Research Directions in Handheld AR

Mark N. Billinghurst and Anders Henrysson

Abstract—**Handheld mobile devices are an exciting new platform for Augmented Reality (AR). Mobile phones and PDAs have the potential to provide AR experiences to hundreds of millions of consumers. However, before widespread use can occur there are some obstacles that must be overcome. In particular, developers must consider the hardware and software capabilities of mobile devices and how these can be used to provide an effective AR experience. They must also develop AR interaction metaphors suitable for handheld AR. In this paper we review current and previous research in the field, provide design guidelines and outline future research directions.**

Index Terms — **Augmented Reality, Handheld, Cell phone**

I. INTRODUCTION

Over forty years ago Ivan Sutherland put on the first head mounted display and saw a virtual wireframe cube superimposed over the real world in front of him [1]. The technology that blends real and virtual imagery became known as Augmented Reality (AR) [2] and in the forty years since here has been a large amount of research and development to explore applications in fields as diverse as engineering, education and entertainment, among others. During that time computing and communications devices have changed. The processing power that once filled a whole room can fit into the palm of the hand and be connected to machines all over the earth.

These technology changes have also changed the hardware available for Augmented Reality applications. Early AR systems were based on desktop computers with custom input and output devices, but by the mid 1990's the first wearable AR systems appeared based on laptop computers and commodity hardware. In the last few years handheld AR systems based on consumer devices have been developed, using handheld personal digital assistants (PDAs) or mobile phones.

The recent progression of AR technology to mobile phones is significant because for the first time there is the opportunity to provide an AR experience on a widely used mobile hardware platform. The basic hardware requirements for AR applications are a computer processor, display and camera; three components that are in the majority of mobile phones sold. In addition, phones have communication capability that naturally supports collaboration.

However, the interface requirements for a handheld AR application are significantly different from desktop systems, and there are a number of design issues that must be addressed. In this paper we review the development of handheld AR systems, particularly focusing on mobile phone based AR. We present some of the work that we have done in the field and provide design guidelines and promising directions for future work. The goal is to provide a guide for current and future researchers in the field.

II. RELATED WORK

Current handheld AR applications are based on more than a decade of research into portable and wearable AR systems. Fig. 1 shows the evolution to mobile phone based AR systems. From Steve Mann and Thad Starner's experiments in the early days of wearable computing [3], it was obvious that body worn computers could be used for mobile AR. For example, Feiner's Touring Machine [4] combined a backpack computer with GPS/inertial tracking systems to overlay context sensitive virtual information in an outdoor setting. Since this first effort, other wearable AR systems was been developed for outdoor gaming [5], education [6], navigation [7], and even enhancing collaboration [8]. From these research systems it was clear that what was carried in a backpack would one day be held in the palm of the hand.
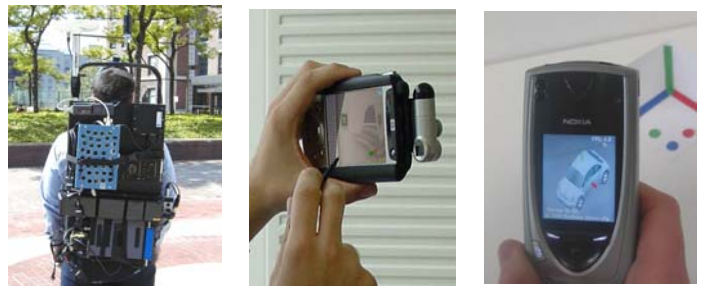

Fig. 1. Evolution of mobile AR from wearable to handheld devices

While some researchers were working on wearable computers, others were exploring the future of handheld AR systems. The first of these simply involved handheld displays that were tethered to computers which did the image processing and display generation. Fitzmaurice's Chameleon [9] allowed a user to see virtual annotations on the real world using a magnetically tracked LCD display, while Rekimoto's Navicam [10] achieved the same using computer vision technology. Rekimoto [11] also showed how a tethered handheld display could provide shared virtual object viewing in an AR setting and enhance face to face collaboration. This was extended with

the AR Pad work of Mogilev [12] who used virtual images to provide cues about shared viewpoints in a face to face setting.

These early handheld applications used displays and input devices that were tethered to a high end computer. As significant computing and graphics power became available on self contained PDA's researchers began exploring their use for AR applications as well. First there was work such as the AR-PDA project [13] and BatPortal [14] in which the PDA was used as a thin client for showing AR content generated from a remote server. This was necessary as the early PDA's did not have enough processing power for stand-alone AR applications.

In 2003 Wagner ported ARToolKit[15] to the PocketPC operating system and developed the first AR application that ran entirely on a PDA [16]. Since that time, he also developed the first stand alone collaborative AR application for the PDA [17]. Unlike the backpack systems, handheld collaborative AR interfaces are unencumbering and ideal for lightweight social interactions.

Mobile phone based AR has followed a similar development path. Early phones also explored thin client approaches. For example, the AR-Phone project [18] used Bluetooth to send phone camera images to a remote sever for processing and graphics overlay, taking several seconds per image. However, Henrysson ported ARToolKit [15] over to the Symbian phone platform [19], and Moehring developed an alternative custom computer vision and tracking library [20]. This work enables simple AR applications to be built, running at 7-14 frames per second. This performance will dramatically improve over the next few years as mobile phones begin to have dedicated graphics hardware.

As the mobile AR hardware platform evolved, so did the AR interaction techniques. The first wearable AR systems used head mounted displays to show virtual graphics overlaid on the real world and developed a number of very innovative techniques for interacting with the virtual data. For example, in the Tinmith system [21] touch sensitive gloves were used to select menu options and move virtual objects in the real world. Kurata's handmouse system [22] allowed people to use natural gesture input in a wearable AR interface, while Reitmayr's work implemented a stylus based interaction method [8].

Handheld AR applications do not typically use head mounted displays, and instead use LCD displays or physical screens. At least one of the user's hands is needed to hold the device so some of the earlier two-handed interaction techniques are not suitable. It is natural in this setting to use stylus input but there are other possibilities as well. In the AR-PAD project [12], buttons and a trackball on the display were used as input. Träskbäck used a tablet-PC and pen input for an AR-based refinery education tool [23], and markers in the environment were used as cues to load the correct virtual content. In Wagner's indoor navigation tool [16] user input is a combination of stylus interaction and knowledge of display position from visual tracking of markers in the environment.

If the AR display is handheld the orientation and position of the display itself can also be used as an important interaction cues. PDA applications such as the Invisible Train [17] interact with the AR content by interacting in the world and with the device itself. In this case, the user moved around in the real world to change the view of the virtual train set and then touched the screen with a stylus to change the position of the tracks. Similarly, in Wagner's AR-Kanji collaborative game [24] the user looked through the PDA screen to view real cards which have Kanji symbols printed on them. When the cards were seen through the screen, virtual models were shown representing the translation of the Kanji characters. These could be manipulated by hand and the model shown from different viewpoints. There was very little stylus input required.

Less research has been conducted on interaction techniques for mobile phone AR interfaces. The mobile phone form factor is similar to that of a PDA, but most phones also have integrated keypads and cameras and provide support for vibration feedback. Phones are also designed for one-handed use. These features allow developers to create interaction techniques ideally suited to mobile phone use. For example, Reimann [25] used the camera input to create a football game in which players could kick a virtual ball with their real foot, trying to score goals against a goal keeper.

This collection of related research shows that with mobile AR systems intuitive interfaces can be developed by considering the physical input and display affordances of the system. We can draw on this research to explore new interaction techniques for handheld AR applications. In the next section we describe our approach to AR interaction design for mobile devices, and in particular an interaction metaphor suitable for AR interfaces on mobile phones.

III. HANDHELD AR INTERFACE METAPHOR

The are several key differences between a handheld AR interface compared to a traditional head mounted display (HMD) based AR system. The virtual content is seen on a screen that is handheld rather than headworn, which means that the display affords a much greater peripheral view of the real world that is not enhanced by AR content. In a handheld device the screen and input devices are connected, for example in a mobile phone the keypad is directly below the screen. Finally, wearable AR systems with HMDs are designed for continuous use, showing virtual content overlaid on the real world. In contrast, mobile phones and PDAs are typically used only for short periods of intensive activity.

This means that interface metaphors developed for HMD based systems may not be appropriate for handheld AR systems. For example, many AR applications are developed following a Tangible AR metaphor [26] where physical objects are used to interact seamlessly with virtual content. Implicit in these applications is the assumption that the user has both hands free to manipulate physical input devices, which will not be the case with handheld devices.

For handheld AR we are exploring a *Tangible Input Metaphor* where the motion of the device itself is used to provide input into the AR application. We assume that the mobile device is like a handheld lens giving a view into the AR scene and that a user will be more likely to move the handheld display than change his or her viewpoint relative to the display. In addition, the small form factor of the mobile phone lets us

support an object-based approach. It is therefore possible to use the mobile phone as a tangible input object itself. Input techniques can be based around motion of the phone.

Mobile phone motion is relatively easy to detect by using the integrated camera. In our work we use the Symbian port of ARToolKit [19] which provides camera pose information relative to a square marker. Simple computer vision techniques can also be used to detect two dimensional camera motion from motion flow cues on the input video. Wang et al. [27] provide a good review of motion detection techniques and describe the TinyMotion library that they have developed for detecting tow dimensional phone motion.

Interestingly, moving the phone handheld display to view virtual content may create the illusion that the display is larger than it really is, and cause an increased sense of object presence; the feeling that the virtual object is really part of the real world. A recent paper by Hwang et al. [28] compared the perceived field of view (FOV) of a moving 5 inch handheld display with fixed displays of various sizes. Users felt that the FOV of the moving display was twice that of a fixed display of the same size, and almost as large as a fixed 42 inch plasma screen. They also reported that users felt that the sense of presence created when they were able to move the 5 inch handheld display was greater than that from the fixed 42 inch plasma screen. These results imply that powerful AR experiences could be created from even small handheld displays, such as in mobile phones, as long as the user is able to freely move the display to interact with the AR content.

IV. SAMPLE APPLICATIONS

In the previous section we described a tangible input metaphor where the motion of the handheld device itself was used to provide input in the mobile AR application. In this section we describe several mobile phone based AR applications that show how this may be implemented. Each of these applications have been reported on in more depth in other papers, so this section provides an overview of current work from which general design recommendations can be drawn.

A. AR Viewing

One of the initial mobile phone AR applications we developed was a simple viewer application. When the user looked at a tracking pattern they could see a virtual object that appeared to be fixed onto the real object (Fig. 2). This involved using a highly optimized version of the ARToolKit computer vision library that was developed for the Symbian OS platform and combined with the OpenGL ES graphics library. To achieve interactive frame rates a fixed-point library was created partly in ARM assembler. To be able to import textured models from a 3D modeling package we used the Deep Exploration tool [ref][1] to convert the exported model to C++ code with OpenGL vertex arrays and then wrote a simple program that converted this into OpenGL ES compatible vertex arrays.

The viewer has been run on both the Nokia 6600 and 6630 phones, with a screen resolution of 176x208 pixels and camera

[1] www.righthemisphere.com

resolution of 160x120 pixels. The 6600 has a 104 Mhz ARM processor and the application ran at 3-4 frames per second, while the 6630 has a 210 Mhz ARM processor and achieved 7 frames per second. Fig. 2 shows a 3,000 polygon virtual model of a car overlaid on video of the ARToolKit marker. The model can be viewed from any angle by moving the phone around the marker, or rotating the marker itself.

Fig. 2. Mobile phone based AR viewing application. See Color Plate 26

B. AR Blocks

Viewing AR content on a mobile phone is interesting, but in most applications there is a need to interact with the virtual content. In order to test AR interaction techniques we developed an AR Blocks application [29] in which users could select and move virtual objects on the mobile phone.

In this application, when users look at a set of ARToolKit tracking markers they see an AR view of a virtual ground plane with several virtual blocks on the ground plane (see Fig. 3.). On the screen is shown a virtual crosshair and users can select a block by positioning the crosshair over it and hitting the joypad controller. When the joypad is clicked the block under the crosshair turns white and is selected and remains selected while the button is held down. While selected, the block is fixed in position relative to the phone and can be moved by simply moving the phone. When the joypad is release the block is dropped back into the AR scene.

Fig. 3. Virtual Blocks Application. See Color Plate 27

C. AR Manipulation Evaluation

In order to test the usability of the manipulation techniques described above we conducted a study in which users tried to

position and orient blocks with different mobile interfaces [30]. The subject sits at a table, which has a piece of paper with a number of ARToolKit makers printed on it. When the user looks through the phone display at the markers they will see a virtual ground plane with a virtual block on it and a wireframe image of another block (Fig. 4.).

Fig. 4. A Virtual Block and Target.

The manipulation study was done in two parts, to explore translation and rotation techniques separately. The goal was to select and move or rotate the block until it was inside the target wireframe block (Fig. 5). In the first part of the experiment we tested the following three ways of translating an object:

A: Object fixed to the phone (one handed): the selected virtual object was attached to the phone and could be moved by moving the phone. The user was not allowed to move the paper tracking maker with their free hand.

B: Button and keypad input: the selected object was moved using button and keypad input. When the keys are pressed the block is translated a fixed amount.

C: Object fixed to the phone (bimanual): as in condition A, the object is attached to the phone, but the user could move both the phone and paper tracking target.

In the second part of the experiment we tested the following techniques for rotating objects:

A: Arcball: the familiar arcball rotation technique

B: Keypad input for rotation about the object axis: the selected object was rotated using button and keypad input.

C: Object fixed to the phone (one handed): the object is attached to the phone and the real phone rotation was used to rotate the virtual object, but the user was not allowed to move the paper tracking maker with their free hand.

D: Object fixed to the phone (bimanual): as in condition B, the object is attached to the phone, but the user could move both the phone and paper tracking target.

When the block was positioned or rotated correctly inside the target wire-frame it changed color to yellow showing the subject that the trial was over. This was determined by measuring the error in position or orientation and stopping the trial once this error value dropped below a certain threshold.

For each trial we measured the amount of time it took the user to complete the trial and also continuously logged the position or rotation of the block relative to the target. After three trials in

one condition we asked the subject to subjectively rate his or her performance and how easy was it for them to use the manipulation technique. Finally after all the were completed we asked the users to rank all them in order of ease of use.
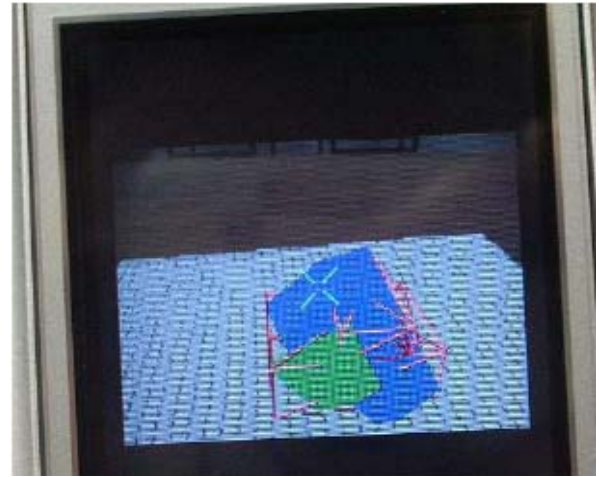

Fig. 5. A Virtual Block and Rotation Target

We recruited a total of 9 subjects for a pilot user study, 7 male and 2 female, aged between 22 and 32 years old. None of the subjects had experience with three-dimensional object manipulation on mobile phones but all of them had used mobile phones before and some of them had played games on their mobile phone. We used a within subjects design so that all the subjects tried all the conditions, although in a counter-balanced order to reduce order effects.

There was a significant difference in the time it took users to position objects depending on the positioning technique they used. Table 1 shows the average time it took the users to position the virtual block in the wireframe target. Using a one factor ANOVA ($F(2,24) = 3.65$, $P < 0.05$) we found a significant difference in task completion times. Users were able to position objects much faster when the virtual object was fixed to the phone.

TABLE I: Average Object Positioning Times

	A: Fixed	B: Keypad	C: Fixed 2
Time (sec)	9.51	17.20	11.95

After each condition subjects were asked to answer the following questions:

Q1: How easy was it for you to position the object?
Q2: How accurately did you place the block?
Q3: How quickly did you think you placed the object?
Q4: How enjoyable was the experience?

Using a scale of 1 to 7 where 1= very easy, 7 = not very easy, etc. Table 2 shows the average results.

As can be seen, the users thought that when the object was fixed to the phone (conditions A and C) it was easier to position the object correctly (Q1) but it they could position the model more accurately (Q2) with the keypad input. A one factor

ANOVA finds a near significant difference in the results for Q1 (F(2,24) = 2.88, P = 0.076) and Q2 (F(2,24) = 3.32, P = 0.053).

TABLE 2: SUBJECTIVE RESULTS

	A: Fixed	B: Keypad	C: Fixed 2
Q1	5.8	5.0	5.8
Q2	5.4	6.0	5.1
Q3	5.9	4.4	5.1
Q4	5.7	4.6	5.3

There is a significant difference in the other conditions. The users thought they could place the objects more quickly when they were attached to the phone (Q3) and the tangible interfaces were more enjoyable (Q4). A one factor ANOVA finds a significant difference in the results for Q3 (F(2,24) = 5.13, P < 0.05) and Q4 (F(2,24) = 3.47, P < 0.05).

There was also a significant difference in the time it took users to orient objects depending on the technique they used. Table 3 shows the average time it took the users to rotate the virtual block to match the wireframe target.

TABLE 3: AVERAGE OBJECT ROTATION TIMES

	A: Arcball	B: Keypad	C: Fixed	D: Fixed 2
Time (sec)	22.5	20.5	42.3	37.0

As can be seen, conditions A (arcball) and B (keypad input) are on average twice as fast as the Tangible Input rotation conditions (C and D). A one-factor ANOVA finds a significant difference between these times (F(3,32) = 4.60, P < 0.01).

Subjects were also asked to answer the same survey questions as in the translation task, except Q1 was changed to:

Q1: How easy was it for you to rotate the virtual object?

Table 5 below shows the average subjective scores for the survey questions. There was no significant difference between these survey responses. The subjects thought that the conditions were equally easy to use and enjoyable.

TABLE 4: SURVEY QUESTION RESPONSES

	A: Arcball	B: Keypad	C: Fixed	D: Fixed 2
Q1	4.7	5.4	4.6	5.1
Q2	4.8	5.2	4.8	5.0
Q3	5.0	4.8	4.1	4.6
Q4	5.1	4.9	4.9	5.6

In addition to survey responses users gave additional comments about the experience. Several commented that when the virtual object was attached to the phone they felt like they were holding it, compared to the case where the keypad was used and they felt that they were looking at a screen. One user said "when the object was attached to the phone, the phone felt more like a tool." They felt like they were more in control and they could use their innate spatial abilities when manipulating the virtual object. In contrast those that preferred the keypad liked how it could be used for precise movements and also how you did not need to physically move themselves to rotate the object about its axis.

The results show that using a tangible interface metaphor provides a fast way to position AR objects in a mobile phone interface. The user just had to move the real phone where the block was to go. The subjects also felt that it was more enjoyable. However, there seems to be little advantage in using a tangible interface metaphor for virtual object rotation. When the virtual object is fixed to the phone then the user often has to move themselves and the phone to rotate the object to the orientation they want, which takes time.

Our results suggest that virtual object positioning based on physical phone motion could be a valuable technique but rotation may be better performed through keypad input about constrained axes.

In the next section we show how the same tangible interaction metaphor can be applied to collaborative AR experiences on a mobile phone.

D. AR Tennis

Mobile AR interfaces can also be used to enhance face to face collaborative experiences. In order to explore this we developed AR Tennis [31], the first face to face collaborative Augmented Reality application developed for mobile phones. In this application two players sit across a table from each other with a piece of paper between them with a set of square markers drawn on it. When the player points the phone camera at the markers they see a virtual tennis court overlaid on live video of the real world (see Fig.6).

Fig. 6. AR Tennis Application

As with the previous applications we have implemented a tangible input metaphor. Players can hit a virtual ball over the net and play tennis against each other. Players serve the virtual ball by hitting the '2' key on the keypad. Once the ball is in play they can hit it back by simply moving their phone in front of it;

there is no need to use the keypad any more. A simple physics engine is used to provide realistic ball motion. Bluetooth wireless networking is used to synchronize the ball movement between phones. Game play is further enhanced with audio and vibration feedback. Each time the ball is hit there is a sound played and the phone that hits the ball vibrates.

The use of an appropriate tangible object metaphor is very important for the usability of mobile phone AR applications. In our case we wanted the player to feel like that the phone was a tennis racquet hitting balls over a virtual net. Once they understood this metaphor it was easy for users to move the phone around the court space to hit the ball. Physical manipulation of a phone is very natural and so provides an intuitive interaction approach for collaborative AR games. This also meant that apart from hitting a key to start serving there was no need to use keypad input while playing the game.

A formal user study was conducted to explore how useful the AR view of the game was, especially in providing information about the other player's actions [31]. Pairs of subjects played the game in each of three conditions:

A: Face to Face AR – where they had virtual graphics superimposed over a live video view from the camera.
B: Face to Face non AR – where they could see the graphics only, not the live video input
C: Non Face to Face gaming – where the players could not see each other and also could see the graphics only. There was no live video background used.

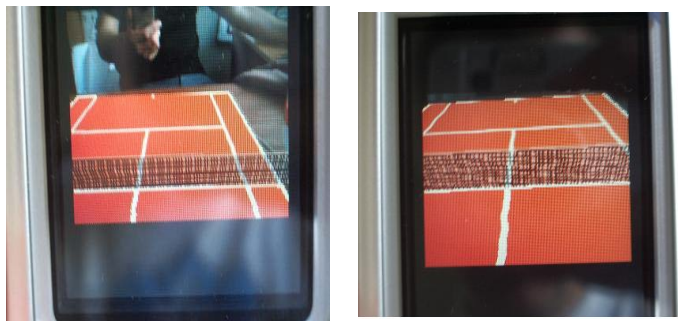Fig. 7 shows a screen shot of the application running with and without the live video background.

Fig. 7. The application with and without a live video background

Twelve pairs of subjects played games in each of the three conditions. At the end of each condition subjects were asked the following four questions:

1/ How easy was it to work with your partner?
2/ How easily did your partner work with you?
3/ How easy was it to be aware of what your partner was doing?
4/ How enjoyable was the game?

Each questions was answered on a Likert scale from 1 to 7 where 1 = Not Very Easy and 7 = Very Easy. Table 5 shows the average scores for each question across all conditions.
As can be seen the responses to questions 4 are almost the same. An ANOVA test on these questions found no statistical difference, meaning that users found each condition equally enjoyable. Interestingly enough, despite simple graphics and limited interactivity this enjoyment score was relatively high.

However there was a significant difference in response to the first three questions. For question 1 ($F_{(2,33)} = 8.17$, $p < 0.05$)

TABLE 5: SUBJECTIVE SURVEY RESPONSES

	A	B	C
Q1	4.83	4.13	3.75
Q2	4.50	4.25	3.83
Q3	5.92	3.17	2.50
Q4	5.17	5.25	5.33

and for question 2 ($F_{(2,33)} = 3.97$, $p < 0.05$). The user felt that there was a difference between the conditions in terms of how easy it was to work with their partner and how easily their partner worked with then. There was also a highly significant difference in response to question 3 ($F_{(2,15)} = 33.4$, $p < 0.0001$). Users felt that it was much easier to be aware of what their partner was doing in the face to face AR condition with the live video background than in the other two conditions which had no video background.

Subjects were also asked to rank the three conditions in order of how easy it was to work together where 1 = easiest and 3 = most difficult. Table 6 shows the average rankings.

TABLE 6: SUBJECTIVE RANKINGS

A	B	C
1.08	2.25	2.67

Again there is a significant difference across ranking values ($F_{(2,33)} = 34.1$, $p < 0.001$). Remarkably, all but one of the users (11 out of 12) ranked the Face to Face AR condition as the easiest to work together in, and then split their opinion almost evenly between the remaining two conditions. This confirms the results from the earlier survey questions.

Players felt they were more aware of what their partner was doing in the face to face AR condition (A) than in the non-AR face to face condition (B) or with remote players (C). They ranked the AR condition as much easier to work together in than in the non-AR face to face condition or with remote game playing. From these results it seems that AR interfaces could provide a greater level of awareness in face to face game play on mobile phones than with more traditional games.

After the experiment was completed subjects were also briefly interviewed about their experience. In general people overwhelmingly felt that seeing the AR view aided the face to face collaboration. Condition C was least favorite, because the collaborator was not visible – either in the phone or in peripheral vision. One subject even said that he didn't feel like he was playing with another person in condition C. Several people also commented on how adding graphics cues such as virtual shadows and a more realistic lit and shaded ball would help with the depth perception.

V.   DESIGN RECOMMENDATIONS

In this paper we have discussed the idea of using a tangible interface metaphor to interact in mobile AR applications. This could be particularly valuable for mobile phone based AR, where the motion of the phone itself can be used for input, and the phone becomes a tangible input object. The use of an appropriate tangible object input metaphor is important for the usability of mobile phone AR applications.

We then described four example mobile AR applications that used the tangible object input metaphor:

A: AR Viewer: a viewer that enabled users to see the model from different positions by moving the phone or marker.

B: AR Blocks: users can select and move blocks in a scene

C: Block Manipulation: a user study to explore AR interaction techniques on the mobile phone.

D: AR Tennis: A collaborative AR application in which the phones acted likes real tennis racquets.

Important lessons can be learned from each of these applications that can provide design guidelines for future handheld AR interfaces. In the case of the AR viewer having the virtual model appear to be attached to a real piece of paper meant that the user can change their view of the model by either moving the mobile phone or the paper. User could even move both the phone and the physical object holding the virtual model at the same time. This flexibility shows the ease of use that comes from combining the tangible input metaphor with the traditional tangible AR approach of attaching virtual content to physical objects.

This is further emphasized in the AR Blocks and Block Manipulation applications. In the case of the AR Blocks users were able to select and move virtual blocks by pressing and holding down a key on the keypad and moving the phone. Positioning virtual blocks by moving the phone was very natural, while the keypad input provided an easy way of switching between selection and viewing modes. The Block Manipulation application and experiment explored the usability of the tangible input metaphor further and found that while users found it more intuitive and faster to position virtual objects using the phone motion; it was quicker for them to orient the object using keypad input. This may have been because of the difficulty of finding a natural way of mapping phone rotation to virtual block rotation, and the problem of losing sight of the phone display while turning the phone. This suggests that techniques for positioning and rotating virtual objects may need to be decoupled into separate methods for rotation and translation.

The AR Tennis application used the tangible input metaphor to enable mobile phones to become virtual tennis racquets. In was natural for users to move their phones in front of the virtual ball, and the users received multimodal feedback while playing the game (audio, visual and tactile vibration cues). A user study with interface also showed the players preferred seeing each other when playing face to face and the AR gaming mode was preferred over the non-AR case. This shows that with the right interaction metaphor, AR interfaces can enhance face to face collaboration and provide new gaming experiences.

We can summarize the lessons learned into the following list of design recommendations for handheld AR interfaces:

- support bimanual input techniques that allows users to interact with real objects while moving the device.
- decouple manipulation modalities to match the affordances of the device, such a using device motion for virtual object positioning and keypad input for rotation.
- support multimodal feedback that uses the audio, tactile (vibration), and graphic abilities of the mobile device.
- use AR overlay to enhance face to face collaboration.

VI.   CONCLUSIONS AND FUTURE WORK

In this paper we have described the trend towards handheld AR devices for mobile Augmented Reality, and reviewed previous work in the area. Although there have been previous mobile AR interfaces, there has been less work on mobile phone based AR applications, and interaction metaphors suitable for handheld AR applications.

We describe the tangible input metaphor, where the motion of the handheld device itself can be used for user input. This is particularly suitable for mobile phone based AR applications where output (LCD screen) and input (camera, keypad) capabilities are combined in one device and can be used one handed. The usefulness of this metaphor is shown by four mobile phone AR applications that we have developed. The user studies we have conducted show that users find the tangible input technique to be intuitive and enjoyable and can prefer AR applications to non-AR applications.

However, this is just the beginning to research that needs to be conducted in handheld AR. As the PDA and phone hardware continues to improve a wider variety of applications should be explored. For example, with higher resolution screens and dedicated graphics hardware mobile devices will be able load complex engineering and medical content, while next generation wireless networks will support high bandwidth collaboration.

There is also a need for more usability testing and development of usability analysis techniques suitable for mobile devices. Although there have been previous examples of handheld AR applications, few of them have been evaluated using formal evaluation methods. Mobile devices are also used in a range of different real world contexts that are significantly different from the research laboratory environment. Methods need to be developed for evaluating handheld AR applications in the context of use.

Finally, there needs to be better development tools available for building handheld applications. Application development for mobile devices, especially mobile phones, is significantly more difficult than desktop development. If handheld AR is going to reach its full potential then it needs to be easier for AR applications to be created.

If these concerns are addressed then mobile devices combined with intuitive AR interfaces could indeed result in widespread commercialization of AR applications.

REFERENCES

[1]   I. Sutherland. The ultimate display, In *Proceedings of the IFIP Congress,* vol.2, 506-508, 1965.

[2]   R. Azuma. A Survey of Augmented Reality, *Presence: Teleoperators and Virtual Environments*,vol.6,no.4, pp. 355 – 385, 1997

[3]   T. Starner, S. Mann, B. Rhodes, J. Levine, J. Healey, D. Kirsch, R. W. Picard and A. Pentland. Augmented Reality through Wearable Computing, *Presence: Teleoperators, and Virtual Environments,* vol. 6, no.4, pp. 386-398,1997

[4]   S. Feiner, B. MacIntyre, T. Höllerer and A. Webster. A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment, *Proceedings of the First International Symposium on Wearable Computers (ISWC 97)*, Cambridge, Massachusetts, USA, pp. 74–81, 1997

[5]   B. Thomas, B. Close, J. Donoghue, J. Squires, P. De Bondi, M. Morris, and W. Piekarski. ARQuake: An Outdoor/Indoor Augmented Reality First Person Application, In *Proceedings of the 4th International Symposium on Wearable Computers (ISWC 2000)*, pp 139-146, Atlanta, GA, Oct 2000.

[6]   V. Vlahakis, N. Ioannidis, J. Karigiannis, M. Tsotros, M. Gounaris, D. Stricker, T. Gleue, P. Daehne and L. Almeida. Archeoguide: An Augmented Reality Guide for Archaeological Sites, *IEEE Computer Graphics and Applications,* vol. 22, No. 5, pp.52-60, September 2002

[7]   R. Suomela and J. Lehikoinen. Context compass, In *Proceedings of the 4th International Symposium on Wearable Computers (ISWC 2000)*, pp.147-154, Atlanta, GA, Oct 2000.

[8]   G. Reitmayr, D. Schmalstieg. Mobile Collaborative Augmented Reality, In *Proceedings of the 2nd IEEE and ACM International Symposium on Augmented Reality (ISAR 2001),* October 2001, New York.

[9]   G.Fitzmaurice. Situated Information Spaces and Spatially Aware Palmtop Computers, *Communications of the ACM,* vol.36, no.7, pp. 38-49 , July 1993

[10] J. Rekimoto, K. Nagao. The World through the Computer: Computer Augmented Interaction with Real World Environments, *User Interface Software and Technology (UIST '95),* pp. 29-38, November 14-17, 1995

[11] J. Rekimoto. TransVision: A Hand-held Augmented Reality System for Collaborative Design, In *Proceedings of Virtual Systems and Multi -Media (VSMM '96),* Sept. 1996

[12] D. Mogilev, K. Kiyokawa, M. Billinghurst and J. Pair. AR Pad: An Interface for Face-to-Face Collaboration, In *Proceedings of CHI 2002.* April 20-25th, Minneapolis, Minnesota, 2002.

[13] C. Geiger, B. Kleinjohann, C. Reimann and D. Stichling. Mobile AR4ALL, ISAR 2001, The *Second IEEE and ACM International Symposium on Augmented Reality,* New York, 2001

[14] D. Ingram, J. Newman. Augmented Reality in a WideArea Sentient Environment, In *Proceedings of the 2nd IEEE and ACM International Symposium on Augmented Reality (ISAR 2001),* October 2001, New York.

[15] ARToolKit website: http://artoolkit.sourceforge.net/

[16] D. Wagner, D. Schmalstieg. First steps towards handheld augmented reality, In *Proceedings of the 7th International Symposium on Wearable Computers (ISWC2003)*, White Plains, NY, USA, IEEE Computer Society, pp.127–137, 2003

[17] D. Wagner, T. Pintaric, F. Ledermann and D. Schmalstieg. Towards Massively Multi-User Augmented Reality on Handheld Devices, In *Proceedings of the Third International Conference on Pervasive Computing (Pervasive 2005),* 8th-13th May, Munich, Germany.

[18] D. Cutting, M. Assad, A. Hudson. AR phone: Accessible augmented reality in the intelligent environment, In *Proceedings of OZCHI 2003*, Brisbane, 2003.

[19] A. Henrysson, M. Ollila. UMAR - Ubiquitous Mobile Augmented Reality, In *Proceedings of the Third International Conference on Mobile and Ubiquitous Multimedia (MUM2004)* pp. 41-45. College Park, Maryland, U.S.A. October 27 - 29, 2004.

[20] M. Moehring, C. Lessig and O. Bimber. Video See-Through AR on Consumer Cell Phones, In *Proceedings of International Symposium on Augmented and Mixed Reality (ISMAR'04),* pp. 252-253, 2004.

[21] W. Piekarski, B. Thomas. Tinmith-Hand: Unified User Interface Technology for Mobile Outdoor Augmented Reality and Indoor Virtual Reality, In *Proceedings of IEEE Virtual Reality (VR 2002)*, Orlando, Fl, Mar 2002.

[22] T. Kurata, T. Takashi Okuma, M. Kourogi and K. Sakaue. The Hand-mouse: A Human Interface Suitable for Augmented Reality Environments Enabled by Visual Wearables, In *Proceedings of the International Symposium on Mixed Reality (ISMR 2001),* Yokohama, Japan, pp.188-189, 2001

[23] M. Träskbäck, M. Haller. Mixed reality training application for an oil refinery: user requirements, In *ACM SIGGRAPH International Conference on Virtual Reality Continuum and its Applications in Industry, VRCAI 2004,* pp. 324- 327, Singapore, 2004.

[24] D. Wagner, I. Barakonyi. Augmented reality kanji learning, In *Proceedings of the 2003 IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 2003),* Tokyo, Japan, IEEE Computer Society, pp. 335–336, 2003

[25] C. Reimann, V. Paelke and D. Stichling. Foot-based mobile Interaction with Games, In *Proceedings of the ACM SIGCHI International Conference on Advances in Computer Entertainment Technology (ACE 2004).* Singapore, June 2004

[26] H. Kato, M. Billinghurst, I. Poupyrev, N. Tetsutani and K. Tachibana. Tangible Augmented Reality for Human Computer Interaction, In *Proceedings of Nicograph 2001.* Nagoya, Japan, 2001.

[27] J. Wang, S. Zhai and J. Canny. Camera Phone Based Motion Sensing: Interaction Techniques, Applications and Performance Study, In *ACM UIST 2006,* Montreux, Switzerland, October 15-18, 2006.

[28] J. Hwang, J. Jung and JG. Kim. Hand-held Virtual Reality: A Feasibility Study, In *Proceedings of ACM Virtual Reality Software and Technology (VRST 2006)* (to appear), *2006*

[29] A. Henrysson, M. Ollila and M. Billinghurst. Mobile Phone Based AR Scene Assembly, In *Proceedings of the Fourth International Conference on Mobile and Ubiquitous Multimedia (MUM 2005),* Christchurch, New Zealand, December 8th – 10th, 2005

[30] A. Henrysson, M. Billinghurst and M. Ollila. Virtual Object Manipulation using a Mobile Phone, In *proceedings of the 15th International Conference on Artificial Reality and Telexistence (ICAT 2005),* Christchurch, New Zealand, December 5th – 8th, 2005

[31] A. Henrysson, M. Billinghurst and M. Ollila. Face to Face Collaborative AR on Mobile Phones, In *Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR 2005),* Vienna, Austria, October 5th – 8th, 2005

Mark Billinghurst (M'02–SM'81–F'87) Ph.D. (Electrical Engineering), University of Washington, Seattle, WA, USA, 2002. MPhil (Applied Mathematics), Waikato University, Hamilton, New Zealand, 1992. BCMS (1st Class Hons.) (Applied Mathematics), Waikato University, Hamilton, New Zealand, 1990.

He is an Associate Professor and Director of the Human Interface Technology Laboratory New Zealand (HIT Lab NZ) at the University of Canterbury, Christchurch, New Zealand. Previously he has worked as a Research Associate at the University of Washington, and has held appointments at the MIT Media Laboratory, British Telecom's Advanced Perception Unit and the ATR Laboratory in Japan. He conducts research in Augmented and Virtual Reality, wearable computing, mobile phones and conversational interfaces.

Dr. Billinghurst is a member of the ACM.