (left) and handheld projector system (right).

Radio-frequency and photo-sensing are complementary approaches for communication. Traditional tags do not use visible or infrared light communication because occlusion and ambient illumination make light unsuitable for data transfer. However, light is suitable for directional methods such as directional selection of tags, and for recovering geometry by employing light rays.

3 Processing Tags

We describe novel functionality that becomes available when a tag-reader is combined with a projector, and a tag is augmented with a photo-sensor. Section 3.1 describes how this allows a user to select individual tags for interaction instead of working in broadcast mode, and Section 3.2 describes how to recover a 3D coordinate frame for the tag system.

3.1 Selecting Tags

Conventional tag communication works by broadcast from an RF-reader, with response from all in-range tags. Limiting the communication to a required tag is traditionally achieved using a short-range tag-reader and close physical placement with the tag. In contrast, we can select tags for interaction at long-range using projected light, while ignoring unwanted in-range tags. The handheld device in Section 2 first transmits an RF broadcast. Each in-range tag is awoken by the signal, and its photo-sensor takes a reading of ambient light, to be used as a zero for subsequent illumination measurements. The projector illumination is turned on. Each tag that detects an increase in illumination sends a response to indicate that it is in the beam of the projector, and is ready for interaction. The user can control the selection by varying the projector beam size, or by shaped projection. A more powerful mechanism for tag selection is given in Section 4.2 using interactive projection.

Figure 4: (a) A tagged surface, (b) One frame during Gray-code projection, (c) Correctly-placed projected augmentation.

3.2 Setting up 3D Coordinate Frame for Tags

Existing tag systems have no functionality for creating a 3D coordinate frame. There are directional RF devices that could be used to provide direction between RF-reader and tag, but they are not appropriate for handheld use, requiring meter-wide antennas, or multiple RF readers distributed in space. While computer vision could potentially be used, it would require visually-distinctive tags, and our goal is to move in the direction of small, indiscernible tags. Instead we propose the following approach.

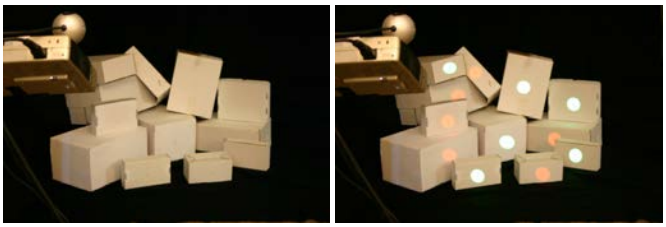

Figure 5: The projector beam reveals the presence of tagged objects

The handheld device is first aimed casually in the direction of a tagged surface like that in Figure 4a. The handheld device sends an RF signal to synchronize the tags, followed by illumination with a sequence of binary patterns – see one example frame in Figure 4b - such that each projector pixel emits a unique temporal Gray-code. The tag records the Gray-code that is incident on its photo-sensor, and then makes an RF transmission of its identity plus the recorded Gray-code back to the RF-reader.

The handheld device's camera also observes the projected light and determines where each Gray-code appears in the camera's view of the surface. Thus, via the Gray-code, we know the following information for the tag – what projector pixel **p** is illuminating it, and what camera pixel **c** is viewing it. Since this is a calibrated projector-camera system, this reduces to a triangulation problem in which we intersect the 3D rays associated with **p** and **c** to find the 3D position of the tag. It is then straightforward to create correctly positioned augmentation on the tagged surface as shown in Figure 4c and Figure 5.

4 Handheld Projector for Interaction and Display

Our other main contribution is to provide a novel user interface that allows a user to interact with a projection. Section 4.1 describes how we attain a stable projection from a handheld projector under hand-jitter, and Section 4.2 describes our method of interactive projection. Although the focus of this paper is on projection in tagged environments, note that interactive projection can work with any display surface, and does not require a tagged surface.

4.1 Stabilized Projection

In creating a handheld projector, an immediate problem that arises is that hand-jitter results in jitter in the projection. The core requirement to deal with this problem is to compute the pose of the projector relative to the display surface. There are multiple ways to approach this. Section 3.2 already introduced a way to find the 3D coordinates of tags on a tagged surface (and hence to find projector pose relative to the surface). However, the computation of projector pose via projected Gray-codes is too slow for correction of hand-jitter.

We have used two approaches to stabilize a projection on a planar display surface:

• The first approach, which we call absolute stabilization, works by recovering full projector pose - location and orientation - relative to the display surface, by utilizing the camera's view of four or more points of known coordinates in general position on the surface (or the equivalent, such as four known lines). Hence, we keep the projection static on the display surface, factoring out motion of the projector completely,

• The second approach, which we call quasi-stabilization, works by recovering projector pose relative to the display surface up to an unknown translation in the plane. It utilizes the camera's view of four projected points on the display surface e.g. the four vertices of the projector image (or the equivalent, such as the four boundary lines of the projector image). This gives projector pose to an unknown translation in the plane, and an unknown rotation around the plane normal. We fix the rotation around the plane normal by using a tilt-sensor to determine the world vertical, and hence to define a unique upright direction for any non-horizontal display surface. This allows us to preserve the projection size and orientation, but the projection translates in the display plane if there is a dominant projector motion parallel to the plane.

Laser Pens – The descriptions below involve the handheld device's camera detecting the four vertices of the current projector image. In fact, we detect the projected laser points from the four laser pens. This improves reliability of detection, because laser points are bright, and avoids a constraint on the form of the projector image. This involves an extra layer of processing because the laser rays are not concurrent with the projector pinhole [Beardsley et al. 2004b], but conceptually the situation is unaltered, so we omit this detail from the algorithms below.

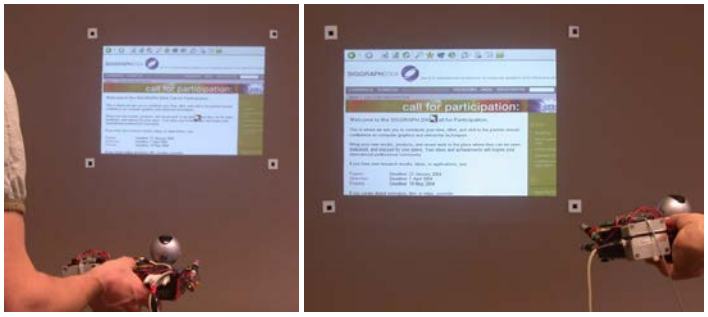

Figure 6: Absolute stabilization projects to a fixed location, regardless of projector position.

Absolute-Stabilized Projection Absolute stabilization, Figure 6, is possible when the camera is viewing four or more fixed surface points in general position, or the equivalent, on a planar display surface. Our goal is to find a homography H_{PS} between the projector image plane and a fixed coordinate frame on the surface. Homography H_{PS} specifies the mapping between each pixel on the projector image plane and the fixed surface coordinate frame. Hence we can use its inverse H_{SP} to transform from a desired (fixed) projection on the surface to the projector image plane, thereby determining the required projector image to give the fixed projection. The steps to find H_{PS} are as follows.

- Take an image with the handheld device's camera. Detect camera image points $\mathbf{x}_i$, for four surface features of known surface coordinates $\mathbf{X}_i$. Compute the homography H_{CS} between the camera image plane and the surface coordinate frame, using the correspondences $(\mathbf{x}_i, \mathbf{X}_i)$.
- In the same camera image, identify the four camera image points $\mathbf{y}_i$, corresponding to the vertices of the projector image. Knowing the projector pixel coordinates for the four vertices $\mathbf{v}_i$, compute the homography H_{CP} induced by the surface between the camera image plane and the projector image plane using the correspondences $(\mathbf{y}_i, \mathbf{v}_i)$.
- Compute the required homography between the projector image plane and the coordinate frame on the display surface $H_{PS} = H_{CS} H_{CP}^{-1}$.

In conjunction with this, we use the inertial sensor on the handheld device to provide a fast update of change in pose of the handheld device. To preserve the projective nature of the algorithm above, the pose change is used to predict a new estimate of the camera image points $\mathbf{x}_i$, and the algorithm then proceeds exactly as above.

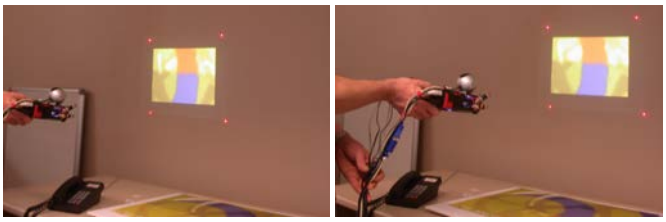

Figure 7: Quasi-stabilized projection translates with the user but maintains an upright rectangle with correct aspect ratio.

Quasi-Stabilized Projection Quasi-stabilization, Figure 7, preserves the form of the projection up to an unknown translation in the plane i.e. it preserves projection shape, size and orientation. But the projection translates in the display plane in accordance with projector motion parallel to the plane. The steps to achieve this are:

- Take an image with the handheld device's camera. Detect the camera image points $\mathbf{x}_i$, for the four vertices of the projector image boundary.
- Having a calibrated projector-camera system, knowing $\mathbf{x}_i$, and knowing the projector pixel coordinates for the four vertices $\mathbf{v}_i$, reconstruct the 3D plane of the display surface in the coordinate frame of the hadheld device. This provides projector pose relative to the display surface up to an unknown translation in the plane, and unknown rotation around the plane normal.
- Use the inertial sensor to get the world vertical, to fix the unknown rotation around the plane normal (for non-horizontal planes).
- Given projector pose up to an unknown translation in the plane, generate a projection of the desired shape, size and orientation on the plane.

The inertial sensor is incorporated into the scheme in a similar way to absolute-stabilization.

4.2 Interactive Projection

The term 'interactive projection' is used to refer to mouse-style interaction with projected data. Functionality to do pointing on a projection could be achieved with a touchpad on the handheld device, or with a separate laser pointer, but these are inconsistent with a compact device and one-handed interaction. We describe how to use simple one-handed direction of the projector to track a cursor across a stabilized projection. The handheld device has two click buttons for selection when the cursor is at a desired point. Interactive projection employs absolute stabilization described in Section 4.1, and therefore requires the display surface to have some texture.

To introduce the mechanism, first note that conceptually we have two types of projection. A conventional mode of projection might

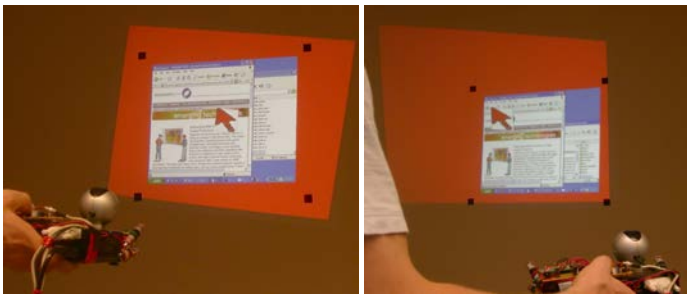

Figure 8: During interactive projection, a moving cursor can be tracked across a projection that is static on the display surface, to do mouse-like interaction with projected data. In these images, the background pixels are intentionally rendered in red to show the extent of projection. The pointer remains at the center of the projector field while displayed desktop remains stable within the surface features.

have a static image on the projector image plane and, as the projector is moved, the projection on the display surface moves in direct correspondence. In contrast, absolute stabilization, introduced in the previous section, has a dynamically-updated image on the projector image plane that results in a static projection on the display surface during projector motion.

Now note that both types of projection can be done simultaneously. Say we apply absolute stabilization to project some content such as a windows desktop environment. The user perceives a stabilized static windows desktop on the display surface. But in addition we always set a graphic for a cursor at the center of the projector image plane. Now as the projector moves, the user sees the stabilized projection of the windows desktop with the cursor tracking across it – see Figure 8. Furthermore the cursor motion follows the motion of the handheld projector, so it is a natural interaction for the user.

With the ability to point, projection becomes interactive. In fact all the standard mouse interactions from a WIMP interface are possible in this new setting. One difference is that in the projector domain, the cursor can be used to indicate either a selection in the projected content, or a selection of something in the physical world (with augmentation giving stabilized feedback about the selection e.g. a rectangular window on a surface). This opens up new types of interaction possibilities.

Interaction with projected content. Section 1 introduced a scenario for user interaction with warehouse inventory. This section describes how interactive projection supports the interaction. **(i) Option selection.** The user wishes to invoke a standard query like 'show inventory close to expiry date'. This can be done by having a menu or an action-button in the stabilized projection, and selecting with the cursor. **(ii) Tag selection.** Section 3.2 described how tags in the projector beam can be highlighted by projected augmentation to indicate their location to the user. To support tag selection, the augmentation could show a selection icon next to each tag - the cursor is tracked to the selection icon, with click to toggle its state. Alternatively the cursor can be used to select a rectangular region of interest by click-and-drag, with a stable view of the selected rectangle providing visual feedback about the selection, to select multiple tags simultaneously. **(iii) Controlling placement of augmentation.** The user can use the cursor to grab individual items in the projected augmentation and move them to a preferred layout.

Interaction with physical texture. The cursor can be used to indicate a point or a region selection for physical features in the world, to be captured by the camera and processed. We present a simple example - copy-paste of a selected part of a surface – in the next section. See further examples in [Beardsley et al. 2004a].

5 Functionality in a Tagged Environment

Section 3 and 4 proposed fundamental techniques that we build on in this Section. Here our contribution is to show practical functions for doing augmented reality and geometric processing in a tagged environment including (a) initializing tags by recovering and storing their 3D position on an object, (b) recovering images of a tagged object's surface texture to provide a basis for creating an augmentation overlay, (c) using copy-paste as a way to show a projected augmentation for surfaces that don't support direct projection.

5.1 Tag Geometry – Geometric Configuration

This section describes how to compute the 3D coordinates of a set of tags on an object. As motivation, assume a newly manufactured object with tags that have been embedded in approximately known positions (say the vertices of the object), but more accurate 3D coordinates are required for subsequent augmentation. Section 3 described how to compute the 3D position of an individual tag relative to the handheld device. But error in individual calculations makes this approach unsuitable to compute the positions of multiple tags in a common frame. We instead use an approach based on structure-from-motion [Hartley and Zisserman 2000].

- For a given placement of the handheld device relative to the object, compute the projector pixels $\mathbf{x}_i$, $i = 1..m$ that illuminate each of the m tags, $m >= 8$, using the method of Gray-code projection in Section 2.1.
- Repeat for N distinct placements of the handheld device. This reduces to the following computer vision problem – compute structure from Nviews of m matched features.
- Compute the essential matrix E between first two views. Initialize a Euclidean coordinate frame and the 3D positions of the tags.
- Process each subsequent view, using a Kalman Filter to update the 3D tag positions.
- Do bundle adjustment to refine the final result.
- Transform the coordinate frame to a world vertically aligned system, using the tilt of the inertial sensor.

Delaunay triangulation is used to create a 3D mesh of tag vertices and hence surface facets. All the computed geometry and connectivity is stored in the tags. A projector subsequently recovers its pose in the object coordinate frame as follows.

- Illuminate the object and recover the projector pixel $\mathbf{x}_i$ illuminating each tag by the method of projected Gray-codes.
- Retrieve the 3D coordinates $\mathbf{X}_i$ stored in each tag.
- This again reduces to a standard computer vision problem (requiring no modification for a projector, which is also a pinhole device) - compute camera pose given 2D-3D correspondences $(\mathbf{x}_i, \mathbf{X}_i)$. Given the projector internal parameters, use four correspondences to compute the projector pose [Zhang 1999].

5.2 Tag Geometry - Deformation

This section describes detection of non-rigid deformation in a set of tags as a way to detect change in the underlying object or surface. This can be useful to detect and highlight change, say detecting pressure damage during shipping, or monitoring deformations due to moisture or heat.

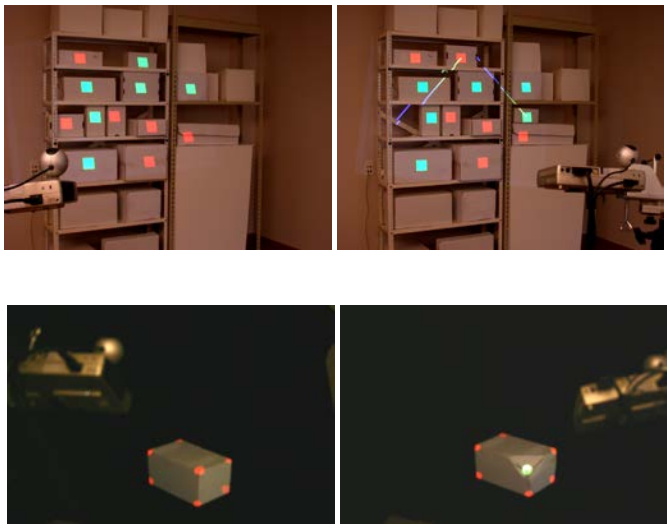

Figure 9: (a) Tagged boxes in original position, (b) Projection indicates change in box locations, (c) Box before deformation, (d) Projection indicates deformation of front vertex.

Each tag contains stored 3D coordinates $\mathbf{X}_i$ acquired pre-deformation, as in Section 5.1. The RANSAC-type procedure [Fischler and Bolles 1981] to detect deformation assumes that the majority of the tags are rigid, and a small number of tags have changed their relative position – we identify the tags that conform to rigidity, in order to obtain the tags that have undergone non-rigidity. The steps are:

- Find the tag positions $\mathbf{x}_i$ on the projector image plane using projected Gray-codes, and retrieve the 3D coordinates $\mathbf{X}_i$ by RF request, to obtain the correspondences $(\mathbf{x}_i, \mathbf{X}_i)$.
- Take a subsample of four correspondences from the full set $(\mathbf{x}_i, \mathbf{X}_i)$.
- Estimate projector pose Γ' from the subsample. Use the computed pose to project all the undistorted 3D tag coordinates $\mathbf{X}_i$ onto the projector image plane. Use the reprojection error between the actual tag positions $\mathbf{x}_i$ and the ideal projections to determine the inlying correspondences to this solution – those having reprojection error $|\mathbf{x}_i - \Gamma'\mathbf{X}_i|$ less than threshold d.
- Continue for n random subsamples, retaining the solution Γ with the largest number of inliers. Record the outliers to this solution as deformed tags.

5.3 Tag Imagery – Local Surface Texture

This section describes how to recover distortion-free image texture for the surface facets of an object (obtained in Section 5.1). As motivation, we might store image texture to provide input for an editor, for creating and positioning the augmentation graphics on a tagged surface. As a second motivation, assume the image texture is stored for subsequent use for change detection on a tagged surface like a control panel.

The steps are:

- Use the method in Section 5.1 to recover the projector pose and hence, by the known calibration, the camera pose relative to the object.
- Using the surface facet information retrieved from the tags, determine the projection of each surface facet on the camera image plane.

Figure 10: Projected augmentation on the box is attached to tags at the vertices, and so adapts automatically to the opening of the lid.

- Extract the image texture for each surface facet. Use a homography between the camera image coordinates of the surface facet, and coordinates for a fronto-parallel projection of the surface facet (obtainable from the known 3D coordinates) to rectify the texture to a fronto-parallel view.
- Store the distortion-free surface facet texture on the tags.

The motivating functions suggested above are now straightforward e.g. given the projector pose in the object coordinate frame by the method in Section 5.1, and the captured image texture with updates to incorporate desired augmentation graphics, it is straightforward to project the image texture back to the correct location on the object as shown in Figure 10.

5.4 Interactive Projection - Copy-Paste

This section describes copy-paste of image texture from a physical surface. We use the handheld device's attached camera to capture texture from a tagged surface, normalize to remove viewpoint dependence, and paste this texture onto a new surface without distortion. As a motivating example, assume we want to project augmentation onto a surface like that in Figure 11, which does not readily support projection. In fact, our experience suggests that even cluttered, dark surfaces can support simple bold projection. But fine detail projection requires another solution. The copy-paste operation is used to capture an image of the surface in Figure 11, simultaneously recording data from the tags embedded in the surface. A distortion-free view of the surface is then 'pasted' to a projection on a clean surface, showing both the image texture and the overlaid augmentation

There are multiple ways to select a physical region for copy, including the method in Section 4.2 for selecting a rectangular ROI, but we outline a simple method here. Use the handheld device to project a full-white projector image onto a planar surface. Due to perspective the rectangular projector image will project as a quadrilateral. Define the copy area as the largest vertically aligned inscribed rectangle inside the projected quadrilateral. The steps are:

- Find the four vertices of the projected quadrilateral in the camera view. Knowing the projector pixels for the vertices, and the camera pixels, and having a calibrated projector-camera pair, compute the quadrilateral in 3D.
- Find the largest inscribed rectangle, S, inside the projection quadrilateral, aligned with the world vertical (obtained from inertial sensor).
- Compute the projection of S to the camera image plane. Store the image texture for S, incorporating a warp to generate a fronto-parallel view of the captured texture.

For pasting at a new location:

- Find an inscribed rectangle on the display surface as above.
- Project the stored texture centered on this rectangle

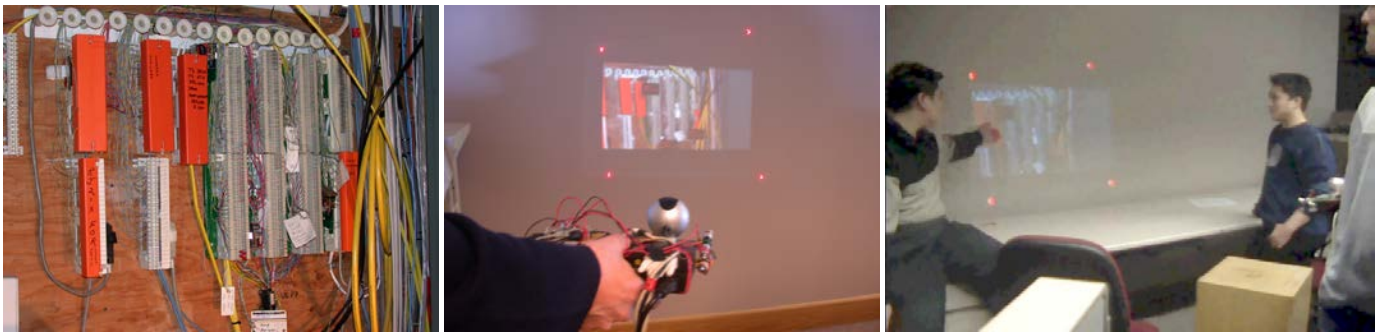

Figure 11: (Left) This surface, with embedded tags, does not readily support projected augmentation. To deal with this, during 'copy', an image is captured, and location and data on the tags is retrieved. (Middle and Right) During 'paste', perspective and photometric distortion is removed, and the image plus augmentation data is displayed on a clean projection surface.

Photometric Compensation Photometric compensation is used to produce a paste image that is as close as possible to the copied texture. There are two relevant factors – attenuation due to distance and orientation, and surface reflectance.

Projector illumination at a specific surface point is a function of projector-surface distance along the ray and incidence angle of the projector ray. The dimmest projector illumination falls on the surface point with the largest incidence angle. This therefore sets the brightness limit for the projection, and we apply a per-pixel attenuation map to the rest of the image to provide an even illumination. For a projector pose $\Gamma = [r_1\ r_2\ r_3\ t]$, with rotation $R = [r_1\ r_2\ r_3]$, the homography between the projector image plane and physical surface is $H_P = [r_1\ r_2\ t]$. A projector pixel x maps to 3D point $X = (X_x, X_y, 0)$ where $(X_x, X_y, 1) \cong H_P^{-1}x$. The angle between the projector ray and surface normal at X is

$$\theta = \cos^{-1}(V \bullet (0,0,1))$$

where, the projector ray is $V = (-R't - X)/|(-R't - X)|$

If the largest incidence angle is θ_m , then attenuation at every projector pixel x is $\cos(\theta_m)/V_z$, where V_z is the z components of the vector V given above. We achieve this attenuation with a simple pixel shader program.

For reflectance compensation, we avoid explicit albedo calculation [Nayar et al. 2003] by noting that the copied texture and the view of the illuminated surface after paste operation are both captured by the same camera. We take an image of the paste surface for two different illuminations from the projector. Given camera intensities I_1 and I_2, for a given camera pixel under the two illuminations L_1 and L_2, and a target paste intensity of I_T, the required linearly interpolated projector illumination is $L = L_1 + k(L_2 - L_1)$ where $k = (I_T - I_1)/(I_2 - I_1)$. For color images, we apply the correction independently in each channel.

6 Results

6.1 Qualitative and Quantitative Results

How well does our implementation work? Typical range between the handheld device and tags in our experiments was about 2m. The method for identifying which tags are within the projector beam, in Section 3.1, works reliably all the time. It would likely have problems in environments where the ambient light undergoes sudden change. The method for computing 3D tag position using Gray-code projection, in Section 3.2, takes about 2 seconds to do the projection. It works about 95% of the time, with the failures being due to noise in the Gray-code detection at the tag.

The method of stabilized projection in Section 4.1 results in a projection that is visibly more stable than raw projection under hand-jitter. However jitter is still present. We have experimented with interactive projection, in Section 4.2, by projecting a standard windows desktop environment onto a wall, invoking a browser, and doing operations like menu selection, clicking on links, and scrolling. The interactions feel somewhat clumsy due to the remaining projection-jitter - a major reason is that we are putting the computed cursor motion directly into a standard mouse-handler, rather than a custom-handler, and mouse-click is often interpreted as mouse-drag due to jitter. Despite this, the standard mouse-type interactions can be successfully carried out on the projection of a desktop environment with the current prototype.

The four functions in Section 5 can all be done reliably.

Projection stabilization in the presence of hand-jitter works by counteractive displacement of the image on the projector image plane, and hence we need a buffer zone around the usable projection area. The pixels outside the usable area are black. This decreases the effective projector field of view. In our case, we used the central one third of the pixels in both dimensions, i.e. about 15 degree horizontal and vertical field of view.

6.2 Inventory Control

The warehouse scenario for inventory management was introduced in Section 1. Here we consider a specific interaction, with results illustrated in Figure 12. A warehouse manager uses the handheld device to annotate food products, and another employee subsequently retrieves the annotation. The manager first uses an RF broadcast to wake up the tags, then uses the handheld device to select and transmit the RF query *find-products* that will expire within two days. Selection is done by projected menu or a projected action-button as described in Section 4.2. A tag that detects it is in the projector beam as in Section 3.1, checks its expiry date and sends an RF transmission of its expiry status back to the handheld device. The projector system determines the 3D location of all illuminated tags as in Section 3.2, and beams green circles on those objects (tags) that are not about to expire, and red circles on others. The manager selects a subset of red-circled objects, by region selection as in Section 4.2, and the circle illumination changes to white. The red and green circles on the other tags remain stabilized on the physical surfaces throughout, Section 4.1. The manager selects and transmits the RF request *mark-selected-objects* for removal, and the selected tags record the request internally. A subsequent user, from a different projector view, selects and transmits the RF query *show-items-requiring-action*. The projector illuminates all annotated items with colored circles, and projected augmentation indicates the actual annotated instructions.

Figure 12: Warehouse scenario implementation. (a) Manager locates items about to expire (marked in red circles) and (b) annotates some of those items (marked in larger white circles) (b) Employee retrieves and views the same annotations from a different projector view.

7 Modes of Tag Deployment

The work has been motivated in terms of the commercially-important application of inventory control. But we believe that photo-sensing tags may have many innovative uses, and here we outline broad modes of deployment. Note these are speculative uses, not work done.

Single tags: A single tag per object is adequate for many situations e.g. consider installed projectors projecting product augmentation in a retail store, say on a wall of sports shoes. Each shoe transmits identity, and the projected display automatically adjusts for the spacing of the shoes, with dynamic update of the display as shoes are removed or repositioned.

Multiple tags on a single rigid object: Multiple tags allow computation of an object's 3D pose. Consider the packing task for objects that are tagged with their physical dimensions stored on tags. Since the system can compute the relative pose of each object, it can automatically suggest the next object to be packed and guide its placement by projecting markers, and track what the user is doing. It is typically difficult to determine what the user is doing with a vision-based system alone. As a second example, an environment can be tagged to support precise navigation and task guidance of an autonomous robot .

Ad-Hoc Distributed tags: Some scenarios can involve fairly ad-hoc distribution of tags. Consider a set of tags, each with a temperature sensor, on a surface - this sparse scalar field can be interpolated and projected back onto the surface to show temperature variation, with user selection of a specific point for a readout of the interpolated temperature value. The same tags can be used to find the deformation in the shape of that surface over time, without using a fixed camera or projector.

Interaction between tagged objects: with tags on multiple objects we can record history of placement among neighboring objects.

8 Discussion

Hardware Issues: Optical communication between projector and tag can be affected by ambient light. Wavelength division multiplexed communication is commonly used to solve this problem (e.g. TV remote and IR photo-sensor). The optical communication also gets noisier as projector-tag distance increases, and as the photo-sensor gets dirty.

Projected augmented reality can be adversely affected by ambient light, is poor on low albedo or specular surfaces, and it does not allow private display. Obtaining a visible projection with a small projector strains power requirements - a possible solution is for the projector to beam an image to a bistable display, like eInk, and then switch off while the bistable display continues to show the intended image.

Interface Issues: The handheld projector system lacks text input. Options include hybrid LCD panels for graffiti input, or a speech interface. New types of text-input solution are also appearing e.g. typing onto a projected laser pattern keyboard with finger placement detected by IR sensing [Canesta 2002].

Future Directions: Industry is already addressing range and power issues for passive-RFID. Photosensing tags provide a new type of imaging using distributed single pixel cameras and can solve a variety of computer vision problems. Tag photo-sensors can be modified in various ways – to do motion sensing, by replacement with 2D cameras, or by being made to be sensitive to a narrow band of wavelength so that the projector can transmit optical signals with wavelength division multiplexing. Light emission on passive tags is difficult due to power constraints, but active tags can use LEDs for optical feedback [Moore et al. 1999]. Active actuated surfaces could support scenarios which involve not just data update on tags, but also change in the physical position of objects.

The trend for projectors is miniaturization. Symbol Technologies has a small laser projector, with two tiny steering mirrors for vertical and horizontal deflection [Symbol 2002]. Siemens has a 'mini-beamer' attachment for mobile phones [Siemens 2002]. LEDs are replacing lamps [Lumileds 2003] and reflective displays are replacing transmissive displays (DLPs, LCOS). Both lead to improved efficiency requiring less power and less cooling. DLP and LCOS projectors can display images at extremely high frame rates, currently 180Hz and 540 Hz respectively, but lack video bandwidth. High frame rate will improve image stabilization and allow tracking of moving tags with rapid binary or colored patterns. At the same time, MEMS will miniaturize inertial sensors.

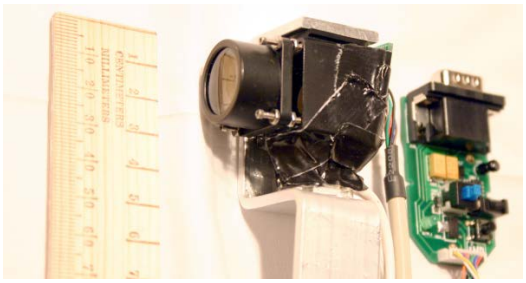

Figure 13: A micro-projector prototype we have built – to indicate the viability of designing a projector for truly handheld use. It is a 1-inch cube using an LED and LCOS imager to create a color animated projection at a distance of about 12-inches.

9 Conclusion

The embedding of computation power and data in everyday environments is now becoming feasible as cost and power problems diminish, and the prototypes in this paper have shown innovative ways to interact with such ambient intelligence. We demonstrated how a wireless RFIG tag system can support geometric operations. We showed a way to extend passive projection of augmented reality to an interactive system, with navigation and update of the augmentation data. The work indicates some of the possibilities for blurring the boundaries between the physical and digital worlds by making the everyday environment into a self-describing wireless data source, a display surface, and a medium for interaction.

Acknowledgements We thank the anonymous reviewers for useful comments and guidance. We thank Cliff Forlines, Rebecca Xiong, Dick Waters, Joe Marks and Kent Wittenburg for stimulating discussions, Mamuro Kato, Takashi Kan and Keiichi Shiotani for providing motivating applications, Debbi vanBaar and many members of MERL for help in reviewing the paper.

References

ABELSON, H., ALLEN, D., COORE, D., HANSON, C., HOMSY, G., KNIGHT, T., NAGPAL, R., RAUCH, E., SUSSMAN, G., AND WEISS., R. 2000. Amorphous computing. In *Communications of the ACM*, vol. 43(5), 74–82.

AZUMA, R., BAILLOT, Y., BEHRINGER, R., FEINER, S., JULIER, S., AND MACINTYRE, B. 2001. Recent Advances in Augmented Reality. In *IEEE Computer Graphics and Applications*, vol. 21(6), 34–47.

BEARDSLEY, P., RASKAR, R., FORLINES, C., AND VANBAAR, J. 2004. Interactive Projection. TR 2004/042, MERL.

BEARDSLEY, P., VANBAAR, J., AND RASKAR, R. 2004. Augmenting a Projector-Camera Device with Laser Pointers. TR 2004/035, MERL.

BIMBER, O., FROLICH, B., SCHMALSTIEG, D., AND ENCARNARO, L. M. 2001. The Virtual Showcase. *IEEE Comput. Graph. Appl. 21*, 6, 48–55.

CANESTA, 2002. Miniature Laser Projector Keyboard. http://www.canesta.com.

FISCHLER, M. A., AND BOLLES, R. C. 1981. Random Sample Consensus: a paradigm for model fitting with application to image analysis and automated cartography. *Commun. Assoc. Comp. Mach. vol. 24*, 381–95.

FOXLIN, E., AND NAIMARK, M. 2002. Shadow Effects of Virtual Objects on Real Surfaces with a Handheld Projector. *Unpublished*.

HARTLEY, R., AND ZISSERMAN, A. 2000. *Multiple View Geometry in Computer Vision.* Cambridge University Press.

HINCKLEY, K., PIERCE, J., SINCLAR, M., AND HORVITZ, E. 2000. Sensing Techniques for Mobile Interaction. In *ACM UIST CHI Letters*, vol. 2(2), 91–100.

HOLMQUIST, L. E., MATTERN, F., SCHIELE, B., ALAHUHTA, P., BEIGL, M., AND GELLERSEN, H.-W. 2001. Smart-Its Friends: A Technique for Users to Easily Establish Connections between Smart Artefacts. In *Ubicomp*, Springer-Verlag LNCS 2201, 273–291.

LUMILEDS, 2003. (Bright LEDs). http://lumileds.com.

MA, H., AND PARADISO, J. A. 2002. The FindIT Flashlight: Responsive Tagging Based on Optically Triggered Microprocessor Wakeup. In *Ubicomp*, 160–167.

MOORE, D. J., WANT, R., AND ET AL. 1999. Implementing Phicons: Combining Computer Vision with InfraRed Technology for Interactive Physical Icons. In *Proceedings of ACM UIST'99*, 67–68.

NAYAR, S. K., PERI, H., GROSSBERG, M. D., AND BELHUMEUR, P. N. 2003. A Projection System with Radiometric Compensation for Screen Imperfections. In *Proc. ICCV Workshop on Projector-Camera Systems (PROCAMS)*.

OMOJOLA, O., POST, E. R., HANCHER, M. D., MAGUIRE, Y., PAPPU, R., SCHONER, B., RUSSO, P. R., FLETCHER, R., AND GERSHENFELD, N. 2000. An installation of interactive furniture . In *IBM Systems Journal*, vol. 39(3,4).

PATEL, S. N., AND ABOWD, G. D. 2003. A 2-Way Laser-Assisted Selection Scheme for Handhelds in a Physical Environment. In *Ubicomp*, 200–207.

PATTEN, J., ISHII, H., AND PANGARO, G. 2001. Sensetable: A Wireless Object Tracking Platform for Tangible User Interfaces. In *Conference on Human Factors in Computing Systems (ACM CHI)*.

PINHANEZ, C. 2001. The Everywhere Displays Projector: A Device to Create Ubiquitous Graphical Interfaces. In *Ubiquitous Computing 2001 (Ubicomp"01)*.

RASKAR, R., WELCH, G., AND FUCHS, H. 1998. Spatially Augmented Reality. In *The First IEEE International Workshop on Augmented Reality (IWAR)*.

RASKAR, R., WELCH, G., LOW, K.-L., AND BANDYOPADHYAY, D. 2001. Shader Lamps: Animating Real Objects With Image-Based Illumination. In *Rendering Techniques 2001, Proceedings of the Eurographics Workshop in London, United Kingdom*.

RASKAR, R., VAN BAAR, J., BEARDSLEY, P., WILLWACHER, T., RAO, S., AND FORLINES, C. 2003. iLamps: Geometrically Aware and Self-configuring Projectors. *ACM Trans. Graph. (SIGGRAPH) 22*, 3, 809–818.

REKIMOTO, J., ULLMER, B., AND OBA, H. 2001. DataTiles: A Modular Platform for Mixed Physical and Graphical Interactions. In *CHI 2001*.

RINGWALD, M. 2002. Spontaneous Interaction with Everyday Devices Using a PDA Workshop on Supporting Spontaneous Interaction in Ubiquitous Computing Settings. In *UbiComp*.

SIEMENS, 2002. Siemens Mini Beamer. http://w4.siemens.de/en2/html/press//newsdesk_archive/2002/-foe02121b.html.

SYMBOL, 2002. Laser Projection Display. http://www.symbol.com/products/oem/lpd.html.

TELLER, S., CHEN, J., AND BALAKRISHNAN, H. 2003. Pervasive pose-aware applications and infrastructure. *IEEE Computer Graphics and Applications* (Jul).

THE COOLTOWN PROJECT, 2001. . http://www.cooltown.com/research/.

UNDERKOFFLER, J., ULLMER, B., AND ISHII, H. 1999. Emancipated pixels: Real-world graphics in the luminous room. In *Proc. Siggraph 99, ACM Press*, 385–392.

VERLINDEN, J. C., DE SMIT, A., PEETERS, A. W. J., AND VAN GELDEREN, M. H. 2003. Development of a Flexible Augmented Prototyping System. In *The 11th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision'2003*.

WANT, R., HOPPER, A., FALCO, V., AND GIBBONS, J. 1992. The Active Badge Location System. *ACM Trans. Inf. Syst. 10*, 1, 91–102.

WANT, R., SCHILIT, B. N., ADAMS, N. I., GOLD, R., PETERSEN, K., GOLDBERG, D., ELLIS, J. R., AND WEISER, M. 1995. An Overview of the ParcTab Ubiquitous Computing Experiment. In *IEEE Personal Communications*, 28–43.

WANT, R., HARRISON, B. L., FISHKIN, K., AND GUJAR, A. 1999. Bridging Physical and Virtual Worlds with Electronic Tags. In *ACM SIGCHI*, 370–377.

WANT, R. 2003. RFID, A Key to Automating Everything. In *Scientific American*, vol. 290(1), 56–65.

ZHANG, Z. 1999. A Flexible New Technique for Camera Calibration. *IEEE Pattern Analysis and Machine Intelligence 22*, 1330–1334.