Olwal, A., and Wilson, A. SurfaceFusion: Unobtrusive Tracking of Everyday Objects in Tangible User Interfaces.
Proceedings of GI 2008 (The 34th Canadian Graphics Interface Conference), Windsor, Canada, May 28-30, 2008, pp. 235-242.

SurfaceFusion: Unobtrusive Tracking of Everyday Objects in Tangible User Interfaces

Alex Olwal

School of Computer Science and Communication, KTH[1]

Andrew D. Wilson

Microsoft Research[2]

ABSTRACT

Interactive surfaces and related tangible user interfaces often involve everyday objects that are identified, tracked, and augmented with digital information. Traditional approaches for recognizing these objects typically rely on complex pattern recognition techniques, or the addition of active electronics or fiducials that alter the visual qualities of those objects, making them less practical for real-world use. Radio Frequency Identification (RFID) technology provides an unobtrusive method of sensing the presence of and identifying tagged nearby objects but has no inherent means of determining the position of tagged objects. Computer vision, on the other hand, is an established approach to track objects with a camera. While shapes and movement on an interactive surface can be determined from classic image processing techniques, object recognition tends to be complex, computationally expensive and sensitive to environmental conditions. We present a set of techniques in which movement and shape information from the computer vision system is fused with RFID events that identify what objects are in the image. By synchronizing these two complementary sensing modalities, we can associate changes in the image with events in the RFID data, in order to recover position, shape and identification of the objects on the surface, while avoiding complex computer vision processes and exotic RFID solutions.

CR Categories and Subject Descriptors: H5.2 [Information interfaces and presentation]: User Interfaces - Graphical user interfaces.
Additional Keywords: RFID, Computer Vision, Fusion, Tangible User Interface, Tabletop, Surface Computing.

1 INTRODUCTION

As the cost of large displays and computing hardware continues to decline, we can expect an increasing variety of computing form factors arrayed throughout our everyday environment. Interactive table systems, for example, exploit large displays in combination with sensing technologies to bring new experiences and interactions to tables, desks and other horizontal surfaces. Many of these systems augment our interactions with familiar everyday physical objects, lending them unique capabilities in the virtual world.

The bridging of the physical and virtual can be achieved in numerous ways. The DigitalDesk [21] tracks and augments paper with an overhead projector and camera, with which the user can interact using pens or their hands. The metaDESK [18] uses a rear-projected surface where vision-based tracking is performed with an IR-camera under the surface. It recognizes objects by the 2D projection of their geometry. SenseTable [9] uses electromagnetic tablets under the surface to sense special "pucks" and supplements them with front-projected graphics. Augmented Surfaces [14] use an overhead camera to track and identify objects with attached visual markers, and adds projected imagery. PlayAnywhere [23] expands on this scenario with a portable setup, where interaction takes place using the hands, or with objects tagged with visual codes for identification. BlueTable [22] and LightSense [8] detect and track spatially aware mobile devices on surfaces to enable context-sensitive feedback and interaction. In order to support simultaneous tracking of multiple devices, these systems rely on the ability of the device to actively communicate with the system. There is also an increasing interest in developing commercial systems that enable interaction with physical objects on interactive tabletops [2][7][10].

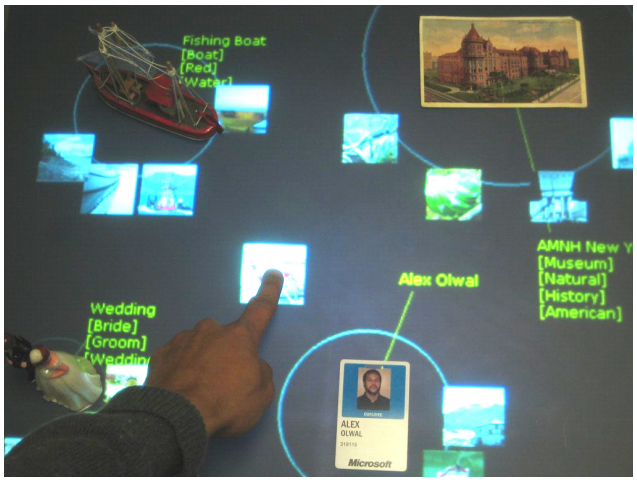

Figure 1: Our tangible user interface allows a user to intuitively interact with physical objects and their associated digital information. The system fuses RFID sensing and activities detected by simple computer vision techniques to identify and locate objects on the table.

Previous work demonstrates robust detection and tracking of everyday objects on an interactive tabletop using various types of fiducial markers and active electronics, such as electromagnetic sensors. For widespread adoption of such interacitive techniques it is desirable to enable the use of real-world artefacts that are only minimally modified.

In the present work we propose a lightweight, unobtrusive and generic sensing framework that enables scenarios where physical everyday artefacts can be augmented and associated with digital information, while supporting intuitive direct-manipulative interaction. (See Figure 1.)

Our contribution is a framework that enables detection and tracking of everyday objects without altering their appearance or employing exhaustive learning processes. We wish to provide as general mechanisms as possible for interactive surfaces, such that they may be used in a wide range of setups. The techniques

[1] PDC, KTH, 100 44 Stockholm, Sweden. alx@csc.kth.se
[2] One Microsoft Way, WA, USA. awilson@microsoft.com

should therefore not be dependent on exotic hardware or complicated configurations.

We make this possible by fusing activity detection in the RFID and computer vision domain, combined with classical correspondence tracking. The work leverages the respective strengths of RFID (identification) and computer vision (location) by combining them in a complementary way, such that arbitrary, unobtrusively tagged objects can be robustly sensed.

We also introduce the *Frame Difference Algebra* (FDA), a set of minimal image processing operations for vision-based detection of scene changes, which allows the fusion framework to avoid complex computer vision-based recognition techniques. Instead, it analyzes the temporal correlation of RFID tags and corresponding objects detected by the camera to establish the identity of shapes in the scene. The combined use of vision and RFID also allows us to create a set of techniques that integrate well with existing rear-projected tabletop systems; these systems often use diffuse projection surfaces that make imaging of fine features difficult. In contrast to most previous work, our sensing technology does not block the camera or projector in such setups.

We believe that by limiting the system to a simple set of techniques, the potential of the fusion with RFID may be illustrated in the most fundamental way. While we acknowledge the many sophisticated object recognition techniques available today, we would like to make as few assumptions about features available from computer vision processes, preferring instead to rely more on the fusion process.

We discuss related work in Section 2, followed by our activity sensing techniques in Sections 3, 4 and 5. The fusion process is described in Section 6. In Section 7, we detail how we extended our fundamental techniques for continuous tracking, and present a Tangible Image Exploration application to demonstrate the functionality of the framework in Section 8. Finally, we provide Future Work in Section 9 and Conclusions in Section 10.

2 Related Work

Vision-based systems have traditionally been popular due to their cost-effectiveness and flexibility in sensing many different types of real objects. While scene changes can be discovered through image processing techniques, recognition of arbitrary objects is significantly harder, and is very limited if we would like to identify objects through a diffuse projection surface which tends to blur fine detail. It is thus popular to simplify recognition by applying visual code markers to the object (e.g., QR code). Such schemes however require that the system have a clear line of sight, restricting how the user may position and orient the object. The markers themselves also alter the object's appearance in an undesirable way, making them inappropriate for truly ubiquitous deployment in everyday objects.

In contrast, radio frequency identification (RFID) tags may be placed on virtually any object and can be identified reliably [20]. We envision that standard consumer product bar codes will in many cases be complemented or replaced by RFID tags in the near future. RFID could be a key component of future ubiquitous computing scenarios — particularly in applications involving a large, or perhaps even virtually unlimited, number of unique objects. Since they can be applied discreetly and unobtrusively, they avoid cluttering the environment with visual markers. While visual codes can support a variety of bit depths, longer codes require more printed space, whereas RFID tags do not.

RFID technology reports the presence of a tag, but does not inherently provide means for locating a tag in space. While RFID technology was not initially designed for localization, several research projects investigate the use of custom or modified RFID readers and tags for positioning purposes, in addition to identification.

In Marked-up Maps [13], for example, a map is instrumented with RFID tags, serving as reference points for an RFID-reader-equipped PDA, which displays context-sensitive information. The ePro board [16] uses 480 readers on a 20×24 matrix with 3×3 cm squares. Parallel processing with 30 units each controlling 16 readers are used to reduce delay. In DataTiles [15], a matrix of RFID readers behind an LCD detects transparent tiles with embedded RFID tags in a 4×3 grid. They serve as graspable interaction devices and also allow interaction with an electromagnetically tracked stylus. The RFIG lamp [12] uses structured light to send unique binary codes to an RFID tag coupled with a photosensor. The RFIG tag transmits a binary code along with other data upon RFID interrogation, such that its position can be established. RFIG tags thus rely on line-of-sight to the reader. Boukraa and Ando [1] describe a system where RFID is used to retrieve a stored appearance model of an object such that computer vision-based registration can be simplified. Rahimi and Recht [11] present the tracking of RFID tags on a version of the SenseTable where 10 antennas are woven into a 30×30 cm surface — each reporting RFID signal strength for the tag. It is shown how a mapping can be learned such that 2D tag position from the 10 sensor readings can be inferred. The tracking of multiple tags is not discussed. Since each tag affects every other tag's signal strength as well as antenna transmission and reception, it is likely that the training task would become increasingly complex with multiple simultaneous tags. Krahnstoever et al. [6] modify an RFID reader to estimate orientation and 3D position of RFID tags, and associate this data with human motion tracking to infer high-level interactions between people and objects in an environment.

Many of these projects rely on multiple or modified RFID readers, active tags, or training processes. Our framework instead focuses on the combination of standard unmodified RFID equipment and vision techniques to avoid such complexity in order to support multiple configurations and off-the-shelf technology.

Existing display systems that are combined with RFID sensing tend to employ multiple short-range antennas covering the surface under the display (e.g., WACOM tablets). Such techniques are inappropriate for a rear-projection tabletop system since their antennas and associated electronics would block both camera and projector.

3 Activity Sensing for Object Detection and Tracking

A general goal of the tracking and detection components in a tabletop system is to recognize objects and track them on the surface. The appearance and interactive behaviour of such objects can be augmented by co-located projection and gesture sensing.

A camera is well suited to discover changes and motion in a video image using image processing techniques such as frame differencing. Vision-based recognition, on the other hand, is a complex and difficult task, especially if it the camera is placed behind a diffuse projection surface. Object recognition systems typically require training examples for each object we wish to recognize — a tedious and time consuming process. The complexity, recognition performance and runtime performance of these techniques varies widely, and often do not scale well with an increasing number of objects. Furthermore, it may be difficult or impossible to identify objects with similar appearance, such as multiple instances of the same consumer product, such as a camera or a mobile phone. For example, how could we, solely based on their appearances, distinguish the identity of two cell phones of the same model, belonging to two different persons?

The vision system also depends on the camera-object distance and on its viewpoint of the objects. Sufficient resolution is critical for detection and recognition, particularly in the case of small objects or small, dense visual codes. Finally, recognition performance tends to be sensitive to varying environmental conditions, such as lighting, or subtle changes in object appearance.

The use of special visually encoded markers is popular in these types of applications, but is not applicable in a general scenario where we want to support arbitrary objects without altering their appearance. Our framework thus focuses on synchronized activity sensing and consists of four parallel processes, which we discuss in the following sections:

1) Detection of activity in the camera image.
2) Detection of RFID tags.
3) Temporal synchronization of vision and RFID activities.
4) Frame-to-frame correspondence tracking for interactivity.

4 Activity Sensing: Computer Vision

For activity sensing, traditional object detection and tracking techniques are unnecessary, as we may instead focus on using computer vision to merely *detect changes* in the scene, such as the addition, removal and movement of objects on the table. This approach minimizes assumptions about lighting conditions, object appearance, tracking and other factors that lead to the complexity of many computer vision techniques.

We are especially interested in finding image capture frames that are representative of a change of state on the surface. Each such *still frame* summarizes the complete, stable state of the objects on the surface. By comparing a still with the previous still, it is possible to deduce whether an object has been added, removed or moved. In the following, we describe image processing operations available to detect such change. These do not rely on correspondence, object recognition or other complex image processing techniques.

A limitation of this approach is that only activity for one object at a time can be detected. But by combining the event detection with frame-to-frame correspondence tracking of the shapes (as discussed in Section 7), we enable fluid interaction and simultaneous manipulation of multiple objects. Our only restriction lies in the possible ambiguity caused by the unlikely event of the user adding two objects at the exact same time.

4.1 Event Detection with Connected Components

Given an image of the tabletop surface, under some assumptions it may be possible to determine the set of objects on the surface through traditional image segmentation techniques using binarization and connected components analysis. Once the set of objects is determined, it is relatively straightforward to detect object activity, particularly when the number of objects undergoing change is small.

A background image is stored when the scene is empty and absolute difference images are calculated from the background image and subsequent images. Candidate objects are detected through connected component analysis: groups of connected pixels are classified as distinct, independent objects. (See Figure 2.)

Add, *remove* and *move* events can be determined by comparing the list of connected components found in the current frame to that of the previous frame, using set difference operations. An increase in the number of objects (by one) indicates the addition of an object to the surface, while a decrease of one corresponds to object removal. Without resorting to object feature matching and recognition techniques, movement is detected as an object being removed and another object (the same) being added (i.e., the number of objects is unchanged). A related approach involves determining that a connected component from the previous frame and another from the current frame correspond to the same physical object if they appear at the same location in the image.

The detection of events using connected components, as described above, is valid only when each connected component corresponds to exactly one object on the surface. In applications involving real world objects, this assumption may not hold. Choosing the binarization threshold can in practice be very difficult, and may depend greatly on the nature of the objects' visual appearance, which in many cases may not be under the designer's control. A single object may lead to multiple connected components. In the following, we present an alternative approach which does not have these requirements.

4.2 Event Detection with Frame Differencing

Changes in the image may be mapped to surface activity by a *Frame Difference Algebra* (FDA) that detects scene changes such as the addition, removal or movement of an object, with a minimum of image processing operations. The FDA allows us to use fast and robust operations to detect scene changes that take place between two still frames.

Three images are used for the FDA calculations (See Figure 3): the background image (BG), the previous frame (P) and the current frame (I). Denoting the pixelwise absolute differencing operator Δ, we observe that $\Delta(I, P)$ leaves areas of the image which just changed. Furthermore, we may compute $\Delta(I, BG)$ and $\Delta(P, BG)$, which contain the objects that exist in the current and previous frame, respectively. We mask with $\Delta(I, P)$, obtaining images $A = \Delta(I, P)$ AND $\Delta(I, BG)$, and $D = \Delta(I, P)$ AND $\Delta(P, BG)$. Image A contains objects not present in the previous frame, but present in the current frame, indicating objects that just *appeared*. D contains objects present in the previous, but not in the current frame, indicating objects that *disappeared*.

The sum of pixels in the difference image indicates if there is a change significant enough for an event to have occurred. With the limitation that only one object can be added, removed or moved at a time, we have the following three cases:

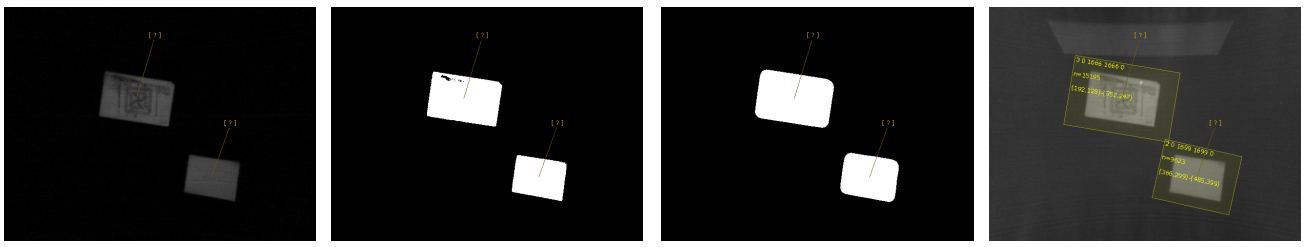

Figure 2: Segmentation. From left to right: a) Background subtraction. b) Binarization.
c) Noise reduction using Integral Image Sums. d) Connected component analysis yields two located objects.

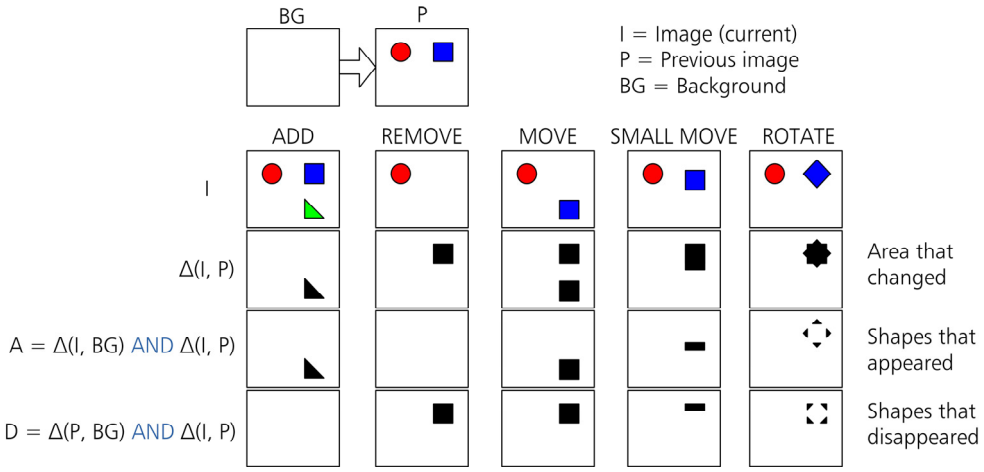

Figure 3: The Frame Difference Algebra uses absolute difference images and binary image operations for robust and fast detection of scene changes under the constraint that only one object is manipulated at a time. The background image (BG), current frame (I) and previous frame (P) are used in the calculations. By comparing the number of shapes in the resulting images with shapes that appeared (A) and disappeared (D), it is possible to infer whether an object was added, moved or removed.

- If sum(A) >> sum(D), then an object has been added.
- If sum(D) >> sum(A), then an object has been removed.
- If sum(A) ≈ sum(D), then an object has been moved.

The system stores an image mask for each new added object (A) and its associated RFID. As shown in Figure 3, the mask will only contain pixels corresponding to the object even if many objects are already on the surface. The masks are currently computed by binarizing a difference image. This requires thresholding, but since the difference image contains only one object (regardless of the number of objects on the surface) this can be set very generously. The moved object is determined by finding which of the stored masks representing the current objects on the surface is most similar to the new mask D. Almost any image comparison operation will do; we use the sum of absolute difference between two binary masks — a small difference will indicate a match. The mask stored for the object is now updated with mask A.

Note that this process does not rely on segmentation or tracking, and easily handles two objects right next to each other before one of them is moved, or when one object is moved right next to another.

The FDA allows us to store the timestamp and location for an object that has been added, removed or moved. Because the FDA involves simple, robust operations on the image, we obtain a fast detection mechanism that is straightforward to implement, and in contrast to many traditional vision-based systems, avoids assumptions on object shape, appearance, position and orientation.

The FDA can naturally be extended for more complex scenarios depending on the requirements of the application. In our case, we find it particularly useful to combine it with continuous tracking of objects, in order to allow fluid interaction and the manipulation of multiple objects. In Section 7, we describe such an extension where the FDA is complemented with vision-based correspondence-based tracking.

5 Activity Sensing: RFID

While our computer vision-based activity sensing provides us with the position and shape of our objects, it lacks means to identify them. The unobtrusive nature of RFID tags, which can be embedded into most any physical object, as well as the robust identification qualities of RFID, motivates the use of this complementary sensing technology. While there has been exhaustive investigations into custom, modified and special-purpose RFID readers in previous work [1][11][12][13][15][16], we find it interesting and practical to leverage commercially available, unmodified RFID readers and antennas, and passive RFID tags.

Whereas RFID tags can be detected reliably in the range of the antennas, our application requires that the tags are *only* detected when they are present on the tabletop surface, and thus visible in the camera image. This requirement ensures that activity detected by the vision system and the RFID reader can be synchronized.

5.1 Frequency of operations

There are four frequency groups in which commercial RFID technology operates; Low Frequency (LF, 125-148 kHz), High Frequency (HF, 13.56 MHz), Ultra High Frequency (UHF, 902-928 MHz) and Microwave (2.4 GHz). Differences in range (a few inches to hundreds of feet) and tag cost ($0.5-$25) can determine suitability to a given application. We find UHF technology appropriate for tabletop systems given the larger tracking/display areas involved – in our case a 32×24 inches surface. Another important motivation for using the longer range UHF is that it is impractical to place sensing technology directly under the surface (as required for short range RFID) for rear-projected systems since it might block the camera and projector.

5.2 Reader

We use the XR400 from Symbol Technologies, a long-range UHF reader made for industrial use, such as the scanning of large pallets and items on conveyor belts. It can use 1–4 read points, where each consists of a transmitting and receiving antenna pair. The XR400 has a read rate of about 1 Hz, which is comparable to the performance of other products. The reader works by transmitting energy and sequentially reading the backscattered energy from the tags. It will sequentially perform reads for each of the read points and within each read point sequentially interrogate the three tag types (class 0, class 1, class 1 gen2). The read rate is therefore affected by the number of active read points, number of tag types currently set to detect, and finally, the number of tags present in the system. We reduce delay in the system by using only one read point and one active tag class. The number of simultaneous tags in our system ranges from 1–10, matching the expected number of objects in a tangible tabletop interface. Additionally, we were advised by the manufacturer to set the reader in “conveyor belt mode” to further increase the performance. While the reader will not report signal strength of a tag, it is possible to attenuate the transmitted energy to limit the effective read range.

5.3 Antenna Configurations

The antenna type must be chosen carefully to support tabletop applications. It is desirable that the antennas can be placed unobtrusively, ideally integrated with the surface, and do not interfere with a rear- or front-projected imaging system.

5.3.1 Wire-loop antenna

In our efforts to limit the RFID readings to the surface, we investigated the use of a design consisting of two custom elongated transmitting and receiving wire loops, placed directly on the surface, on opposite sides of the area to be monitored. (See Figure 4.) With full gain we obtained a working design where the tags were detected on the surface only, as desired. The sensing area measured approximately 32×12 inches, which is about half the size of our target display. Readers that support multiplexed transmission and reception on the same antenna could potentially address this problem by doubling the sensing range. Our reader requires a separate transmit (TX) and receive (RX) antenna, which we place on opposite sides, whereas a multiplexing reader could have two TX/RX antennas placed on opposite sides such that each antenna need only cover half of the surface. The advantage of the wire loop antenna is that it effectively restricts the RFID sensing to just the surface, but our design was not powerful enough to cover a sufficiently large area.

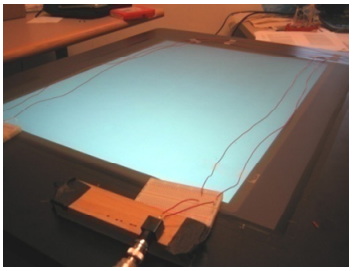

Figure 4: The rear-projected setup with our custom-made wire-loop antennas. RFID tags are detected only on the surface and in an area about half the size of the display due to limited range of the antenna.

5.3.2 Area antennas

Our second configuration, shown in Figure 5, has the advantage of using commercially available *area antennas*. A transmitting and receiving area antenna is placed opposite one another, under the surface and angled towards the center, such that they monitor the display. These antennas are powerful and work well with the reader set to 20% gain. It is important that the antennas are placed at a sufficiently steep angle, otherwise an object that is held directly above the surface for an extended period of time might be prematurely detected.

The area antennas are significantly more powerful than our wire loop antenna and can easily cover larger surfaces.

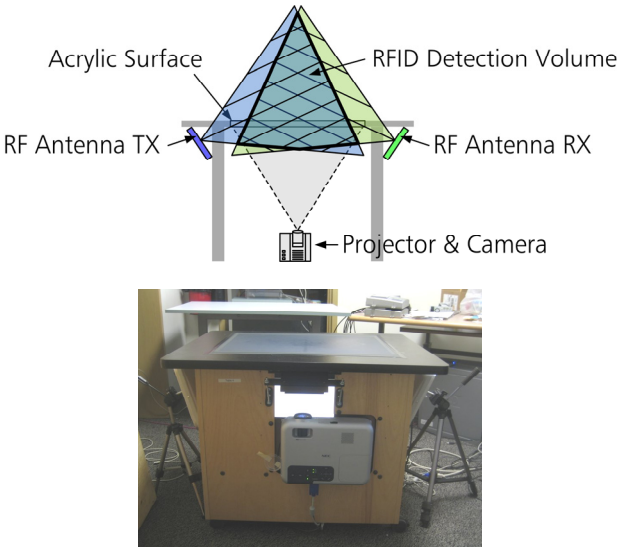

Figure 5: Area antennas can be unobtrusively added to front- and rear-projected systems. RFID tags are detected in the overlapping volume of the transmission (TX) and reception antenna (RX).

6 Fusion Framework

The vision-based event detection gives us accurate shape and position information for objects in the scene. Simultaneously, our RFID sensing accurately identifies present objects. The fusion framework provides a mechanism to synchronize the information from these two modalities such that each object can be identified.

We employ a database of events where detected vision and RFID events are continuously stored. When a new object appears, disappears or is moved, as detected by our image processing techniques, we add a timestamped entry, with a reference to the corresponding still image in the FDA. Similarly, we store a timestamped event when an RFID tag appears or disappears. Our database thus contains all state changes that have occurred, such that the state of objects on the surface may be retrieved at any time.

When a new event occurs in either modality, a matching process searches backwards in time for a corresponding, unmatched event in the other modality. If found, the two events are marked as

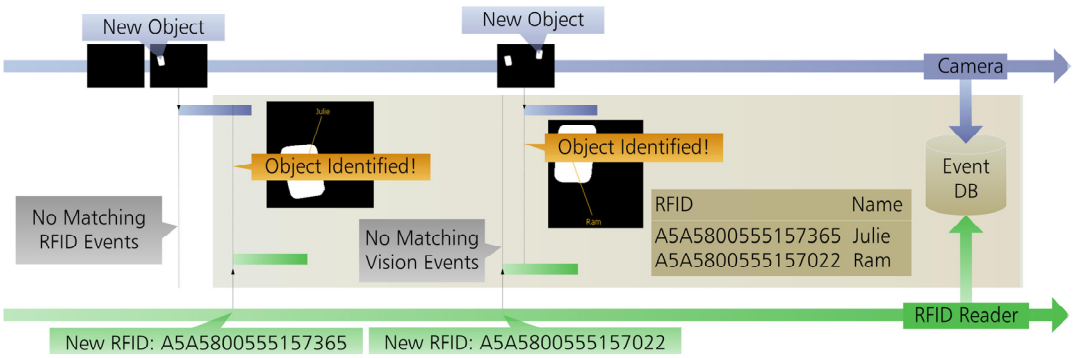

Figure 6: The fusion pipeline. As new events appear, the system tries to match them with previously unmatched events in order to associate localize and identify shapes on the surface.

matched and the still image from the FDA is associated with the RFID data. (See Figure 6.) Upon identification we perform a lookup in an object database, where the RFID is associated with additional metadata, such as the name of the object. That information can now be displayed at the location of the shape, as indicated by the FDA.

This fusion process suffers if a match was made incorrectly or missed altogether due to a great difference in time between RFID and vision events, RFID read failure or large spurious events in the vision system. We note that this has not been a problem in our experiments, due to the use of still images, each of which represents the stable state of surface objects. We also emphasize that typical tabletop scenarios do not have a high rate of objects being placed and removed from the surface, for practical reasons. It is thus likely that the user will introduce a smaller number of objects to work with. But while the working set might be kept small, there is a simultaneous need to be able to support a very large number of objects without having to prepare them or train the system for each.

7 Continuous Tracking

While our activity sensing provides us with a mechanism to identify and track objects in the scene, it limits the activity to one object at a time and only in-between still frames. These restrictions allow the approach to work in front-projection systems. While there are ways in which we can enable interaction in a front-projection system, the required computer vision reasoning becomes significantly more complex [23].

Our rear-projected system allows fluid tracking of objects while they are in motion, since the hand does not occlude objects it interacts with (from the camera's point of view) and shapes touching the surface can be robustly and reliably detected.

We employ frame-to-frame correspondence tracking to associate moving objects with the same ID they had in the previous frame. Correspondence is determined by computing the distance of a given object to every other object on the surface, such that a shape in the new frame inherits the ID of the closest shape in the previous frame, given that it is not a newly introduced object. The correspondence can be extended with more sophisticated methods, such as common pattern and template based tracking techniques.

The continuous tracking effectively addresses limitations of the FDA, such that multiple objects can be simultaneously manipulated on the surface, enabling responsive and fluid interaction.

8 Application: Tangible Image Exploration

In the spirit of Tangible User Interfaces [5][17][19], we developed a prototype tabletop application, which in contrast to previous work, uses visually unaltered objects with minimal instrumentation. It is based on the previously described techniques and makes use of our fusion framework, continuous tracking and touch screen interaction in a rear-projected setup, as shown in Figure 7.

In our application, imagery is downloaded from the Internet and projected next to various objects as they are placed on a table, as shown in Figure 1 and 8. To copy an image of interest, users place a personal item on the table, such as a badge, and drag the image to it. The image is then copied to the user's personal folder on the network. The next time the badge is placed on the table, the previously stored images appear.

When a new object is detected by the fusion module, we perform a lookup in an object database that stores information about the object type. Currently there are three types of tagged objects; *Query*, *container* and *operator* objects.

Query objects use pre-stored parameter values such as associated keywords. When a query object is detected on the table, the keywords are retrieved from the database and used in a search on the online Flickr photo database (http://www.flickr.com/). Matching images are then downloaded and appear around the object. Users can interact with the images by changing their size and moving them, as well as dragging them to other objects on the table.

Container objects act as a physical handle to a collection of digital images. They can also be used as symbolic links to physical storage, such as a shared network folder or a USB drive. They present all images currently stored in the represented location, and as new images are dragged to them, they are copied to the represented physical storage.

Operator objects execute a specific function on a dropped image. An ashtray, for example, can represent a trashcan, such that an image is deleted when dragged to it.

There are several ways in which our prototype application could benefit from additional functionality. In order to support more complex queries, we need mechanisms for authoring tags and keywords, as well as introducing more operators. Similarly, while the current prototype is limited to performing queries returning existing photos stored on Flickr, it would be beneficial to allow photos to be transferred directly from a portable device such as a digital camera or a camera phone, in the spirit of BlueTable [22]. New photos can spill out onto the table when the camera is placed on it, and associated to other objects (thus assigning tags to the photo), or deleted. We are also interested in other applications, such as a tabletop slideshow controlled by the configuration of the objects placed on the table. The relative position of the various objects on the surface can be used to build a database query, where we can extend the discrete control in previous work [17] with our continuous 2D parameter space.

There are many ways in which we can make use of the digital surface as a platform for extending and augmenting physical objects, both thanks to the availability of a large display, and also given new means for interaction through a multi-touch sensitive surface. Editing documents and pictures on a device would, for instance, be much easier if we could extend the interface to the surface. Linked service manuals that dynamically visualize the

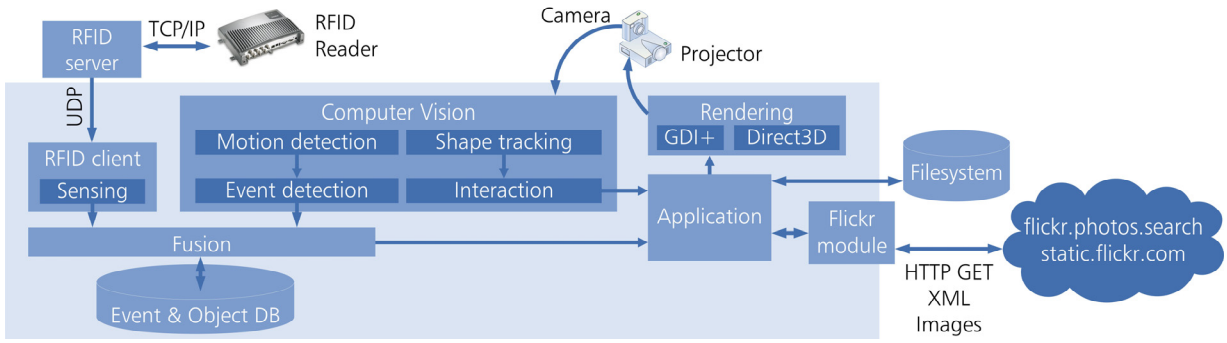

Figure 7: Application system architecture, including object database.

functionality associated with parts on a camera is another example.

9 FUTURE WORK

The simple fusion technique presented in this paper is clearly not limited to RFID and vision modalities. While we would like to improve the performance of these existing modalities, it may be interesting to consider complementing RFID and computer vision with additional sensing modalities. We emphasize that the modularity of the approach demonstrates that the combined power of multiple modalities can achieve satisfactory results with less effort.

9.1 Extracting data using the RFID reader

RFID readers can potentially provide information that goes beyond the detection of a tag's presence. Such capability may be useful in multimodal fusion systems.

Previous work [11][6] has shown ways in which RFID readers can sense additional properties, such as orientation or the location of a tag in space. Currently, such approaches require training or access to more data than is typically exposed in commercial systems. While the RFID technology we used is designed for common commercial applications and does not expose a number of low-level features, we find its availability as a commercial product compelling. It would be especially advantageous to use a more advanced commercial reader that provides signal strength, as this would allow more sophisticated reasoning about sensed objects on the surface. Higher read rates would improve overall system performance and interactivity. The ability to transmit and receive on the same antenna could allow twice the number of read points, increase sensing range and simplify antenna design.

We explored a set of features that might be useful in inferring data about objects in the system, especially in combination with our fusion framework. Most of these features are readily available without modification to the reader or tags:

9.1.1 Signal strength

Signal strength might be used as a coarse indication of distance, refined by subsequent fusion with other modalities, or for detecting interaction with the tagged object [3].

9.1.2 Response rate

Most commercial readers do not report signal strength, including the device we used. Fishkin et al [3] however discuss how signal strength can be approximated with the response rate, the number of successful responses divided by the number of attempted polls.

9.1.3 Time-multiplexed gain attenuation

Some readers provide software control over gain attenuation at runtime. We achieved a radar-like functionality by increasing the energy over a number of reads. The major drawback of a 1 Hz update rate is that a scan with 10 different energy levels takes 10 seconds, which is prohibitive for most interactive applications.

9.1.4 Multiple antennas/readers

Depending on the available data, one can use multiple antennas and readers with varying position, orientation, gain and other parameters in order to extract more information about the tags being read. For example, signal strength from multiple antennas could be used for coarse position triangulation of a tag.

9.2 Exploiting RFID tag specific properties

There are many factors that determine whether a tag is read successfully. Occlusion, tag geometry and orientation are examples of issues that can affect how much energy the tag can absorb and reflect through backscattered energy. This could bring additional factors to help the fusion process.

9.2.1 Geometry

Tag antenna design varies greatly and is critical to how well the tag absorbs and reflects energy. Besides using different designs, we can also modify the performance, by cutting off half of the antenna, for example. We found this to be useful when it is desirable to limit the reading range of certain tags. A related possibility is to place multiple tags with varying geometry on an object, and use the resulting variation in sensitivity as an indication of signal strength.

9.2.2 Orientation

Tag geometry also plays an important role for how detection performance varies with orientation. Tag detection is typically less reliable when the (flat) tags are oriented perpendicular towards the antenna, rather than face-on. We also discovered that elongated tags are not as robustly read as symmetrical tags with a 90 degree orientation when used with our wire-loop antenna. Multiple orientation-sensitive tags on an object could both increase robustness and provide an indication of orientation.

9.2.3 Occlusion

Like any RF technology, sensing degrades in the presence of liquids or metal. Given that the human body is largely composed of water, we have been able to reliably block a tag from being read by occluding it with our hand. By tracking the hand in the camera image, we might correlate that motion with the varying readability of the blocked tags, such that the tags will also act as sensors.

9.2.4 Memory

The on-board memory on RFID tags has already surpassed many applications' requirements for storing the identification number and we can expect it to continue growing. By having a passive tag with general purpose on-board storage we could make use of this to aid the recognition process. For instance, we envision that the interactive surface could update the object's tag with detected tracking features as it learns new properties about the object. Instead of storing the object features in a central repository (as in [1]), we could store all information directly with the object itself. Even simple information, such as shape, size and color, could provide valuable information to the fusion framework, such that objects could be more robustly disambiguated on the surface.

9.3 RFID sensing for tabletop systems

We have experimented with both reader parameters and tag properties and conclude that the most important features are interactive rates of operation and robust tag detection on the surface only. Response rate and time-multiplexed gain attenuation are thus not particularly useful, given the severely reduced update rate. Signal strength is perhaps the most promising technique, but is unfortunately still not exposed in most commercial readers. We also note that the use of multiple readers, and the performance-affecting tag properties, requires training specific to the particular configuration and tags used. The control of tag geometry did however prove useful in order to minimize accidental tag detection above the surface.

9.4 Global fusion

Building upon our current fusion framework, we envision a global fusion process that is capable of resolving ambiguities over a history of activity, such that the reasoning becomes an integrated,

Figure 8: Our tangible image explorer enables interaction with unobtrusively tagged physical objects on interactive surfaces.

online probabilistic process. Specifically, motion and RFID activity may be modelled for each pixel of the input image. Each tag would have a probability distribution in the image, where each pixel stores the likelihood of its association with a specific RFID tag. These probability images would be continuously updated as new events occur, allowing them to update the model, resolve ambiguities and repair incorrect associations.

10 Conclusions

We have presented an approach that combines RFID technology and image processing to support tangible interactions with visually unaltered everyday objects on interactive surfaces. We use a camera to detect shapes and motion in the video image, whereas an RFID reader senses tag presence. By synchronizing these two sensing modalities in time, we can associate a located shape with the ID provided by the RFID reader. The ID can be used to index into a stored table of known objects far larger than what is practical with most visual codes. Our approach takes advantage of each modality's strength; the vision component monitors the camera image for activities of interest, while the RFID component monitors the RF domain to sense tags. We also note that the unobtrusive nature of our techniques allows them to coexist with other approaches, such as fiducial tracking, for example if additional robustness and redundancy would be desired.

The fusion of these complementary sensors allows the use of a single standard RFID reader and robust vision techniques. The frame difference algebra, for example, makes very few assumptions about the nature and appearance of objects, and is thereby widely applicable. Likewise, the use of standard RFID equipment allows the unambiguous identification of multiple physical objects of almost any type, using inexpensive and unobtrusive tags that we expect to be ubiquitously embedded in future products, replacing today's visual barcodes. We believe that this approach provides new opportunities in bridging physical and virtual worlds, using interactive surfaces and everyday physical objects.

References

[1] Boukraa, M. and Ando, S. Tag-based vision: assisting 3D scene analysis with radio-frequency tags. Image Processing 2002 (2002), I-269-I-2722.

[2] Dietz, P. and Leigh, D. DiamondTouch: a multi-user touch technology. UIST '01 (2001), 219-226.

[3] Fishkin, K., Jiang, b., Philipose, M. and Roy, S. I Sense a Disturbance in the Force: Unobtrusive Detection of Interactions with RFID-tagged Objects. IRS-TR-04-013. (2004).

[4] Gonzalez, R.C., and Woods, G. Digital Image Processing. Addison Wesley (1993).

[5] Ishii, H. and Ullmer, B. Tangible bits: towards seamless interfaces between people, bits and atoms. CHI '97 (1997), 234-241.

[6] Krahnstoever, N., Rittscher, J., Tu, P., Chean, K. and Tomlinson, T. Activity Recognition using Visual Tracking and RFID. WACV/MOTIONS '05 (2005), 494-500.

[7] Microsoft Surface. http://www.microsoft.com/surface/. (Sep 2007).

[8] Olwal, A. LightSense: Enabling Spatially Aware Handheld Interaction Devices. ISMAR '06 (2006), 119-122.

[9] Patten, J., Ishii, H., Hines, J., and Pangaro, G. Sensetable: a wireless object tracking platform for tangible user interfaces. CHI '01 (2001), 253-260.

[10] Philips Entertaible. http://www.research.philips.com/initiatives/entertaible/. (Sep 2007).

[11] Rahimi, A, and Recht, B. Estimating Observation Functions in Dynamical Systems using Unsupervised Regression. NIPS (2006).

[12] Raskar, R., Beardsley, P., Dietz, P., and van Baar, J. Photosensing wireless tags for geometric procedures. Commun. ACM 48, 9 (2005), 46-51.

[13] Reilly, D., Rodgers, M., Argue, R., Nunes, M., and Inkpen, K. Marked-up maps: combining paper maps and electronic information resources. Personal Ubiquitous Comput. 10, 4 (2006), 215-226.

[14] Rekimoto, J. and Saitoh, M. Augmented surfaces: a spatially continuous work space for hybrid computing environments. CHI '99 (1999), 378-385.

[15] Rekimoto, J., Ullmer, B., and Oba, H. DataTiles: a modular platform for mixed physical and graphical interactions. CHI '01 (2001), 269-276.

[16] Sugimoto, M., Kusunoki, F,. and Hashizume, H. Supporting Face-to-face Group Activities with a Sensor-Embedded Board. CSCW Workshop on Shared Environments to Support Face-to-Face Collaboration (2000).

[17] Ullmer, B., Ishii, H., and Jacob, R. Tangible Query Interfaces: Physically Constrained Tokens for Manipulating Database Queries. INTERACT'03 (2003), 279-286.

[18] Ullmer, B. and Ishii, H. 1997. The metaDESK: models and prototypes for tangible user interfaces. UIST '97. (1997) 223-232.

[19] Ullmer, B., Ishii, H., and Glas, D. mediaBlocks: physical containers, transports, and controls for online media. SIGGRAPH '98 (1998), 379-386.

[20] Want, R., Fishkin, K. P., Gujar, A., and Harrison, B. L. Bridging physical and virtual worlds with electronic tags. CHI '99 (1999), 370-377.

[21] Wellner, P. Interacting with paper on the DigitalDesk. Commun. ACM 36, 7 (1993), 87-96.

[22] Wilson, A. D. and Sarin, R. BlueTable: connecting wireless mobile devices on interactive surfaces using vision-based handshaking. Graphics Interface 2007 (2007), 119-125.

[23] Wilson, A. D. PlayAnywhere: a compact interactive tabletop projection-vision system. UIST '05. (2005), 83-92.