

LUMAR: A Hybrid Spatial Display System for 2D and 3D Handheld Augmented Reality

Alex Olwal

School of Computer Science and Communication
KTH (Royal Institute of Technology)
100 44 Stockholm, Sweden

alx@csc.kth.se

Anders Henrysson

VITA, Linköping University,
ITN, Campus Norrköping
601 74 Norrköping, Sweden

andhe@itn.liu.se

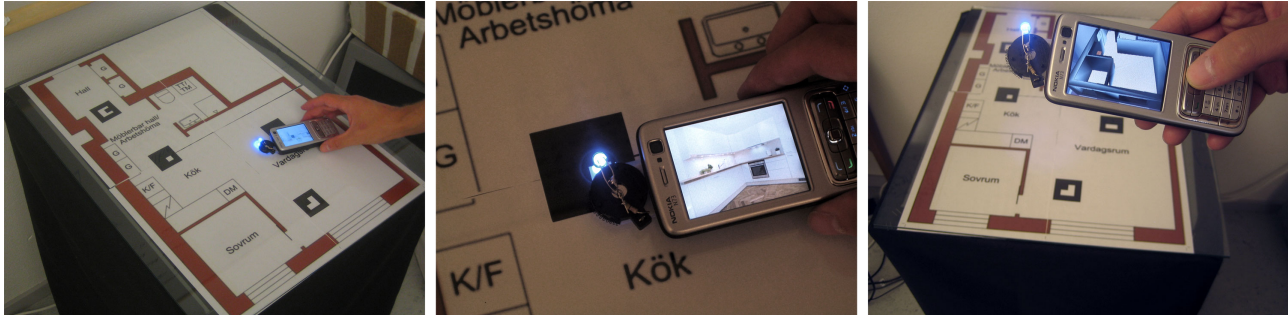


Figure 1. The LUMAR system incorporates exocentric LightSense tracking with the egocentric, ARToolKit-based, UMAR framework, resulting in a hybrid system for tracking a handheld spatial display in 2D and 3D. This allows a commercial off-the-shelf cell phone to display context-sensitive 2D information and 3D augmentations.

ABSTRACT

LUMAR is a hybrid system for spatial displays, allowing cell phones to be tracked in 2D and 3D through combined egocentric and exocentric techniques based on the LightSense and UMAR frameworks. LUMAR differs from most other spatial display systems based on mobile phones with its three-layered information space. The hybrid spatial display system consists of printed matter that is augmented with context-sensitive, dynamic 2D media when the device is on the surface, and with overlaid 3D visualizations when it is held in mid-air.

CR Categories: H.5.1 [Multimedia Information Systems]: Artificial, augmented, and virtual realities; H.5.2. [User Interfaces]: Graphical user interfaces, Input devices and strategies; I.3.6 [Methodology and Techniques]: Interaction techniques.

Keywords: Spatially aware, portable, mobile, handheld, cell, phone, augmented reality, mixed reality, ubiquitous.

1 INTRODUCTION

We are moving towards a world where we, through our portable devices, can be connected everywhere, with instant access to a wealth of digital content. Digital content is becoming truly ubiquitous and is increasingly taken for granted.

Simultaneously, laptops are no longer exclusive instruments for businessmen and engineers, but are becoming a tool for creativity, socializing and communication – a

necessity for everyone. The form factor and means of interaction however tend to make them socially inconvenient to use at any given time in public spaces. Instead we rely on handheld computers, such as PDAs and cell phones, which have gained an enormous popularity in recent years.

The advancement in technology has made it possible for today's phones to become powerful Internet-connected handheld computers with integrated calling, messaging, PDA and camera functionality. The focus of connected portable devices has shifted from laptops to PDAs to cell phones. Their position as the preferred portable device is continuously being strengthened and makes it increasingly harder to motivate the user to bring other devices with them, such as PDAs, cameras, voice recorders and music players. Still, user interfaces are constrained and the phone's interaction with the outside world and the surroundings is limited.

There are ways of accessing digital information associated with physical objects, for instance using bar code or RFID readers that use the obtained ID of a tag for a lookup in an online resource. A more sophisticated approach using computer vision was presented by Föckler et al. in the PhoneGuide project [5], where a captured image of an object is recognized on the device such that associated media can be presented to the user. But this query-type interaction does not allow seamless integration of our handheld device with the physical environment, and we are thus typically forced to resort to conventional cumbersome and indirect

approaches. Such approaches where a scan triggers an event are significantly different from *spatial displays*, where the handheld device is continuously tracked in the 3D space with multiple degrees of freedom.

We discuss related work in Section 2 and present our hybrid tracking system, its components and our implementation in Section 3. We discuss three-layered information spaces in Section 4, followed by our experimental setup and application in Section 5. We describe various application scenarios that our techniques enable in Section 6. Finally, Conclusions and Future Work is provided in Section 7.

2 RELATED WORK

The concept of spatially aware displays was introduced by Fitzmaurice et al. [3][4] and Rekimoto [10] with their Chameleon and NaviCam systems. These proof-of-concept implementations illustrate interesting interaction metaphors made possible with a tracked handheld display that explores a virtual information space, grounded in a physical environment.

A small tracked display could for instance show a small section of a larger map. As the display is moved, its relative position will determine what part of the larger map is visible. The display thus becomes a viewport into the larger virtual environment.

This metaphor becomes significantly more interesting when the virtual information space is related to physical information sources in the environment. Fitzmaurice for instance shows how a tracked display held above a physical map can give the user access to various digital map layers.

Rekimoto's NaviCam system utilized a display that is equipped with a camera on its back. It uses computer vision to detect special markers in the environment and uses them to identify objects, as well as tracking its position and orientation in the environment. The displayed camera image can thus be overlaid with relevant annotations that communicate virtual information.

These two pioneering systems utilize two distinct approaches for tracking the display and they differ not only in technology, but also conceptually.

As it has recently become possible to, on commercial handheld computers and cell phones, realize the proof-of-concept metaphors introduced by the Chameleon and NaviCam an even clearer distinction between the two tracking approaches can be seen.

2.1. Egocentric tracking

NaviCam attached a camera onboard the handheld system, which resulted in an egocentric, inside-out tracking of markers in the environment. While in its original design it relies on special markers in the environment, one can

imagine those algorithms being replaced with natural feature tracking.

Recently, several researcher groups have realized the benefit of using the onboard camera on off-the-shelf camera cell phones. Its use has been investigated in numerous projects for egocentric tracking of markers and features in the physical environment.

Custom 3D markers were used by Moehring et al. [7] in a system developed for the Symbian platform. A cell phone could track its relative position in relation to small 3D coordinate systems that were cut from paper.

The Ubiquitous Mobile Augmented Reality (UMAR) framework [6] makes the popular video-see through AR framework, ARToolKit¹, available on Symbian phones, while the Studierstube group ported ARToolKitPlus² for use on Windows Mobile devices [11]. ARToolKit recovers the relative transform from the camera to 2D markers in the environments and allows the location of the 2D markers to be overlaid with 3D graphics.

The advantages of egocentric tracking based on ARToolKit-type approaches are:

- 1) The device tracks itself such that no significant infrastructure is required aside from the 2D markers. There is thus much more flexibility in how and where the device can be used.
- 2) Six degrees of freedom (3DOF position and 3DOF orientation) is recovered and enables rich interaction possibilities. Thus the capabilities of spatial displays can be fully utilized.

The disadvantages include:

- 1) Computer vision algorithms are resource and computationally intensive, which reduces the performance and leaves less CPU cycles for the application. This is currently especially critical for a cell phone.
- 2) The device can not be too close or too far away from the surface, as the camera needs to see the markers in order for the tracking to work. If the device is placed on the surface, then the camera will be blocked. If the device is too far away from the surface, the reduced size and resolution of the markers in the acquired camera image will make it increasingly harder to track them. These two limitations give rise to an interaction style where the device must be held in mid-air, which might be inappropriate for extended time of use, due to possible risk of fatigue.

¹ <http://www.hitl.washington.edu/artoolkit/>

² http://studierstube.icg.tu-graz.ac.at/handheld_ar/artoolkitplus.php

- 3) Cluttered environment. While natural feature tracking could be possible in the future, current egocentric techniques depend on specific markers to be able to recover the device's absolute position in relation to the marker. The markers can be visually distracting and take up space. It is thus important that they are carefully integrated with the environment.

2.2. Exocentric tracking

The Chameleon system can be viewed as an exocentric, outside-in tracking system, where the display is tracked in an electromagnetic field through a sensor attached to the display. The position and orientation of the display is recovered through the use of a heavily instrumented environment.

Recent projects use exocentric approaches where external sensors detect and track the devices in an instrumented environment, and the recovered location is wirelessly communicated to the device.

Bluetable [12] is a computer vision system that can detect a cell phone on an interactive surface through camera tracking in the IR domain. When phone-like devices are detected on the surface, the system sequentially connects to the different Bluetooth devices in the environment, in an attempt to find a service that runs on the visually detected phone. For each found candidate, the system instructs the device to identify itself through optical communication, either by blinking on the IR port or by blinking the light on the display. The system can thus initiate a handshaking procedure and identify the location of devices on the surface using a combination of communication through Bluetooth and vision techniques.

LightSense [8] uses outside-in tracking to locate LED lights on cell phones. The ultra-bright LED, which normally serves as a photo light to assist taking pictures in dim lighting conditions, is here used as an easily identifiable active marker that can be detected by an external camera or photosensor. LightSense demonstrates how a combination of a computer/camera or microcontroller/photosensors can be used to locate the device on the surface. The device's distance from the surface can be estimated through the size of the detected light spot. The recovered location is continuously sent to the phone over Bluetooth.

The advantages of such exocentric techniques are:

- 1) All computation of device position takes place in the instrumented environment. The phone only needs to deal with a low-bandwidth wireless communication for receiving its coordinates.
- 2) Tracking can be done at better resolution and at higher frame rates given virtually unlimited computation and sensing resources. Multiple cameras or sensors can be used for increased performance, since there are in the-

ory no restrictions on size and weight compared to on-board approaches.

- 3) The device can be conveniently placed and interacted with directly on the surface, which minimizes the risk for fatigue.
- 4) Intrusive markers can be avoided, which leaves the physical environment intact and hides the technology.

The disadvantages:

- 1) The system relies on a heavily instrumented environment with its own computational resources.
- 2) It is significantly harder to recover 6 degrees of freedom tracking for the device, since a precise measure of its orientation is hard to establish. It is also more difficult to determine its distance from the surface. However, 2D tracking on the surface works very well, which includes 2D position and potentially also 1D orientation in the plane.

2.3. Egocentric + Exocentric = Hybrid

As we have discussed above, the two tracking approaches have their respective advantages and disadvantages. In this paper we introduce a hybrid tracking system for mobile devices, where we get the best of both worlds. The device's camera is used for egocentric 3D tracking above the surface. When the device is placed on the surface, and the ability to use the camera is disabled, the device is tracked in 2D by an exocentric system. (See Figure 1)

3 HYBRID TRACKING

Our exocentric and egocentric components are based on LightSense [8], and the ARToolKit-based UMAR framework [6], resulting in the hybrid LUMAR (LightSense + UMAR) system for spatial displays.

The setup consists of an interactive table with printed material that has integrated ARToolKit markers. Under the paper map, the LightSense camera monitors the surface for the presence of phone LEDs.

3.1. LightSense

LightSense is responsible for 2DOF position tracking on the surface. It consists of two servers running on PCs and a LightSense client on the cell phone, as shown in Figure 2. The LightSense Image Processing component detects the presence of a bright LED in the camera image. A chain of image processing filters finds the brightest spots in the image, filters out noise, and finally uses ellipse-fitting to establish the shape of the light from the LED. The center of the ellipse corresponds to the location of the phone in the camera image. The Image Processing component is implemented in C++ with a custom modular filter frame-

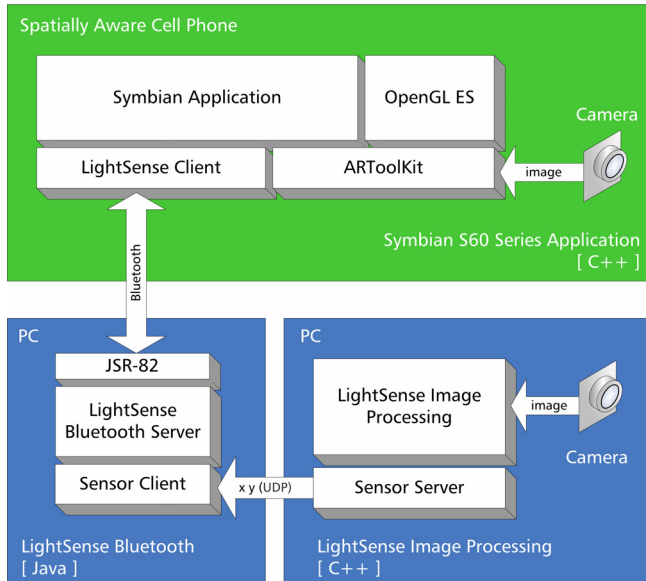


Figure 2. The LUMAR system combines egocentric and exocentric camera tracking from the UMAR and LightSense systems. While most of the work in LightSense is done with an external camera and processing on a separate PC, the UMAR tracking takes place on the device using a port of ARToolKit and OpenGL ES. This separation allows the two tracking approaches to seamlessly coexist in the system.

work on top of the Open Source Computer Vision library, OpenCV³.

The coordinates are sent over UDP to the LightSense Bluetooth Server, a Java application that uses JSR-82, the Java Bluetooth standard, to communicate with the phone.

If the cell phone does not have a built-in LED or the camera API does not provide sufficient control, an external LED can be attached, as in the case of the Nokia N73 phone used in our prototype setup.

3.2. UMAR

The UMAR system provides full 6DOF position and orientation tracking above the surface. Lacking any dedicated position and orientation sensors, the phone has to use its camera for egocentric tracking. For this purpose, Henrysson ported ARToolKit to the Symbian platform [6]. The performance was improved and reached interactive frame rates when the camera pose estimation algorithm was rewritten using fixed point representation. This is an essential step since most phones still lack floating point units. Floating point arithmetic is up to two order of magnitude slower when emulated in software. Further enhancements include frame-to-frame coherency thresholding to avoid expensive tracking calculations and DESP smoothing between pose estimations to stabilize tracking.

From identified markers, ARToolKit uses corner and edge information to calculate the camera pose matrix relative to a coordinate system centered in the marker. The resulting

matrix is used in the OpenGL ES graphics pipeline to align the physical and virtual cameras.

3.3. Combining LightSense and UMAR

Since LightSense is an exocentric tracking system, only data relayed by the Bluetooth server need to be received by the phone. The phone requests data by sending a message. The server sends the current vector and listens for the next request.

In our hybrid tracking setup the mobile phone camera is always on and 3D mode is entered when a marker is identified. This strategy means that there is little or no risk that the phone would accidentally enter 3D mode while remaining on the table surface. ARToolKit separates marker detection and pose estimation where the former has a low computational cost. The tracking performance in 2D is therefore not affected.

When no marker is visible, the system goes back into 2D mode. Since LightSense provides a crude z-parameter the phone could remain in 3D mode if at a distance from the surface. This would be useful since markers might be lost and camera feedback is useful to guide the user until a marker is framed.

With LightSense being 3DOF and ARToolKit 6DOF there will be a seam in tracking if the phone is rotated when being lifted from the surface before any marker is detected. If there is a one-to-one mapping between print and 2D graphics, the phone must already be aligned with the coordinate axis for the magic lens metaphor to hold since no rotation is currently detected.

3.4. Interaction modes

The hybrid tracking allows fluid transition between 2D and 3D modes of interaction.

In the 2D mode, we use the phone as a tangible cursor in the physical information space that the map represents. It is convenient and intuitive to put down the phone on top of objects of interest for immediate feedback with context-sensitive information in the display.

Lifting the device, we move into 3D mode where the tracking transitions into full isomorphic 6DOF interaction — i.e. direct, motion-based 3D input. The information in the space can now be properly visualized with 3D graphics and overlaid on the live video feed from the camera.

The interaction modes provide fluid single-handed operation both as the device is moved within each mode as well as when it transitions between 2D and 3D modes. Buttons and joypads on the device complement the interactive aspects of each mode.

³ <http://sourceforge.net/projects/opencvlibrary/>

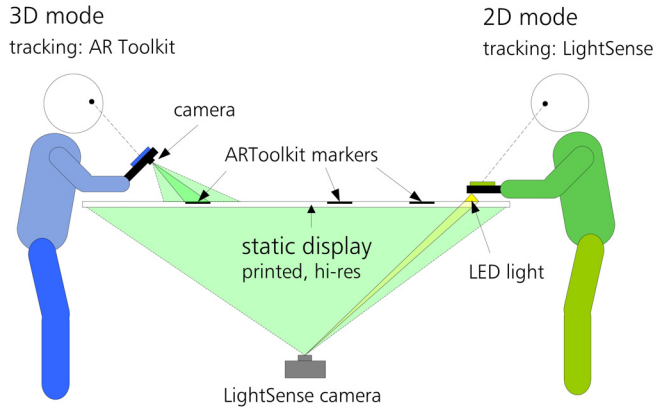


Figure 3. The LUMAR system enables fluid transitions between 2D and 3D modes. It combines the individual strengths of the respective tracking and visualization modes.

4 THREE-LAYERED INFORMATION SPACES

In visualization there is a trade-off between providing high amount of detailed information and facilitating fast navigation. In dynamic systems the user is often provided with a coarse overview of the data and means to zoom in on areas of interest. Data is often segmented into layers that can be hidden when other types of information is requested. In, e.g., Google Earth, the user is presented with a view from space from which it is possible to zoom in on any part of the globe. On the surface level, 3D terrain and buildings are added and information layers can be enabled to provide road information or photographs.

In static systems, such as printed information boards, this interactive level of detail is not possible. Static levels can be provided on separate surfaces, but this requires a switch of context with associated seams and cognitive load. In many public scenarios there is no room for more than one level of detail. In, e.g., public street maps, a high level of detail is often provided together with a legend saying: “You are here”. This is a common strategy since most pedestrians want to know their current position. In contrast, public transportation maps provide a coarse overview of routes since focus is on finding the right train as quickly as possible, given a known destination. Despite rapid deployment of displays ranging in size from handheld devices to wall-mounted TV-screens, printed information will remain a dominant media given its low cost, and ease of handling and distribution.

3D information is essential for real-world navigation and conceptualization of objects. The use of the third spatial dimension further exploits human cognitive abilities to interpret visual information. Printed media used for maps and advertisement lack this capability, while stationary 2D displays require dedicated 3D interaction devices or various 2DOF modes to set the view parameters. A dedicated 3D system is however not a good option since much multimedia is inherently 2D. Browsing the web or watching a video in a 3D environment makes little sense and often

degrades quality or introduces fatigue when one dimension must remain fixed.

Some previous systems combine different media in a similar fashion. In MagicBook [2], 3D content is added to book pages. It uses both AR and VR. The user can view objects popping up from the pages in AR mode. Alternatively, the user can become immersed in the 3D world by entering the VR mode. Additionally, there exist simpler approaches where printed markers are overlaid with 2D graphics as in the original UMAR paper [6]. The Marked-Up Maps project [9] demonstrates how an RFID-sensing PDA can display electronic information related to a region of interest in a map with embedded RFID tags. The device is tracked in 2D using a grid of tags under the paper surface.

In this work we combine the respective strengths of printed media, 2D multimedia and 3D interactive graphics, as shown in Figure 3. The key observation is that a mobile phone provides a relatively high-resolution display for 2D and 3D content and enables intuitive motion-based 6DOF interaction when tracked. Modern mobile phones are able to display most media types, including videos and web pages. They provide mobile Internet access and an assortment of short range communication techniques. Instead of placing expensive and sensitive displays in public areas, we exploit the fact that most people carry a small interactive display with them at all times. To compensate for the small size of the mobile phone screen, we use the Magic lens metaphor [1], enabled by our exocentric tracking infrastructure. Our 6DOF egocentric tracking further enhances this setup by allowing interactive motion-based 3D navigation and interaction. It is important for the interaction techniques to blend as seamlessly as possible when transitioning between 2D and 3D mode. With our hybrid tracking strategy, we achieve this in form of motion based interaction. The choice of AR over VR for 3D visualization allows printed information to be combined with 3D graphics.

The general application scenario consists of a high resolution print that gives an easily navigated overview and ensures that the tangibility of the media and its possible use as a traditional display is preserved. The high-resolution print can be important in several applications, where the superior level of detail is vital, yet impossible to reproduce with a digital display. Subsequent 2D information layers are digital and viewed by placing the phone on the surface of the print. They can be either continuous, providing higher information density than the print for the same region of interest, or they can be segmented into areas, each associated with various kinds of 2D multimedia, such as videos or web pages. When the phone is lifted, a transition occurs into 3D mode, where 3D information serves as an augmentation of the print or as a continu-



Figure 4. The Nokia N73 phone with an attached external LED. Although the N73 has a built-in LED, it is not controllable by software, and we thus attach an external LED for tracking by the LightSense system.

ous 2D layer now registered in 3D. Thus there is a close spatial relationship between 2D and 3D information. The phone acts as a looking-glass that magnifies the information content.

Markers are often considered obscuring, but with three tiers (print, 2D and 3D) the print can have low information density and markers will be less intrusive than if no 2D digital layer was used.

5 INTERACTIVE APARTMENT EXPLORATION

We decided to implement an apartment ad scenario as a first prototype application for the LUMAR system.

Floor plans are typically used to communicate the layout of rooms and their relative size in apartment ads. They are however of little use when a potential buyer wants to visualize what it is like to be inside the apartment or navigate the information associated with the different rooms, for instance.

From an Internet real estate site we obtained an apartment plan and images of the interior. An enlarged print of the plan was placed on the LightSense table and four AR-ToolKit markers were placed in empty spaces of the print, ensuring that they were not obscuring any details.

We associated one photograph with each room. This was easily achieved by simply moving the phone over the plan and recording the coordinates at room corners. The procedure is analog to producing an HTML image map.

Using the floor plan and images, we made a 3D model of the apartment in the 3D modeling application Google SketchUp⁴. Each marker's offset relative to the model's origin was measured manually.

In this case no mapping between LightSense and AR-ToolKit coordinates was necessary. However, if the 2D information had been continuous and the 3D information superimposed on it we could have easily aligned the two coordinate systems and calculated the scale factor between pixels used by LightSense and millimeters used by ARToolKit.

The mobile phone was a Nokia N73. It features a 220 MHz CPU and a 3.2 MP camera with Carl Zeiss optics. For tracking and rendering we use a video resolution of 320×240 pixels. Since the camera API did not give us sufficient control of the built-in flash, we attached an external LED for 2D tracking (See Figure 4.) We ported UMAR to Symbian 9.1 and wrote a COLLADA importer for 3D scene loading.

Placing the phone over a room on the surface will load and display an image of the room. When the phone is moved across the surface, corresponding images are loaded automatically, as shown in Figure 5.

When the phone is lifted and a marker is framed, a 3D model of the apartment is superimposed on the printed plan and registered with it in 3D. (See Figure 6.) The user can now move the phone freely to view the 3D model from an arbitrary direction provided that at least one marker is fully visible. The registration makes it easy to identify which object in the 3D model corresponds to a certain legend in the plan.

Putting the phone back on the surface makes the application switch to 2D mode, as soon as no marker is

⁴ <http://www.google.com/sketchup/>

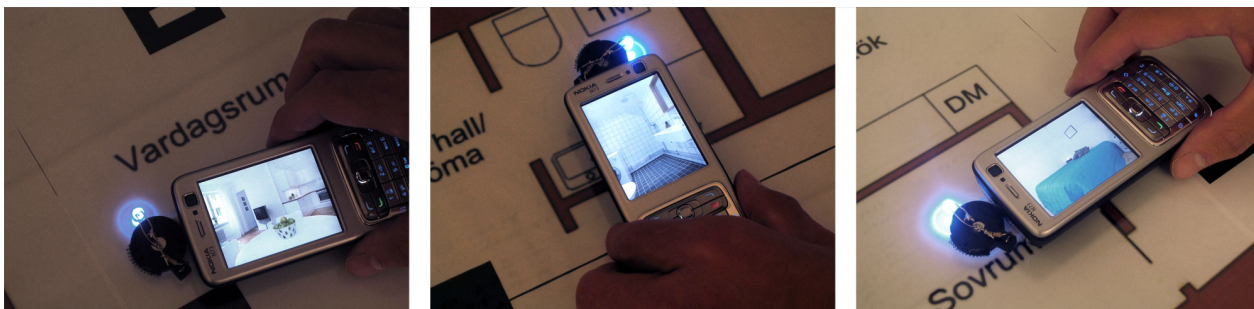


Figure 5. 2D mode. Context-sensitive information, in the form of images from the various rooms, is shown to the user as the phone is held over the corresponding location on the floor plan.

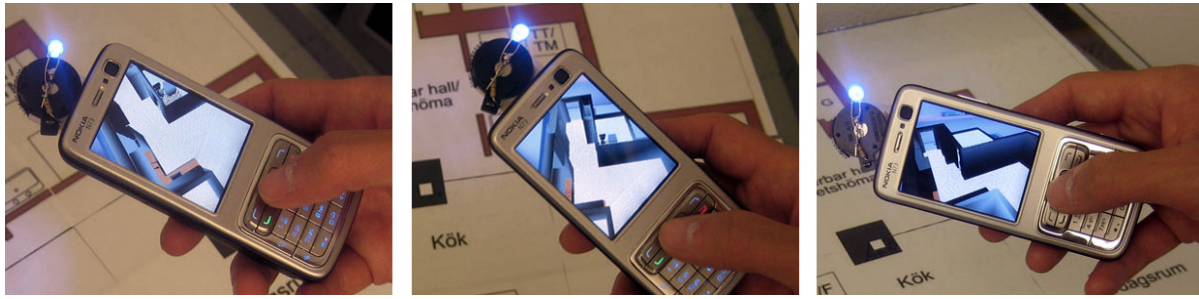


Figure 6. 3D mode. A 3D model of the apartment is overlaid over the live video feed, allowing the user to view the apartment from different angles and get a good feeling for its structural properties.

visible. The setup works as intended and results in a smooth and fluid transition between 2D and 3D by simply lifting the phone from the table.

Currently we do not utilize z-values from LightSense but they could be used to detect if the user intends to switch back to 2D or if the markers were accidentally lost. If they were lost, viewfinder feedback should be displayed to make it easier to find and frame a marker.

6 APPLICATION SCENARIOS

There are many application scenarios that commonly make use of printed material, where 2D and 3D digital information could add significant and relevant information, as well as enabling the possibilities for interesting services. We described our implementation of one such scenario in our previous discussion of the interactive apartment exploration. Additional example application scenarios and the information as it relates to the printed matter, the 2D dynamic display and 3D visualizations, are outlined in this Section.

6.1. Super map

Maps are often designed with fast navigation in mind. Thus the photometric resolution is very low with one color for each type of terrain. To make roads salient blocks of buildings are typically reduced to single objects.

In our system, a phone at a desired location could display up-to-date aerial image data where moving the phone across the map surface would control the panning. Most services associated with digital maps, such as route planning, could be supported.

Lifting the phone from the surface would transform the aerial image into a texture on a 3D terrain model registered with the physical map. Detailed 3D buildings with textured facades would replace the flat projections.

6.2. Information visualization

Information visualization is a field where we often like to combine high-density displays with dynamic data. Here, the print serves as an ideal grounding information space, and where the lower resolution 2D and 3D visualization

can provide focus and detail with dynamically updated content.

Stock prices, for instance, can be plotted over time in a chart to give a coarse overview of their development. By moving the phone along the chart, additional data such as market index or comparable stocks can be displayed. In 3D, the plotted curve can be extruded along the z-axis with the height value corresponding to trade volume.

6.3. Arena seat booking

Arena maps and folders are often used to help users select, book and find their seats. Such tangible matter is invaluable and cannot be readily replaced by solely digital information. It could however be advantageous to combine the physical and virtual.

For example, to find a seat at an arena, the user could move the phone over a printed section map to identify individual seats, their availability, price and number.

Moving into 3D, the user could view seats of interest highlighted in a 3D-model of the stadium. Similarly to MagicBook [2], the user could fly into the model and get a feel for the view from a particular seat to ensure visibility.

Returning to 2D mode, the user could book a seat through a web service by a simple point and click action.

6.4. Public transportation

Bus route maps often mark only a selection of the stops along the route. By moving the phone along the route all individual stops can be shown. Other potential data is the estimated time of arrival given the current traffic situation and the local geographical neighborhood map at each location. Live video feeds from traffic cameras could also be displayed at relevant locations.

A wide area 3D overview could provide current bus locations from GPS and ideal locations from time

table data, to visualize delays. Color or size could be used to indicate, e.g., the current speed of a bus.

6.5. Product catalog

Printed ads are all around us trying to get our attention. They however often lack detailed product information, and to consume the product the buyer must remember the brand and model name. The Internet contains a rich set of data for various products, such as price comparisons, reviews, images, 3D models and videos. These information sources are however inaccessible from the printed media.

By moving the phone over a product list, e.g., a car catalogue or restaurant menu, detailed information can be shown and on-line price comparison services can be consulted with just a click away. The product can also be visualized in 3D and extraneous images and videos can be shown to the user.

Products with changing prices can also be dynamically updated and immediately displayed to the user.

7 CONCLUSION AND FUTURE WORK

We have demonstrated a hybrid tracking system that combines printed media with digital 2D multimedia and 3D graphics, using handheld AR. This is important since all three media types have their individual strengths and together contribute to empower our human perception. The single-handed interaction is based on motion in both 2D and 3D modes which minimizes the interaction seams.

In the present setup, all information resides on the phone at start. In a real-world scenario, both 2D and 3D data could be downloaded to the phone on the fly, which would require service discovery and data transfer. Since we know beforehand what the print looks like we could provide feature data in addition to 2D and 3D information. That could allow us to replace or extend the ARToolKit tracking with natural feature tracking, resulting in potentially even smoother transitions between 2D and 3D modes.

As described previously [8], the LightSense system could also be made flatter with a matrix of photosensors under the surface, or using camera arrays and wide angle lenses.

After our implementation and exploration of three-layered information spaces, we find it interesting to in future work explore other possible application scenarios, as well as their associated interaction techniques.

ACKNOWLEDGEMENTS

The second author is funded by Brains and Bricks, at the University of Linköping as well as receiving supervisory support from CUGS, the Swedish National Graduate School in Computer Science.

REFERENCES

- [1] Bier, E. A., Stone, M. C., Pier, K., Buxton, W., and DeRose, T. D. 1993. Toolglass and magic lenses: the see-through interface. In *Proceedings of the 20th Annual Conference on Computer Graphics and interactive Techniques SIGGRAPH '93*, pp. 73–80, 1993.
- [2] Billinghamhurst, M., Kato, H., and Poupyrev, I. The MagicBook—Moving Seamlessly between Reality and Virtuality. *IEEE Comput. Graph. Appl.* 21, 3 (May. 2001), pp. 6–8, 2001.
- [3] Fitzmaurice, G. W., Zhai, S., and M. Chignell. Virtual Reality for Palmtop Computers. *ACM TOIS* 1993, pp. 197–218, 1993.
- [4] Fitzmaurice, G. W. Situated information spaces and spatially aware palmtop computers. *Commun. ACM* 36, 7 (Jul. 1993), pp. 39–49, 1993.
- [5] Föckler, P., Zeidler, T., Brombach, B., Bruns, E., and Bimber, O. PhoneGuide: Museum Guidance Supported by On-Device Object Recognition on Mobile Phones. In *Proceedings of International Conference on Mobile and Ubiquitous Computing (MUM'05)*, pp. 3–10, 2005.
- [6] Henrysson, A. and Ollila, M. UMAR: Ubiquitous Mobile Augmented Reality. In *Proceedings of the Mobile and Ubiquitous Multimedia (MUM '04)*, pp. 41–45, 2004.
- [7] Moehring, M., Lessig, C. and Bimber, O. Video See-Through AR on Consumer Cell Phones. In *Proceedings of International Symposium on Augmented and Mixed Reality (ISMAR'04)*, pp. 252–253, 2004.
- [8] Olwal, A. LightSense: Enabling Spatially Aware Handheld Interaction Devices. In *Proceedings of International Symposium on Mixed and Augmented Reality (ISMAR 2006)*, pp. 119–122, 2006.
- [9] Reilly, D., Welsman-Dinelle, M., Bate, C., and Inkpen, K. Just point and click?: using handhelds to interact with paper maps. In *Proceedings of Human Computer interaction with Mobile Devices and Services (MobileHCI '05)*, pp. 239–242, 2005.
- [10] Rekimoto, J. and Nagao, K. The world through the computer: computer augmented interaction with real world environments. In *Proceedings of User interface and Software Technology (UIST '95)*, pp. 29–36, 1995.
- [11] Wagner, D. and Schmalstieg, D. First Steps Towards Handheld Augmented Reality. In *Proceedings of the 7th IEEE international Symposium on Wearable Computers (ISWC 2003)*, pp. 127–135, 2003.
- [12] Wilson, A. and Sarin, R. Connecting Wireless Mobile Devices on Interactive Surfaces Using Vision-Based Handshaking. In *Proceedings of Graphics Interface (GI 2007)*, pp. 119–225, 2007.