Bimanual Tangible Interaction with Mobile Phones

Darren Edge
Microsoft Research Asia
5F Beijing Sigma Center, 49 Zhichun Road
Beijing, China
Darren.Edge@microsoft.com

Alan F. Blackwell
University of Cambridge Computer Laboratory
William Gates Building, 15 JJ Thomson Avenue
Cambridge, UK
Alan.Blackwell@cl.cam.ac.uk

ABSTRACT
In the context of tangibility, mobile phones are rapidly becoming sensor-rich handheld computers with the potential to take better advantage of our physical capabilities and our lifetime of experiences interacting both in and with the world around us. In this paper, we analyse four different ways in which mobiles can be used to represent and control digital information, showing that each resulting interaction style is characterized by a unique coordination of the user's attention and two hands in relation to the mobile device. We present our analysis in terms of a framework that can be used to critically examine future schemes of bimanual interaction with mobile phones.

Author Keywords
Mobile Interaction, Bimanual Interaction, Augmented Reality, Tangible User Interfaces

ACM Classification Keywords
H5.2 Input devices and strategies

INTRODUCTION
The desire to combine the power of computing with our familiarity of the physical world is a central concern of research into tangibility, whether we see our creations as "disappearing" ubiquitous technologies [22], "graspable" user interfaces [4], "legible" tangible user interfaces (TUIs) [9], or "shareable" platforms for tangible interaction [8].

In this paper we examine the tangible interaction potential of an everyday artefact: our mobile phone. Since mobile phones are handheld devices and interaction typically involves both hands, we revisit the literature on bimanual skill, or how we coordinate the use of our two hands according to their preferred roles [6]. Although such bimanual coordination was a pressing concern for "graspable" interaction [4], in which the user's two hands operated "function handles" specialised both to the task and the preferred roles of the hands, the subsequent emphasis on supporting collaborative interaction with "shareable" tangibles [8] has generally decreased the degree to which those tangibles are specialised for use in a particular hand. As a consequence, more recent work on TUIs has tended to treat bimanual interaction as something that comes about naturally, rather than something to be designed for. However, if we are to consider styles of interaction in which one hand is holding a mobile phone, we need to think explicitly about which hand it should be, what the other hand should be doing, and why.

We are also interested in investigating less conventional uses of mobile phones, specifically those with embedded cameras. Mobile camera phones have the potential to augment physical objects in a manner reminiscent of the tangible "lens" metaphor, "a physically instantiated window which allows haptic interaction with 3D digital information bound to physical objects" [9]. By streaming a phone's incoming camera images to its display, it becomes a "video see-through lens". This loop can be intermediated by image processing and computer vision algorithms that analyse the image stream for certain objects or gestures, appropriately inserting virtual representations and controls into the observed scene. While such "video see-through lenses" are not a conventional structural form for TUIs, the technique is common in the parallel field of Augmented Reality (AR).

Although the stereotypical AR application is based on an immersive, head mounted display, the recent shift towards mobile AR [17] emphasises the practical benefits of using the almost universal mobile camera phone as an augmentation device. These benefits also represent a huge economy for tangible interaction: rather than augmenting our environment with interactive surfaces everywhere, we appropriate the single device we already carry with us. Our concern is therefore with the travelling or casual computer user who is temporarily at rest, either standing or seated. Such a user might naturally take their phone from a pocket, hold it in the hand or place it on a convenient surface, and interact directly for a period of time before moving on.

We structure the paper in terms of four different ways in which people can make use of both hands in interaction supported by mobile camera phones. For each of these four interaction styles, we present existing systems that illustrate the underlying relationships between the user's hands, mobile phone, and environment. We then demonstrate the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
TEI 2009, February 16–18, 2009, Cambridge, UK.
Copyright 2009 ACM 978-1-60558-493-5/09/02 ...$5.00.

potential for composing different elements of these styles to create hybrid schemes of bimanual interaction. We conclude with a comparison of the usability properties of these four basic interaction styles.

THEORY OF BIMANUAL INTERACTION

Guiard's work on the asymmetric division of labour between the two hands abstracts away from the biomechanical and physiological complexity of skilled, coordinated bimanual activities, by modelling the hands as abstract motors that tend to operate in series [6]. This is the basis of Guiard's *kinematic chain* model, which accounts for the following phenomena commonly observed in the diverse activities of handwriting, violin playing, golf swing performance, sewing, driving a screw, and others:

1. *Precedence*. The non-dominant hand[1] precedes the dominant hand in two-handed interactions.
2. *Reference*. The non-dominant hand sets a spatial frame of reference for actions of the dominant hand.
3. *Scale*. The dominant hand performs actions on a finer spatial-temporal scale that the non-dominant hand.

These principles account for our everyday actions in the real world, in which bimanual cooperation entails both hands operating on the same physical objects, in the same physical space. However, as our everyday actions become increasingly mediated by technology, we need to consider the potential for the actions of our two hands to be *conceptually linked* in our minds, even though our hands might be operating on physically independent devices. Experiments suggest that the Guiard *reference* principle is robust against changes in the kinaesthetic reference frame of the two hands, provided there is adequate visual feedback [1]. The implication for tangible interaction with mobile camera phones is that two hands can operate on distinct physical objects in disjoint physical spaces and still co-operate in the performance of a common task.

BIMANUAL INTERACTION WITH MOBILES

The Guiard principles provide us with an abstract model of bimanual skill around which we can design interactions with an appropriate division of labour between the two hands, with each performing in a manner consistent with observed behaviour in both natural and artificial settings.

If we are to design bimanual control schemas specifically for interaction with mobile camera phones, we need to consider the potential of such phones from the perspectives of both *representation* and *control*. In terms of representation, the conventional use of mobile phones is to interact with *virtual applications* displayed on-screen. As outlined in the introduction, however, mobile camera phones also have the potential to act as "video see-through lenses" to *augmented reality*. The decision whether to index information in screen coordinates or world coordinates is determined not by the roles of the hands, but by the goal of the application. A similar two-way choice is also seen in the potential means of control. This can be achieved *on phone*, making direct use of the many built-in physical controls (buttons, joysticks, touch-screens, etc.), or *in space*, using the whole phone to point at objects or make gestures in the air (perhaps using physical controls indirectly, to protect against unintentional action performance). Each combination of design choices for the representation and control of information with mobile phones characterizes a particular *bimanual interaction style*. We use a "painting" metaphor to present and compare such styles, since the familiar pointing and mark-marking aspects of painting can be understood independently of the technological means of mark production, and yet generalized beyond simple painting to encompass selection, gesture, and widget manipulation. Our framework for bimanual interaction with mobile camera phones, based on this analysis of representation and control, is shown in Figure 1.

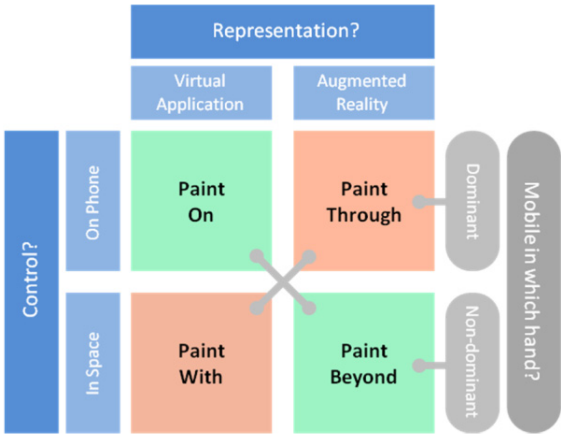

Figure 1. Framework for bimanual interaction with mobiles

We now present the four bimanual painting styles delineated by our framework, in each case showing how the two hands cooperate in their preferred roles to achieve interaction qualities appropriate to the application domain.

Paint On

The use of a stylus or finger to "paint" on a phone screen that represents a canvas-like 2D virtual application is the standard means of mobile interaction arising out of the WIMP paradigm of *Windows*, *Icons*, *Menus*, and a *Pointer*. Although multi-touch computing (popularised by the Microsoft Surface and Apple iPhone) has the potential for rich gestural interaction that is more direct and appealing than basic pointing, it is an interaction technique that extends WIMP rather than a paradigm that replaces it.

In both cases, the "paint on" interaction style mirrors our physical interactions with pen and paper. The non-dominant hand begins with coarse framing actions that appropriately

[1] The dominant hand is the one preferred in single hand interactions such as writing, and is most commonly the right hand.

orient the phone/paper, while the dominant hand follows with accurate mark-making actions on the surface of the phone/paper. Although this interface might not be considered a "TUI" in the sense that it doesn't physically embody digital information [20], the interactions it supports are undeniably tangible in that they exploit familiar physical behaviours (drawing/writing) and support mobile users who need to interact anywhere in space.

Example systems

The Augmented Maps system [13] demonstrates how a stylus-operated PDA, similar in form and function to many mobile phones, can present a GUI front-end to both printed and projected "assets" of augmented physical maps. Such assets are *space-multiplexed* across the map – all asset locations are visible all of the time. The PDA is then used as a *time-multiplexed* controller, "picking up" whichever map asset it is placed on top of (as viewed by an overhead camera), and then being used to manipulate that asset in "paint on" style. The system, shown in Figure 2, used a PDA to conveniently adjust parameters in a flood simulation application projected onto a physical paper map.

Figure 2. Tangible control of an augmented map

In addition to conventional button pressing, other systems have attempted to extend the physical interaction vocabulary of mobile phones through augmentation with embedded sensors. One approach is using internal accelerometer data to recognize "taps" on the phone's casing, and interpreting these taps based on the interaction context (phone in pocket or resting , ringing from an alarm or message, etc.) [16]. Another approach is to increase the interaction fidelity of keypad buttons, additionally detecting touch contact with keys and the force of key presses [7]. Both approaches extend the "paint on" interaction style.

Paint With

In the previous section, we described how a mobile phone with a touch screen can be used as a virtual canvas that users can "paint on". By estimating the motion of a phone based on the *optical flow* observed in the stream of camera images, a mobile phone can also be used as a virtual brush that users can "paint with" in space.

In the one-handed solution, the effects of stroking in the air can be visualized on the phone's screen, albeit at a reduced scale. However, the difficulty of visually tracking the moving phone for virtual feedback can be eliminated with the introduction of a second mobile display device (such as a UMPC, PDA or additional mobile phone). In this two-handed solution, the user can focus his visual attention on the display device held steady in his non-dominant hand, while his dominant hand performs fine-grained painting with the "optical brush" of the camera phone.

Example systems

Here we report on an undergraduate research project investigating bimanual interaction possibilities with two consumer handheld devices – the Nokia 770 "Internet Tablet" and the Nokia N80 Smartphone. The N770 has a 7cm high-resolution touch screen that supports stylus interaction, a few navigation buttons, and support for Bluetooth peripherals. The N80 has a 3MegaPixel camera, phone and navigation buttons, a "shutter" button and support for both Bluetooth and WiFi communication. The manufacturer's expectation for these devices is apparently that a customer might buy both, and that they would be complementary (the N770 has neither a SIM card nor camera, and its user interface places Bluetooth connection to a phone as one of the foreground configuration tasks).

In the design experiment, we used a re-implementation of the TinyMotion [21] optical flow algorithm to estimate continuous motion of the N80 phone, with movement being communicated to the N770 over Bluetooth. We subsequently supported the archetypical "paint with" interaction style, with the N80 Smartphone used as a straightforward "optical brush" in an actual painting application running on the N770 tablet (Figure 3a).

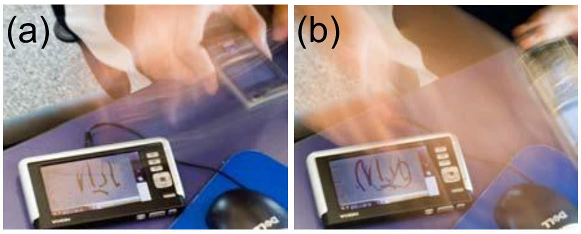

Figure 3. Painting with paired mobile devices: dominant hand paints with phone (a); dominant hand paints with stylus (b)

Although simple "mousing" with a camera phone needs little instruction, this is not the case for more complex gestural interfaces. Learning the sensing capabilities of a device, the gestures detected from the sensor data, the correct performance of these gestures in different situations (e.g. seated, standing, walking), and also learning the mappings from input gestures to output effects, all contribute towards a steeper learning curve for users. However, recent work has demonstrated the potential for teaching users what accelerometer-based interfaces can and can't detect, by combing visual feedback in terms of transformations of a virtual 3D "active cube" shown in the phone's display, and haptic feedback in terms of vibrotactile pulses that punctuate detected gestures [12].

Paint Through

The two bimanual interaction styles presented so far have had *virtual applications* as the target of "painting" activity. In this section, we analyse the potential of a mobile camera phones to act as a video see-through "lens" that users can "paint through", such that physical actions made on the lens (key-presses, screen-touches, etc.) are reflected in virtual changes to real-world objects as they appear through the lens. The change is not made in reality, but in an *augmented reality*. In such systems, the accurate targeting of objects in the world is facilitated by the "lens" being held by the dominant hand, such that it can examine and select objects brought before it with the non-dominant hand.

Example systems

In earlier research [19], we used camera phones as handheld interfaces to digital content linked to from circular barcode-style visual tags printed inside a brochure (Figure 4).

Figure 4. Linking from tangible surfaces

That earlier combination of brochure-plus-phone, when interpreted as image-plus-camera, suggested a bimanual interaction style in which images such as photographs that were meaningful to the user might be "re-connected" to the camera that took them, but without any obvious technical apparatus. Rather than visual tags, we used an algorithm that matches images based on a tree of low-level "keypoints" that are particularly salient [11]. The algorithm is insensitive to small degrees of image masking, tilt, orientation and lighting conditions, so that an image considered by the user to be unique is recognised as such when held in front of the camera. Although this student project simply identified the photos placed before a mobile camera phone, the combination of automatic object identification and manual phone input (key presses, joystick shifts, screen taps/strokes) both ensures actions are intentional and extends the interaction vocabulary.

Paint Beyond

The example systems of the previous section can be seen as creating a "lightweight" augmented reality, with virtual augmentations appearing as coloured 2D shapes and text superimposed on the underlying video feed. In more "heavyweight" augmented reality, printed 2D markers are linked to more complex 3D augmentations that are rendered to appear as "real" objects anchored to the markers. Augmentation is not by introducing virtual *annotations* to the video feed, but by introducing virtual *objects*. The appearance of these richer augmentations depends on the viewing distance and angle, and these can equivalently be changed by moving the camera relative to the marker, or by moving the marker relative to the camera. Whereas the former method encourages "paint through" interaction, with the camera phone deftly manoeuvred in space by the dominant hand, the latter method encourages a switch of the phone's position to the non-dominant hand.

With the camera phone "lens" held steady in the non-dominant hand, the dominant hand is free to manipulate the 2D marker in the physical world: rotating, tilting, and shifting the marker to gain a full hemispherical view of the virtual augmentation. Moreover, the detection of hand postures and gestures in the vicinity of markers can allow users to directly manipulate the associated augmentations. Rather than the dominant hand being used to indirectly "paint through" the mobile device, it is now used to "paint beyond" the imaginary surface defined by the camera's orientation in space.

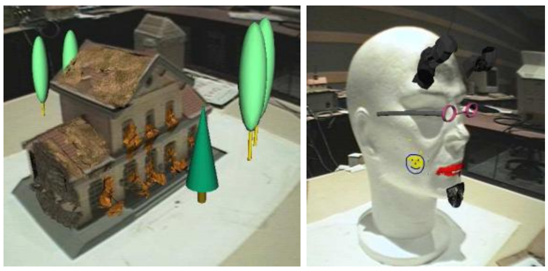

Figure 5. Authoring in augmented reality

Example systems

For "authoring" in augmented reality, Grasset et al. [5] have developed a suite of interactive tools based on painting, grabbing, and gluing both real and virtual objects (Figure 5). The user interacts with a tracked pen that they move between a tool palette, an experimental "scratch" area, and the working surface on which the result is created. The pen is naturally held in the dominant hand, while the non-dominant hand assists by adjusting the position and orientation of the tool palette and working area as required. Although this system was developed for augmentation via a Head Mounted Display (HMD), it is possible to imagine a mobile camera phone replacing the HMD and the user's hand (combined with appropriate gesture recognition) replacing the pen and tool palette. This approach can then be seen as creating a physical analogue of a "tool lens" [10] – one of multiple "see through tools" [2] for two-handed desktop interaction.

DISCUSSION

Composition of Bimanual Interaction Styles

Each of the four interaction styles presented in this paper can be seen as a building block for more complex schemes of bimanual control. We examined the potential for such combinations in a second design experiment, building on our ability to "paint with" a Nokia N80 Smartphone. As with the system shown in Figure 3a, we chose painting as our application. However, rather than using the dominant hand to move the "optical brush" in space, with the resulting strokes visible on the Nokia N770 tablet, we decided on an interaction style that made full use of the available display space and means of control. In our hybrid interaction style, the dominant hand uses a stylus to "paint on" the N770, which rests on any convenient surface. This frees the non-dominant hand to coarsely and casually adjust visual stroke parameters using the N80 Smartphone. For continuous manipulation of stroke width, the N80 is tilted upwards and downwards – making use of optical flow in "paint with" style. For discrete selection of colour, the N80 is moved from side to side in a "scrubbing" gesture that "picks up" a colour from the environment. This latter technique combines the notions of "paint through" and "paint with", but with effects visible in the virtual painting application running on the N770, rather than in an augmented reality visible through the N80. The overall effect is to provide interactive functionality at least equivalent to the N770's native painting application, but requiring no onscreen controls. Furthermore, simultaneous bimanual actions allow creative effects that are not possible when a single stylus alternates between paint and control tasks. The application in use is shown in Figure 3b.

A number of mobile tangible interfaces from the literature also incorporate more than one of our interaction styles. The most popular combination is a two stage interaction, with a "paint through"-style acquisition of an AR marker followed by "paint with"-style gesturing. The CyberCode system [14] used camera-enabled mobile devices to point at one of many physical AR markers spread throughout the environment, triggering the virtual display of a contextual radial marking menu around the physical marker. Subsequent motion of the device in the direction of a menu option makes the selection. This interaction pattern is extended further in the Vidgets system [18], which uses human-readable markers to depict the nature of the supported "paint with" gesture – e.g. tilt to scroll/slide the virtual media, rotate to fast forward/rewind, and so on. A thorough analysis of all combinations of device movement relative to a fixed AR marker has also been published [15]. These combinations are described in terms of "interaction primitives", which may be static (pointing, rotation, tilting, distance, stay, and keystroke) or dynamic (sweeping and relative movement).

Comparison of Bimanual Interaction Styles

We can analyse the abstract usability properties of the four basic styles of bimanual mobile interaction using the *Tangible Correlates of the Cognitive Dimensions* [3]. These are a vocabulary for describing the usability properties of tangible notations, in such a way that highlights the tradeoffs between different design candidates. The key tradeoffs manifest in our 2×2 design space of interaction styles are presented in Figure 6.

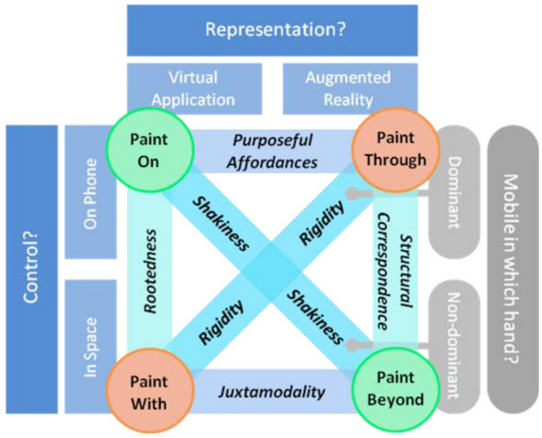

Figure 6. Usability properties of interaction styles (opposite sides and diagonals of the square indicate tradeoffs)

Purposeful Affordances versus Juxtamodality

In the "paint on" and "paint through" interaction styles using the phone's existing physical controls, corresponding systems inherit the `purposeful affordances` of these controls as they were designed to be used. Physical interaction is thus simple, familiar, and reliable. However, whereas on-phone interaction requires the user's attention to be directed towards the phone itself, this is not the case with "paint with" and "paint beyond". These two styles both allow the dominant hand to paint in space while the non-dominant hand holds the representation steady – the user's visual and haptic attentions are divided in the physical world, but combined via visual feedback in the virtual representation. This property is called `juxtamodality`, and supports computational intermediation between physical actions and their digital effects (like using a mouse to point on a screen). For "paint with" interactions, this might be to *reduce* the effects of small, involuntary actions (such as those arising from hand tremor or a bumpy ride) by making the spatial scale of interaction larger than the touch screen of a mobile device. For "paint through" interactions, this might be to *increase* the effects of small, voluntary actions by interpreting them as gestures that have associated virtual actions.

Rootedness versus Structural Correspondence

In the "paint on" and "paint with" interaction styles, the use of virtual applications as the means of representation means that interactions are not rooted to any particular physical location – they have low `rootedness`. This "use anywhere" characteristic make these interaction styles the most general. Representations in augmented reality do not share this characteristic – they are typically anchored to particular physical contexts. An advantage of "paint through" and "paint beyond" interaction styles in this respect is that they can exhibit high `structural correspondence` between the representation and the problem domain. These interaction styles are more specialised, suitable for activity domains that are spatial in the geographic sense, such as social geotagging, or spatial in the contextual sense, such as using physical markers to index and interact with digital media in a spatially-distributed way.

Shakiness versus Rigidity

The use of the dominant hand to "paint through" or "paint with" a phone it is holding makes use of its fine-grained interaction capabilities. This mapping means that interactions encounter low `rigidity` – the interaction styles do not resist changes to the underlying representation. Interaction can be both casual and one-handed, if required. However, this can also increase the potential for accidental change, or `shakiness`. The degree of `shakiness` can be reduced by requiring both hands to operate simultaneously, as an indication of intentional action. This is the case with both "paint on" and "paint through", where changes only occur when the actions of the dominant hand are made in the frame of reference of the mobile phone, held and actively positioned by the non-dominant hand.

CONCLUSIONS

In this paper we have presented a framework for asymmetric bimanual interaction with mobile phones, supported by examples of mobile systems from the literature and our own design explorations. The framework can be used to critically examine bimanual schemes of interaction with mobile phones, in terms of the relationships between the user's two hands, their attention, their mobile phone as a means of representation and control, and their physical environment.

ACKNOWLEDGMENTS

This research was supported by The Boeing Company, by Nokia, and by the Engineering and Physical Sciences Research Council undergraduate bursary programme. We are grateful to Gareth Bailey, Vincent Chen, Mattias Linnap, and Hugh Warrington for their work on the prototype systems described in this paper.

REFERENCES

1. Balakrishnan, R. & Hinckley, K. (1999). The role of kinesthetic reference frames in two-handed input performance. UIST '99, 171–178
2. Bier, E.A., Stone, M.C., Fishkin, K., Buxton, W. & Baudel, T. (1994). A Taxonomy of See-Through Tools. CHI'94, 358-364.
3. Edge, D. & Blackwell, A.F. (2006). Tangible Correlates of the Cognitive Dimensions for Tangible User Interface. JVLC, 17(4), 366–394
4. Fitzmaurice, G.W. (1996). Graspable user interfaces. PhD thesis, University of Toronto
5. Grasset, R., Boissieux, L., Gascuel, J.D. & Schmalstieg, D. (2005). Interactive mediated reality. AUIC'05, 21–29
6. Guiard, Y. (1987). Asymmetric Division of Labor in Human Skilled Bimanual Action: The Kinematic Chain as a Model. J. of Motor Behavior, 19 (4), 486–517
7. Holleis, P., Huhtala, J., and Häkkilä, J. 2008. Studying applications for touch-enabled mobile phone keypads. TEI'08,15–18
8. Hornecker, E., Marshall, P. & Rogers, Y. (2007). From entry to access: how shareability comes about. DPPI'07, 328–342
9. Ishii, H. & Ullmer, B. (1997). Tangible bits: towards seamless interfaces between people, bits and atoms. CHI'97, 234–2411
10. Kabbash, P., Buxton, W. & Sellen, A. (1994). Two-handed input in a compound task. CHI '94
11. Lepetit, V. & Fua P. (2006). Keypoint Recognition Using Randomized Trees. IEEE Trans. Pattern Anal. Mach. Intell., 28(9), 1465-1479
12. Linjama, J., Korpipää, P., Kela, J. & Rantakokko, T. (2008). ActionCube: a tangible mobile gesture interaction tutorial. TEI'07, 169–172
13. Reitmayr, G., Eade, E. & Drummond, T. (2005). Localisation and Interaction for Augmented Maps. ISMAR'05
14. Rekimoto, J. & Ayatsuka, Y. (2000). CyberCode: designing augmented reality environments with visual tags. DARE'00, 1–10
15. Rohs, M. & Zweifel, P. (2005). A conceptual framework for camera phone-based interaction techniques. Pervasive Computing 2005, number 3468
16. Ronkainen, S., Häkkilä, J., Kaleva, S., Colley, A. & Linjama, J. (2007). Tap input as an embedded interaction method for mobile devices. TEI'07, 263–270
17. Schmalstieg, D. & Reitmayr, G. (2005). The World as a User Interface: Augmented Reality for Ubiquitous Computing. CEMVRC'05
18. Tokunaga, E., Kimura, H., Kobayashi, N. & Nakajima, T. (2005). Virtual tangible widgets: seamless universal interaction with personal sensing devices. ICMI '05. 325–332
19. Toye, E., Sharp, R., Madhavapeddy, A., Scott, D., Upton, E. & Blackwell, A. (2006). Interacting with mobile services: An evaluation of camera-phones and visual tags. Pers. & Ubiq. Comp., Jan 2006, 1–10
20. Ullmer, B. & Ishii, H. (2001). Emerging Frameworks for Tangible User Interfaces. In J.M. Carroll (ed.) HCI in the New Millennium. Addison-Wesley, 579–601
21. Wang, J. & Canny, J. (2006). TinyMotion: Camera Phone Based Interaction Methods. CHI'06 Ext. Abstracts
22. Weiser, M. (1991). The computer for the 21st century. Scientific American, 265(3), 66–75