# Real World Object Annotation for See-Through Displays

*Alois Ferscha and Markus Keller*

Institut für Praktische Informatik
Universität Linz
ferscha@soft.uni-linz.ac.at, markus.keller@students.uni-linz.ac.at

## Abstract

A tremendous growth of the Internet over the past years together with the growing availability of wireless communication technologies in the wide, local and personal area, and the maturing of miniaturized Internet hardware technologies like tiny web servers has fertilized the emergence of "smart appliances": wirelessly ad-hoc networked, mobile, autonomous special purpose computing devices, providing largely invisible support and services to users. Such smart appliances, usually operated under intuitive, rather implicit user interaction – i.e. they interact with their environment via a variety of sensor on the input side, and a vast of actuators on the output side – have started to populate the "real world" with "hidden" or "invisible" services, thus building up an "invisible world" of services associated with real world objects. With the embedding of invisible technology into everyday things, however, also the intuitive perception of "invisible services" disappears. We believe it has potential advantages to support the perception of smart appliance services via novel interactive visual experiences. We have developed and built DigiScope, a see-through based visual perception system for "invisible worlds" to support interactive theater experience in mixed reality spaces. In a case study, we demonstrate the use of DigiScope to percept the "invisible services" of a smart appliance which we have developed: SmartCase – a multisensor, wireless, embedded "carry-on" Internet appliance.

## 1 Introduction

A vast manifold of heterogeneous, small, embedded and mobile devices has emerged in the pervasive and ubiquitous computing landscape, characterized by the autonomy of their programmed behaviour, the dynamicity and context-awareness of services and applications they offer, the ad-hoc interoperability of services and the different modes of user interaction upon those services (Estrin et al., 2002). This is mostly due to technological progress like the maturing of wireless networking, exciting new information processing possibilities induced by submicron IC designs, low power storage systems, smart material, and motor-, controller-, sensor- and actuator technologies, envisioning a future computing service scenario in which almost every object in our everyday environment will be equipped with embedded processors, wireless communication facilities and embedded software to percept, perform and control a multitude of tasks and functions. Since many of these objects will be able to communicate and interact with global networks and with each other, the vision of "context-aware" (Dey, 2001) smart appliances and smart spaces (Roman & Campbell, 2000) – where dynamically configured systems of mobile entities by exploiting the available infrastructure and processing power of the environment – has become a reality. The individual utility of such services stems from being personalized, i.e. user centered and dynamically adapted to user preference, location aware, i.e. multimodal and multifunctional with respect to the environment, and time dependent, i.e. dynamic and with timely responsiveness. Common in the trend of context aware environments – often referred to as "smart appliances" or "smart spaces" – is that they interact with the user in a pro-active, autonomous, sovereign, responsible and user-authorized way. Common is also that the provision of their services is based on their ability of being aware of the presence of other objects or users, and being sensitive, adaptive and responsive to their needs, habits and emotions. Their services tend to become ubiquitously accessible via natural interaction. Embodied into real world objects like furniture, clothing, crafts, rooms, etc., those services are usually "invisible", i.e. cannot be perceived visually or orally, thus leaving the user "un-aware" of their presence.

We believe that the utility of many of those "invisible services" of smart appliances being better exploited when presented to the user in a more intuitive and natural way, thus raising the need for a better perception of smart environments by the user. To support people living in the real world populated with a variety of digital artefacts as created by the digital components in a smart environment, when acting, perceiving and interacting with objects in their environment, we propose a see-through based theatre experience of visual perception, seamlessly merging the artefacts of the real and the digital world. In Section 2 we present DigiScope, the 6DOF visual see-through tablet we have developed to support an intuitive "invisible service" – or more generally: "invisible world" – inspection. In Section 3 we present a use case for DigiScope. The invisible services of the smart appliance "SmartCase" – which has been developed as a demonstrator for our contextware framework (Ferscha, 2002) – will be inspected via DigiScope.

## 2 The DigiScope See-Through Invisible World Inspection System

The driving motivation for the design and implementation of a 6DOF see-through inspection tablet was to provide a seamless interface between real and digital artefacts, in order to create new and better human experience when observing the services of smart environments. Towards a seamless visual perception of the coexistence of a real world and its virtual augmentation, we exploit the metaphor of digital annotations for real world objects, and display these annotations along the line of sight to real world objects that are seen through the tablet.

The challenge for digital annotation visualization via a see-through display is to cope with the changes and interactions among physical and virtual world objects in a certain environment or scene in real-time. To be able to provide the necessary user fidelity, the monitoring and managing of objects in the scene demands mechanisms for real-time identification, positioning and tracking of physical objects, and needs adequate models of the objects of interest (Wand, Weiser, Mynatt 1998). As an example for identification and annotation of real world objects with virtual hyperlinks, some researchers have developed systems that sense physical entities in the environment and map them to a web browser (Pradhan et al., 2001) (Kindberg & Fox, 2002). Sensing technologies used for such linking and annotation purposes are infrared beacons, radio frequency ID tags (RFID) and bar codes, while the interaction with services in "smart spaces" is implemented based on the use of hand-held devices for the perception, resolution and manipulation of URIs (universal resource locators) of real world objects. Other approaches are using augmented reality system technology to enhance a person's view within a smart space by visually presenting the digital and real artefacts respectively. A typical approach is to annotate real objects with detailed virtual information in the users view (Okuma, Kurata & Sakaue, 2002), (Rekimoto & Ayatsuka, 2000), (Liu, Georganas & Boulanger, 2002). Another possibility is to augment the real world with virtual objects like in (Bobrich & Otto, 2002), where a map is extended with a 3D surface of a landscape to give the user a more realistic impression. Augmented reality (AR) within a “smart space” can also provide alternative opportunities to manipulate digital artefacts (Poupyrev et al., 2002).

We follow the augmented reality approach to display digital annotations of a physical reality, where the reality is perceived behind a see-through tablet. The tablet can be controlled by the spectator (user) in position (3D) and orientation (roll, pitch and yaw) respectively, while annotations are displayed on the tablet as decorations of the reality. Even invisible or hidden objects in the reality (consider a shirt or a key ring hidden in a suitcase) can be made “visible” with appropriate annotations. The implementation options for such a see-through AR based smart environment inspection system ranges from various kinds of video see-through displays to optical see-through displays. Video see-through AR displays typically consist of a HMD (head-mounted display) that displays a real scene video that is captured in positional sync with the orientation of the viewers head (tracked with a tracking system). The user gets the impression of being immersed into an environment by viewing both the real world scene (as a real-time video) and the annotation (as an overlay with graphical augmentations of the scene) on a small display placed in front of his eyes. Optical see-through AR displays avoid the use of real-time video to present the scene, but superimpose the real world with annotations that are displayed with projections using mirrors and/or semi transparent glasses. Yet another option are retinal scanning displays or virtual retinal displays (Viirre, Pryor & Nagata 1998), where a low power laser projects the scene augmentation directly onto the spectators retina. According to (Azuma et al., 2001) “there is still no see-through display that has sufficient brightness, resolution, field of view, and contrast to seamlessly blend a wide range of real and virtual imagery”. Above that, for smart environment inspection the limited fidelity as imposed by the HMD precludes user acceptance beyond occasional use. We therefore decided to avoid any approach involving an HMD, but instead build on holographic see-through technology. A recent non-HMD see-through AR display is presented in (Schwald, Seibert & Weller 2002), where a transparent 17” active-matrix LCD (TFT) screen is used as an output device. The transparency of the screen will however only be acceptable, if the observed scene is sufficiently illuminated.

For all those reasons, we have developed our own non-HMD see-through “mixed reality” (MR) display which we call DigiScope, because it serves a microscope for the digital annotations in real world scenes. DigiScope is the first approach to use a portable holographic screen as an MR inspection device. With DigiScope, the user is handling a holographic display tablet just like a 6DOF window that opens a view into the virtual world. The tablet is an optical see-though display which allows for a very natural viewing and scene inspection. To implement correct views into the scene, the angle and perspective of the DigiScope is being tracked, instead of tracking the position and orientation of the user. Thus the user is freed from any system hardware obstacles like HMDs, stereoscopic glasses, trackers, sensors, markers, tags, pointers and the such. DigiScope builds on hologram technology (www.dnp.dk) manufactured in a thin film and laminated to a transparent acrylic plate. The DigiScope responds to light rear projected in an angle from 30°-35° and is very selective about the direction of the arriving light: ambient light or even sunlight is ignored, black images appear transparent on the DigiScope. This allows to implement any scene annotation as a non-black overlay projected onto the DigiScope. To support free navigation in the secen, the DigiScope can be fully tilt and rotated in space by hand. The projecting beamer is fixed in the right projecting angle within a 6DOF mounting frame (see Figure 1), and is used to project the computer generated image encoding the scene annotation. For tracking the position and orientation of the DigiScope frame, an Intersense IS-900 (www.isense.com) is used. The mounting frame also holds also a video camera (Sony DFW-X700) for optical marker recognition and a touchpad for user interactions.

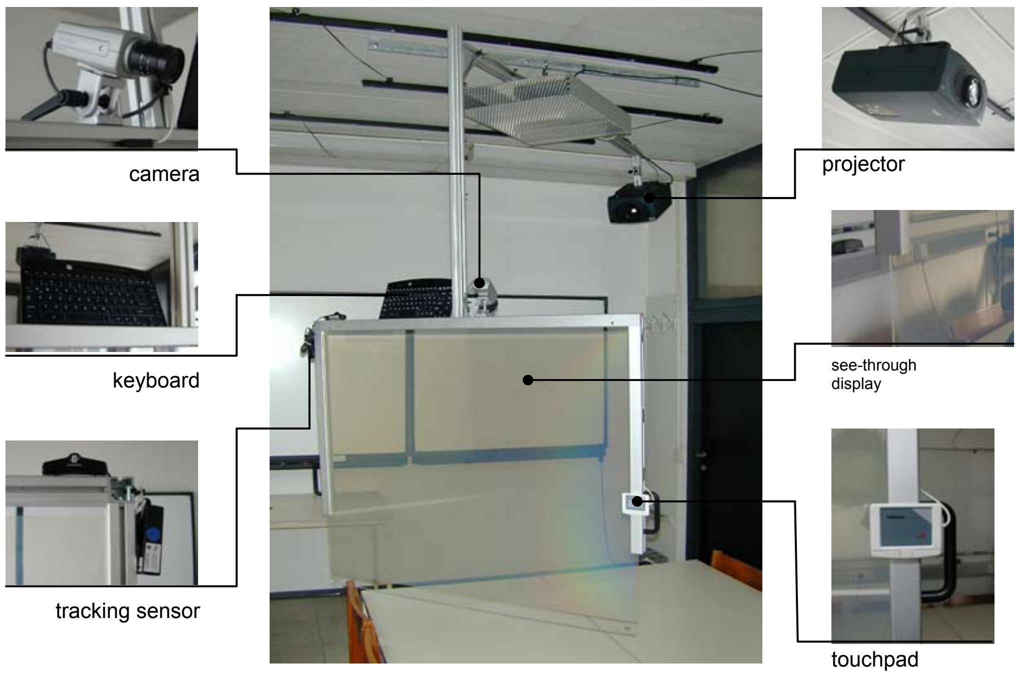



**Figure 1:** DigiScope frame with hardware details

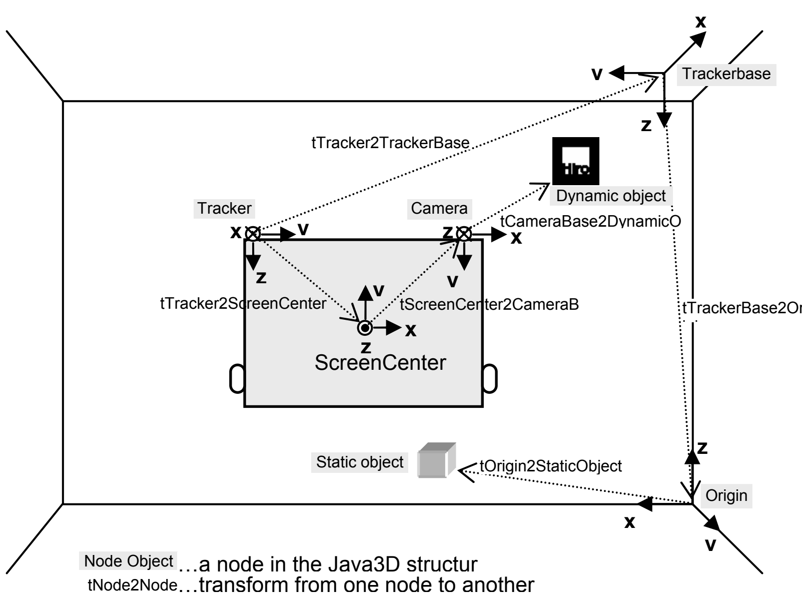
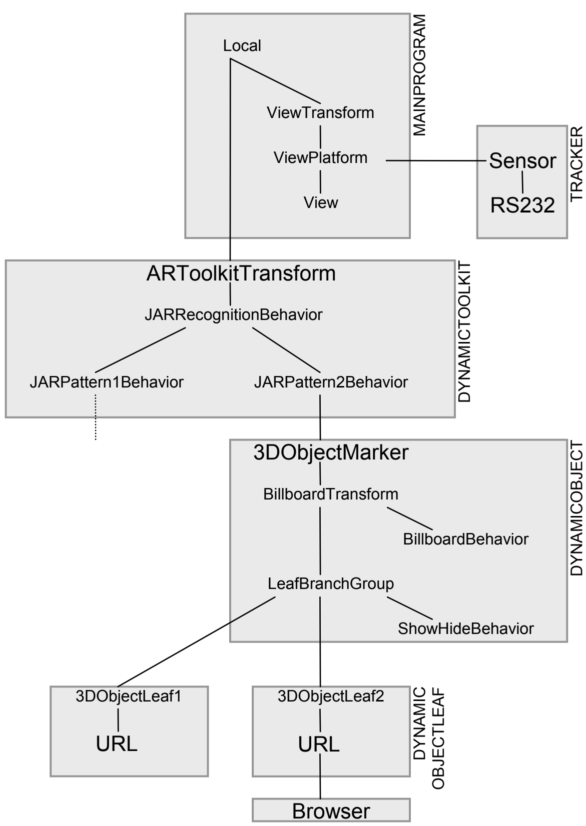



**Figure 2:** DigiScope hardware mapped to software archtiecture

The DigiScope software architecture is based on standard building blocks for AR application frameworks: (i) a 6 DOF tracking library for position and orientation tracking of the DigiScope frame, (ii) Java and Java3D for 3D scene modelling, rendering and implementing user interaction, and (iii) the ARToolkit (www.hitl.washington.edu) for visual object tracking and scene recognition. For visual object recognition based on optical markers the JARToolkit package (www.c-lab.de/jartoolkit), a Java wrapper class for the ARToolkit, has been used. In the current implementation, for the objects in a smart environment we distinguish (i) static objects which are not being (visually) tracked at runtime, but have a more or less static position in the scene, and (ii) dynamic objects, tagged with optical markers, tracked during in position and orientation at runtime. For scenes with an arbitrary number of static and dynamic objects, the DigiScope software frameworks at runtime creates visual presentations of predefined object annotations together with their corresponding hyperlinks into a 3D scene, which is in turn rendered by the Java3D engine and displayed (projected) onto the holo-screen. Object annotations can be any Java3D node, ranging from plain text to simple and very complex geometry, together with complex appearance attributes (e.g. for color, light, texture, etc.), and can exhibit arbitrary interpolator behaviours. The scene is rendered so as to give a true perspective view from the assumed position of the observers eye, the so called ScreenCenter (see Figure 3). Annotations are rendered perpendicular to the viewer, and Web-linking from within the scenegraph allows to use any information on the WWW for object annotation, and to access it during inspection. An apparent digital annotation of real world objects is the visual impression delivered to the viewer.




**MAINPROGRAM**: The main Java3D program. It contains the 3D universe and the virtual viewpoint of the user.

**TRACKER**: InputDevice class to interact with the tracker system to retrieve the position and orientation of the screen. This is necessary to provide the correct view.

**DYNAMICTOOLKIT**: Wrapper class for the ARToolkit. It handles the camera image and calculates the transform matrix for each optical marker. Each marker gets its own JARPatternBehavior.

**DYNAMICOBJECT**: Presents a 3D object at the position of the optical marker on the see-through display. Additional information can be shown or hidden by selecting the 3D object with the cursor. Appearing additional text information is always perpendicular to the viewer, because of the BillboardBehavior.

**DYNAMICOBJECTLEAF**: Additional information object which holds a describing text, an image, a 3D object and an URL. Selecting one of the 3DObjectLeafs will pop up the associated HTML page.

**BROWSER**: Shows the HTML file.

**Figure 3:** DigiScope software architecture

# 3 Inspecting the SmartCase with DigiScope

A context aware smart appliance, SmartCase, has been developed to illustrate our contextware software framework (Ferscha, 2002). The hardware for the SmartCase demonstration prototype uses an embedded single board computer integrated into an off-the-shelf suitcase (see Figure 4), which executes a standard TCP/IP stack and HTTP server, accepting requests wirelessly over an integrated IEEE802.11b WLAN adaptor. A miniaturized RFID reader is connected to the serial port of the server machine, an RFID antenna is integrated in the frame of the suitcase so as to enable the server to sense RFID tags contained in the SmartCase. A vast of 125KHz and 13,56 MHz magnetic coupled transponders are used to tag real world objects (like shirts, keys, PDAs or even printed paper) to be potentially carried (and sensed) by the suitcase. The SmartCase itself is tagged and possibly sensed by readers integrated into home furniture, car or airplane trunks, conveyor belts etc., so as to allow for an identification and localization at any meaningful point in space of the application. In addition, the SmartCase is equipped with optical markers so as to enable visual recognition and tracking with the ARToolkit framework.

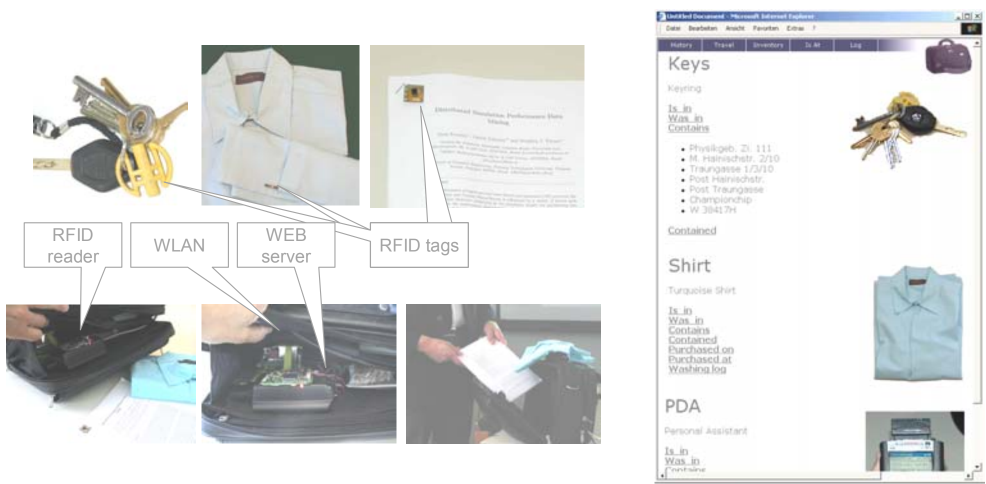



**Figure 4:** SmartCase technologies and inventory inspection via the WWW

Within our contextware software framework (Ferscha, 2002), an application specific abstraction of the real world is generated from three generic classes for persons, things, and places (McGarth & Mickunas, 2001). The reification of physical world objects and their relation among each other is expressed in RDF (Lassila & Swick, 1999). The person object represents the concepts about a person that are necessary to link their physical properties and activities to the virtual world. The thing object encapsulates the basic abstraction for objects relevant to the application and has instances like shirt, bag, etc. The place object is used to represent the essential features of a physical location like an office room or a waiting lounge. Places, things and persons may be related in a manifold of ways. The contextual interrelatedness among those object instances is expressed by a set of (bilateral) object relations within the resource description framework (RDF). Real world scenarios – like the situation of the SmartCase – are expressed as RDF statements, i.e.

denoted as subject, predicate and object triples. A subject in RDF is represented by a resource, that could be a URL, a simple HTML file, a URI reference etc., a predicate is represented by a property describing a resource or the relation among resources, and an object is represented by a property value. To represent state changes in the real world, a set of sensors (like the RFID reader embedded into the frame of the SmartCase) are used to dynamically modify a set of relations that hold among the corresponding RDF resources at runtime. Such relations are for example the **`owns`** relation to expresses ownership of a real world object by another, the **`contains`** and **`is_in`** relations to expresses geometrical containment of an object within another one, the **`contained`** and **`was_in`** relations trace the history of containment etc. A unique ID associated with every real world object is the ID encoded in its RFID tag. It is sensed by the RFID reader which triggers a script to update the involved object RDFs (Inserting e.g. the shirt into the suitcase would cause the RFID reader to identify the shirt tag and automatically update (among others) both the shirts RDF relation **`is_in`**, as well as the suitcases RDF relation **`contains`** by cross-referring URIs. Considering now the **`contains`** relation of the object SmartCase as an "invisible" inventory service, then, once an object (e.g. shirt) has been put into the SmartCase, this service can be queried to check whether the shirt is in the case or not. A straightforward way to access this information would be via a classical http interface to the embedded web-server, as shown in Figure 4, left. Observed via the DigiSpace however, changes to the SmartCase inventory – or in other words: the **`contains`** relation – are displayed as a graphical annotation of the real world. Figure 5 shows the process of inserting the shirt (3), a bunch of keys (4) and a PDA (5) into the SmartCase and the respective visual impression. A web-link associated with the shirts annotation is accessed via a touchpad click in (6), spawning a web browser with the shirts URL.

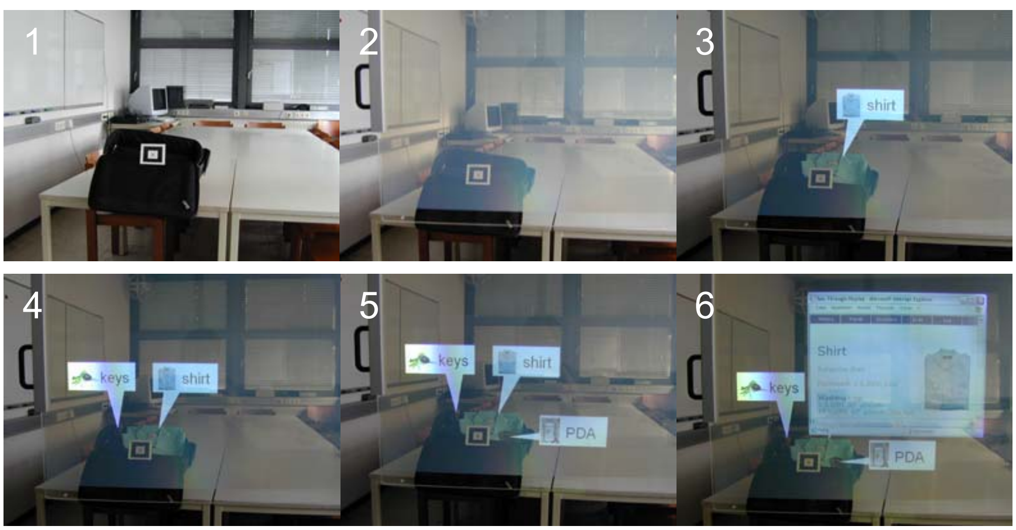



**Figure 5:** SmartCase inventory inspection via DigiScope

# 4 Conclusion

We have approached the emerging problem of developing intuitive interfaces for the perception and inspection of natural human environments populated with an increasing number of smart appliances in the pervasive and ubiquitous computing landscape together with their "invisible" services. Our DigiScope envisions a new type of MR interface with two main features: (i) a new exploration experience of the physical world seamlessly merged with its digital annotations via a non-obtrusive MR interface, and (ii) an integration of ubiquitous context-awareness and physical hyperlinking at the user interface level.

## References


Azuma, R., Baillot, Y., Behringer, R., Feiner, St., Julier, S., & MacIntyre, B. (2001). "Recent Advances in Augmented Reality", Computers&Graphics Vol. 21, No. 6 (pp 34-47).

Bobrich, J. & Otto, S. (2002). "Augmented Maps", IAPRS Vol. 34 Part 4, Ottawa, Canada.

Dey, A. K. (2001). Understanding and Using Context. Personal and Ubiquitous Computing, Special Issue on Situated Interaction and Ubiquitous Computing, 5(1).

Estrin D., Culler, D., Pister, K. & Sukhatme, G., (2002). ", Pervasive Computing 1(1):59-70.

Ferscha, A., (2002). "Contextware: Bridging Virtual and Physical Worlds", Reliable Software Technologies, AE 2002, Lecture Notes in Computer Science LNCS 2361, pp 51-64, Berlin.

Kindberg, T. & Fox, A. (2002). System Software for Ubiquitous Computing, IEEE Pervasive Computing, vol. 1, no. 1, Jan-Mar 2002, pp. 70-81.

Lassila, O. & Swick, R. R. (1999). "Resource Description Framework (RDF): Model and Syntax Specification. Recommendation, World Wide Web Consortium.

Liu, P., Georganas, N. D. & Boulanger, P. (2002). "Designing Real-Time Vision Based Augmented Reality Environments for 3D Collaborative Applications", Proceedings of the IEEE Canadian Conference on Electrical & Computer Engineering 2002 (pp. 25-31), Winnipeg, Manitoba.

McGrath, R.E. & Mickunas, D.E. (2001). "An Object-Oriented Framework for Smart Spaces", submitted to OOPSLA2001.

Okuma, T., Kurata, T., & Sakaue, K. (2002). "VizWear-3D: A Wearable 3-D Annotation System Based on 3-D Object Tracking using a Condensation Algorithm", IEEE Virtual Reality Conference 2002 (pp. 295).

Poupyrev, I., Tan, D. S., Billinghurst, M., Kato, H., Regenbrecht, H. & Tetsutani, N. (2002). "Developing a Generic Augmented Reality Interface", IEEE Computer 35 (3) (pp. 44-50).

Pradhan, S., Brignone, C., Cui, J-H., McReynolds, A. & Smith, M. (2001). "Websigns: Hyperlinking Physical Locations to the Web", IEEE Computer, Vol. 34, No. 8 (pp. 42-48).

Rekimoto, J. & Ayatsuka, Y. (2000). "CyberCode: Designing Augmented Reality Environments with Visual Tags", Proceedings of DARE2000 (pp. 1-10), Elsinore, Denmark.

Roman M. & Campbell, R. H. (2000). GAIA: Enabling Active Spaces. In 9th ACM SIGOPS European Workshop. September 17th-20th, 2000. Kolding, Denmark. http://choices.cs.uiuc.edu/2k/papers/sigopseu2000.pdf

Schwald, B., Seibert, F., & Weller, T. (2002). "A Flexible Tracking Concept Applied to Medical Scenarios Using an AR Window", Poster at ISMAR2002.

Viirre, E., Pryor, H., & Nagata, S. (1998). "The Virtual Retinal Display: A New Technology for Virtual Reality and Augmented Vision in Medicine", In Proceedings of Medicine Meets Virtual Reality, San Diego, California, USA, (pp. 252-257), Amsterdam IOS Press and Ohmsha.

Wand, F., Weiser, M. & Mynatt, E. (1998). "Activating Everyday Objects", Proceedings of the 1998 DARPA/NIST Smart Spaces Wordshop (pp. 7/140-7/143), Gaithersburg, Maryland.