Proceedings of the Sixth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 2007)
Nara, Japan, Nov. 13-16, 2007, pp. 47-50.

Visual Hints for Tangible Gestures in Augmented Reality

Sean White Levi Lister Steven Feiner

Columbia University

ABSTRACT

Tangible Augmented Reality (AR) systems imbue physical objects with the ability to act and respond in new ways. In particular, physical objects and gestures made with them gain meaning that does not exist outside the tangible AR environment. The existence of this new set of possible actions and outcomes is not always apparent, making it necessary to learn new movements or gestures. Addressing this opportunity, we present *visual hints*, which are graphical representations in AR of potential actions and their consequences in the augmented physical world. Visual hints enable discovery, learning, and completion of gestures and manipulation in tangible AR. Here, we discuss our investigation of a variety of representations of visual hints and methods for activating them. We then describe a specific implementation that supports gestures developed for a tangible AR user interface to an electronic field guide for botanists, and present results from a pilot study.

CR Categories and Subject Descriptors: H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems—*Artificial, augmented, and virtual realities*; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*GUI*

Additional Keywords: Tangible augmented reality, visual hints, gestures, electronic field guide

1 INTRODUCTION

Tangible augmented reality (AR), which combines physical input devices with overlaid imagery, allows physical objects and the actions that can be performed on them to be overloaded with new meaning [7]. However, the interactions possible in such systems are not always obvious. While menus in a windowing system reveal the set of supported actions, interactive user-interface (UI) documentation that addresses the richer domain of tangible AR has not been well explored. Tangible objects can reveal affordances [6] based on their physical morphology, yet more complex manipulations and gestures are not always readily apparent. Much like manual gestural UIs, the ephemeral nature of tangible gestures makes them difficult to discover and properly learn [8].

In this paper, we investigate *visual hints* (Fig. 1), graphical representations in AR of potential UI actions associated with the physical world. Visual hints can potentially improve discovery and learning of gestures and manipulation in tangible AR. Our contribution focuses on developing and formalizing visual hints and presenting results from our implementation, use, and comparison of different kinds of visual hints. We start with a discussion of related work. We then present conceptual model for visual hints. Next, we discuss the space of potential interactions specific to activating and presenting visual hints. We follow this with a description of several instantiations of the technique and a report on a pilot study we ran to compare different methods of activation and presentation. We close with a discussion of the general applicability of the technique and future research directions.

* email: (swhite, ljl2116, feiner)@cs.columbia.edu

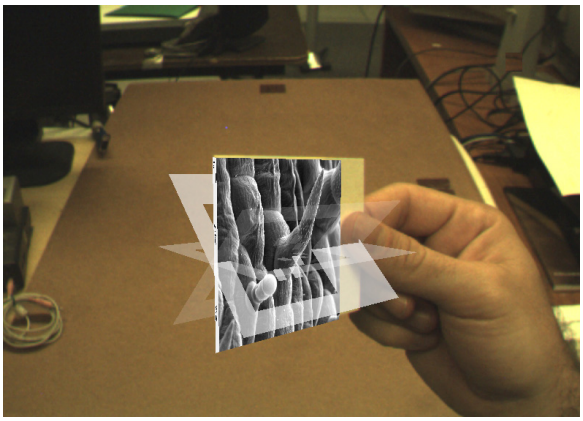

Fig. 1: A visual hint for a "twirling" gesture, represented by ghosting. (Viewed through a video see-through display.)

2 RELATED WORK

Recognition of the problem of discovering and learning gestural UIs is not new. Kjeldsen [8] observes that one problem with the use of hand gestures in UIs is learning the set of possible gestures, because "there are no memory aids for what to do next." Although he does not explore this, he suggests that pose icons that pop up during pauses might address the problem.

There is a long history of displaying annotations or related information to provide help in 2D graphical UIs, especially when the user pauses or hovers over an object, including Microsoft Tooltips, Apple Balloon Help, and contextual help systems [14]. Kurtenbach et al.'s marking menus [9] display the set of possible menu choices only when the user pauses. Recently, Ramos et al. [13] presented a browsing mode that displayed four possibilities for completing pen gestures that combine selection and action.

In AR, Tan et al. [15] developed Tangible Tooltips and Tangible Bubble Help in their Tiles system. Tooltips, displaying textual help, were activated when a tracked card was tilted. Tangible Bubble Help was activated when an activator card was brought in proximity to another card. Tooltips, in general, document *what* happens once you take an action, such as pressing a button, but do not show *how* to perform the action. In contrast, visual hints provide spatial and temporal guidance on executing a gesture.

Work on maintenance/assembly instructions has represented 3D physical actions graphically (e.g., as arrows) on desktops [1, 2] and in AR [3], but does not address gestural interaction.

3 VISUAL HINTS

An example of one type of visual hint is shown in Fig. 1. In our tangible AR UI to an Electronic Field Guide (EFG) for botanists [17], cards with fiducial markings can be manipulated to transform representations of plant species displayed on the card (called *virtual vouchers*) through semantic zooming, magnification, and change of species. One particular gesture involves a twirling or flipping motion that was readily learned through in-person demonstration, but was not apparent to uninitiated users of our original UI. In this example, the visual hint shows the user the motion needed to execute the gesture.

More generally, visual hints incorporate a variety of methods for representing actions that can be taken, ways to complete them, and their outcomes. Our implementation also supports activating/deactivating hints and cycling through them. Visual hints can be applied to aspects of the environment or specific tangible objects. Here, we investigate visual hints for gestures and manipulation in tangible AR.

3.1 Tangible Gestures

Tangible gestures involve the manipulation of tangible UIs. They provide a simple way to interact with information relevant to a particular object or abstraction. While the space of gesture and tangible UI taxonomies is broad, we focus on gestures that Quek et al. refer to as manipulative or semaphoric [12]. *Manipulative* gestures tightly couple the target of manipulation and the gesture. *Semaphoric* gestures are symbolic and typically represent a stylized dictionary of static or dynamic gestures in a system.

Several questions arise in exploring such gestures. How does one discover the gestural affordances of an object or environment? How does one learn the correct movement of a gesture? As a gesture is performed, how does one know it is being completed correctly? Although physical affordances often represent these aspects of a system, they may not always be present or sufficient to reflect the expanded capabilities imbued by AR.

3.2 Representation

There are many ways to represent a visual hint. For example, a textual explanation, a static diagram, or an animated extension of a tangible object can all provide information about the gesture. These representations fall into a design space that can be characterized along multiple dimensions, including media type, dynamics, anchoring, and proximity. Anchoring and proximity are particularly important. Hints can be anchored to an object or they can be anchored to, for example, the screen, the user's body, or the world, regardless of the position of the object. This is distinct from proximity, where a hint can be displayed close to or distant from an object. Close proximity, which has been shown to improve learning in multimedia [10], can also create the illusion that a hint is part of an object. We consider a variety of representations in our design space, some examples of which follow.

Textual hints (Fig. 2a) can be read and perceived quickly, but depend strongly on shared meaning and understanding. For example, the English word "flip" means different things in different cultures, has no standard gestural representation, and would need to be translated into other languages.

Diagrammatic hints (Fig. 2b) provide a spatial image of the appropriate gesture. Although there are some domain-dependent diagramming standards, images typically are designer-dependent. Their location and proximity can also vary. For example, a set of diagrams may be anchored to the top of the display or anchored to an object and presented in close proximity to it.

Ghosted hints (Fig. 2c) represent the action of the gesture in 3D space, starting from the current position of the object, and traversing through a series of ghost images that follow the trajectory of the gesture. *Ghosting* is a well-known illustrative technique in comics [11] and in manual and automated [4] graphic design, in which an object is rendered semitransparent to represent its past or future state, or to allow other objects that it would obscure to be viewed through it.

Animated hints are similar to ghosted hints, but replace the ghosted image with an animated representation of the movement trajectory. Tversky et al. suggest that animation can improve learning under certain conditions [16].

Composite hints integrate multiple simpler hints. In a later section, we discuss the implications for showing a set of possible actions and results instead of a single possible action.

3.3 Activation and Deactivation

To avoid visual overload, we do not show visual hints all the time, and, we limit those that are shown when visual hints are enabled. Thus, there must be ways to activate and deactivate hints (e.g., through an implicit or explicit user action), and to determine which relevant hints to display. Activation and deactivation can be accomplished using tangible AR interaction, or through an additional modality, such as voice. We implemented a variety of activation methods, which we describe here.

Pausing or lack of motion. Inspired by the use of pausing to enable marking menu display, this can act as either an implicit or explicit activation technique. Pausing works well when it does not normally occur as part of the action, as in skilled selection from a menu, but can be problematic when it is part of the action, as is the case in our system: Users often hold the virtual voucher still, so they can study the veins or edges of a leaf, and this could inadvertently trigger display of visual hints when it isn't required.

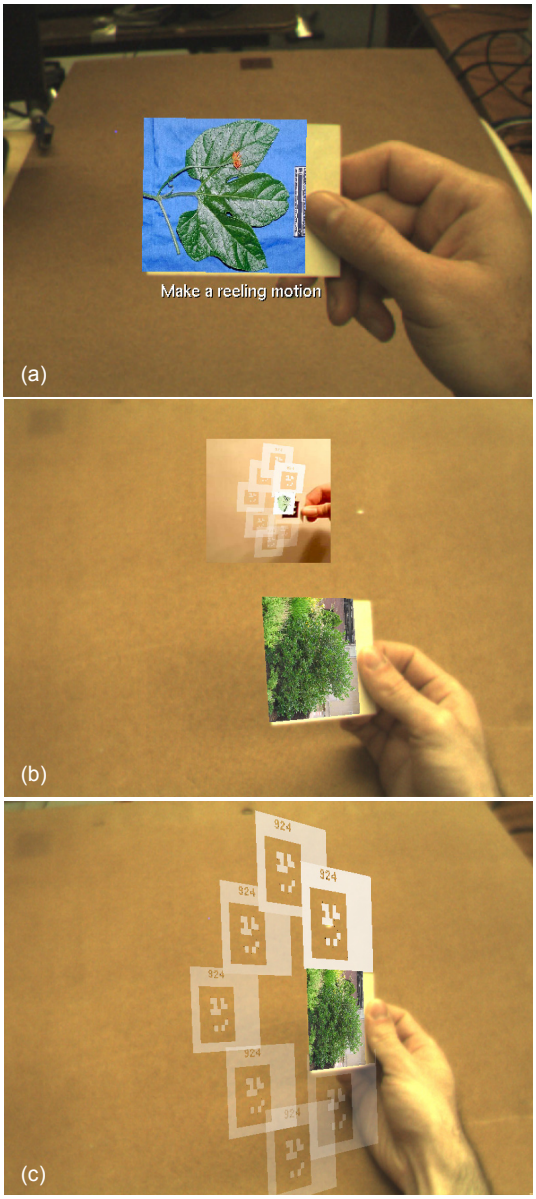

Fig. 2: A visual hint for a circular "reeling" gesture, represented through (a) text, (b) diagram, and (c) ghosting. (Viewed through a video see-through display.)

Combining pausing with a specific orientation and distance from the user can alleviate some inadvertent activation.

An activation gesture. For example, shaking the virtual voucher in or out of plane [7] suggests an attempt to discern its hidden properties, as in shaking a gift box to hear what is inside.

A key or button press. This conventional activation approach creates a more clearly defined visual hints mode. However, pressing a button outside the tangible AR space of interaction can bring the user away from the task at hand.

We considered, but did not implement, other methods, such as proximity of the hand and voice activation. Proximity of the hand to a tangible AR object can be used to activate visual hints, but our experience from mock-ups is that once the gesture has been shown, the hint will remain present even if no longer necessary.

Visual hints can also reflect completion of an action. For example, as a user makes the correct circular reeling motion following the path of ghosted images, each ghosted image can be removed.

4 Implementation

We implemented the methods described in the previous sections in our tangible AR EFG prototype [17]. Our system uses a Sony LDI-D100B 800×600 resolution, color, head-worn display, on which we mounted a Point Grey FireFly MV camera to capture the scene for 6DOF fiducial tracking and biocular (non-stereo) video see-through AR. In the mobile system, the display and camera are connected to a Sony U750 hand-held computer running Windows XP, mounted on a fanny pack worn over the shoulder. For our stationary user study, the camera and display were connected to an Apple Macbook with 1GB RAM and a 2Ghz Intel Core Duo CPU. A clipboard and individual rigid paper cards mounted with fiducials provide tangible objects for interaction.

The software is developed in C++, with OpenGL graphics, ARTag [5] 6DOF optical fiducial tracking, and grammar-based gesture recognition. Individual movements such as left, right, up, down, forward, back, and rotation are continuously captured, with those movements below a calibrated threshold omitted to avoid jitter artifacts. Movements are combined in a history that is parsed and fit to state machines representing each of the gestures. If a gesture is matched within a threshold time limit, the gesture is accepted and added into a gesture history. The test system runs at 24 frames per second for both capture and rendering. (Gesture recognition is less accurate at lower frame rates.)

5 Pilot Study

To gain a better understanding of how to represent and activate visual hints, we ran a pilot user study. Seven participants (three male, four female), 19–34 years old, were recruited from our university population through e-mail lists and fliers. Each had prior familiarity with computers and was compensated $10.

The task was to understand, learn, and perform the correct gesture with a fiducial card, given specific stimuli for the gesture. An automated test suite was used to present seven different individual and composite visual hints to the subject for each of three gestures, for 21 total combinations in a within-subject design. Order of presentation was randomized. The seven visual hint types were text (T), diagram (D), ghost (G), animation (A), ghost+animation (GA), ghost+text (GT), and ghost+text+animation (GTA). Diagrams were screen-anchored. All other conditions were object-anchored. Text and diagrams always faced the user. The three gestures were reeling, twirling, and movement into a target area, each of which was used to cycle through 2D images of a plant species presented as if printed on the fiducial card.

Subjects sat at a table and were videorecorded to capture both their gestures and the images they viewed. Prior to the study, the experimenter explained the task to the participants and subjects were given a preliminary trial prior to the actual trial. Subjects wore the head-worn display and held a fiducial in their hand. After each visual hint was presented, the subject was asked to identify and perform the gesture, and completion time was logged. At the end of the trial, subjects were introduced to shaking and pausing activation methods for visual hints. After completing the automated tests, the subject was asked to fill out a questionnaire to provide ranked preferences, Likert-scale responses, and qualitative comments on the visual hint and activation approaches.

5.1 Results and Discussion

One important observation about our within-subject pilot study is that there were significant learning effects. Once a user knows the gesture, they can apply that knowledge to other hints about the same gesture. This implies that time to completion analysis is only valid for the first instance of a gesture being observed and correctly completed. However, the qualitative results from observations of the subjects and questionnaire results are useful.

Subjects were asked to rank the seven techniques in order of preference (Fig. 3). Looking at a chi-square for all seven categories, $X2=16.8$, $df=6$, $p=0.01$, showing that the distribution of scores between the different options is significant. GA and GT ranked the highest, followed by GTA. This is mirrored in results of ranked comprehension, also in Fig. 3. We noted a variety of issues during observation and from participant comments, which we describe below.

Text can indeed be ambiguous because of language or culture. As one subject wrote, "I don't know what a 'reeling' motion is." One participant also interpreted twirling as a single motion, moving back and forth, rather than a continuous motion. Even words such as "target" and "left" proved ambiguous. We also found that object-anchored text hints could drop below the edge of the screen when the fiducial was held towards the bottom of the viewable image (which could be fixed by adaptive layout). One subject commented that they would prefer text for known gestures and animation with ghosting for unknown gestures.

Diagrams were less culturally linked, but were misinterpreted by some subjects as showing 2D motion in a plane. This is most likely because of the nature of our diagrams, which were displayed on 2D surfaces. It is possible that different diagrams or stereo imagery would have produced better results.

Ghosting proved to be successful at illustrating the required movement. However, the subjects raised two important issues. First, the ghosted image is anchored to the fiducial, so as subjects tried to perform the gesture, the ghost moved with the fiducial, much like a dog chasing its tail. This could be resolved by keeping the ghost fixed after a test gesture is started or by providing completion feedback. A second issue was directionality. A series of images that vary in transparency can be interpreted in two op-

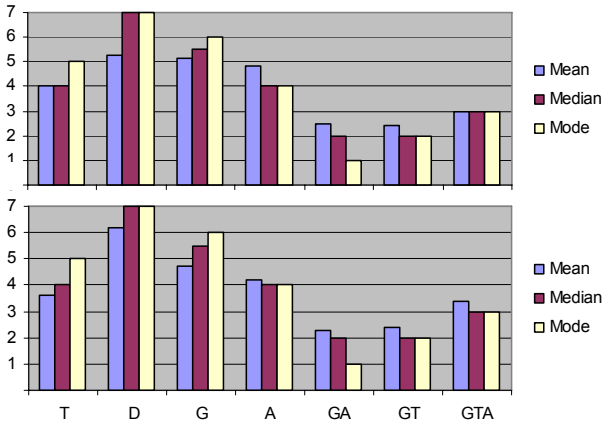

Fig. 3: Ranked preference (top) and comprehension (bottom) for each representation technique. 1 is best.

posing ways. In comics [11], motion from the past to the present is often shown as a series of images that transform from transparent to opaque, such as those of a fist moving through the air to a punch. The past is transparent, and the present is opaque; that is, time moves forward from the transparent to the opaque. However, when we represent motion from the present to the future, we can either emphasize the forward movement of time and interpolate from transparent to opaque or keep the present as opaque, interpolating opaque to transparent. This ambiguity can cause confusion, although the trajectory of the gesture is clear. (If directionality matters, an arrow can be added, creating a composite hint.)

Animation by itself resolves the issue of directionality, but could also be confusing because the subject is forced to keep the trajectory modeled in their mind. Furthermore, while animation is useful in clarifying the movement of a gesture over time, the speed could be taken either literally or as an abstraction of movement. One subject treated the animation speed literally, trying to match the rate of the animated visual hint, even though the gesture could potentially be made faster.

The combinations of ghosting with text or animation proved to be the most preferred. With animation, the subject saw both the directionality and the trajectory of the gesture; however, the additional movement could be distracting. With text, a potential benefit arose from reinforcing a particular name or label for a gesture.

In their responses to shaking versus pausing for activation, four participants preferred shaking, while three preferred pausing, with one commenting that "it's easier to shake than pause."

Note that the graphic representation used in our visual hints is an image of the fiducial, not the plant. In our experience, the fiducial was less visually complex in terms of texture and color, as compared to the plant, and thus less distracting. Use of the plant image also created confusion between the hint and the virtual object. However, use of other imagery is worth further study.

5.1.1 System Discussion

In implementing and using different types of visual hints and activation mechanisms, we made several observations. Our visual hints primarily represent simple gestures that require a single motion or trajectory to complete. Visual hints for complex or compound gestures might still run into limitations given the necessary reuse of visual space around a tangible object. In ghosting and animated visual hints, we found that the space around the fiducial could support only a single gesture (because of the relatively large spatial extent of the hint and small field of view of our display). In contrast, diagrammatic and textual representations support simultaneous display of hints for multiple gestures, by placing multiple diagrams or text strings adjacent to each other.

One might ask why not design the tangible object itself to have visually obvious affordances, so that the set of gestures and manipulations are clear and the physical morphology constrains the user to these gestures? While this is a noble goal, objects may change functionality depending on their context. In a sense, visual hints acknowledge the object's affordances, as defined by Gibson [6], making them more readily perceived. That is, an object can be manipulated in many different ways, but only certain actions have meaning in the context of the AR system.

We also found that while our focus was on the proper way to execute and complete the gesture, it was important to reflect the actual function of the gesture, documenting its consequences.

6 Conclusions and Future Work

We have presented visual hints for discovering, learning, and assisting with the completion of gestural interactions in tangible AR systems. We investigated the space of representations and activation mechanisms through prototyping, implementation, user testing in a pilot study, and reflection on that use.

While we have focused on tangible AR, the concept of visual hints applies to manual gestures in general, which have a steep learning curve [12]. We would like to expand this work to explore common activation techniques across AR systems that would reveal visual hints about environments, as well as gestural UIs. Results from our pilot study suggest that the activation techniques could be broadly applicable. We also believe that authoring visual hints could be simplified and automated through learning by example. Easy authoring could make visual hints more readily available across systems.

Finally, we would like to incorporate real-time analysis of the correctness of more subtle gestures. Sufficiently good tracking can provide data for recognizing subtle gestures and providing feedback on how well they are executed.

7 Acknowledgements

We thank Jason Kopylec and Randall Li for their help developing the tangible AR UI to the EFG; and Chris Smith, Hrvoje Benko, Eddie Ishak, Steve Henderson, Ohan Oda, Charles Macanka, Lauren Wilcox, and Anette von Kapri for discussion and feedback. This work was funded in part by NSF Grant IIS-03-25867 and gifts from Microsoft Research and NVIDIA.

References

[1] Agrawala, M., Phan, D., Heiser, J., Haymaker, J., Klingner, J., Hanrahan, P., and Tversky, B., "Designing effective step-by-step assembly instructions," *ACM Trans. Graph.*, vol. 22, 3, 828-837, 2003.

[2] Feiner, S., "APEX: An Experiment in the Automated Creation of Pictorial Explanations," *IEEE Comp. Graphics and Applic.,* vol. 5, 11, 29-37, 1985.

[3] Feiner, S., Macintyre, B., and Seligmann, D., "Knowledge-based augmented reality," *Communic. ACM*, vol. 36, 7, 53-62, 1993.

[4] Feiner, S.K. and Seligmann, D.D., "Cutaways and ghosting: satisfying visibility constraints in dynamic 3D illustrations," *The Visual Computer*, vol. 8, 5, 292-302, 1992.

[5] Fiala, M., "ARTag, a fiducial marker system using digital techniques," *Proc. CVPR 2005*, San Diego, CA, 2005, 590-596.

[6] Gibson, J.J., *The Ecological Approach to Visual Perception*: Lawrence Erlbaum Associates, 1986.

[7] Kato, H., Billinghurst, M., Poupyrev, I., Imamoto, K., and Tachibana, K., "Virtual object manipulation on a table-top AR environment," *Proc. IEEE ISAR*, Munich, Germany, 2000, 111-119.

[8] Kjeldsen, R. and Kender, J., "Toward the use of gesture in traditional user interfaces," *Proc. Automatic Face and Gesture Recognition*, Killington, VT, 1996, 151-156.

[9] Kurtenbach, G., Sellen, A., and Buxton, W., "An empirical evaluation of some articulatory and cognitive aspects of marking menus," *Human Computer Interaction*, vol. 8, 1-23, 1993.

[10] Mayer, R.E., *Multimedia Learning*: Cambridge Univ. Press, 2001.

[11] McCloud, S., *Understanding Comics: The Invisible Art*: Harper Paperbacks, 1994.

[12] Quek, F., McNeill, D., Bryll, R., Duncan, S., Ma, X., Kirbas, C., McCullough, K.E., and Ansari, R., "Multimodal Human Discourse: Gesture and Speech," *ACM Trans. Comput.-Hum. Interact.*, vol. 9, 3, 171-193, 2002.

[13] Ramos, G. and Balakrishnan, R., "Pressure Marks," *Proc. CHI*, San Jose, CA, April 28 - May 3, 2007, 1375-1384.

[14] Shneiderman, B., *Designing the User Interface: Strategies for Effective Human-Computer Interaction,* Addison Wesley, 2004.

[15] Tan, D.S., Poupyrev, I., Billinghurst, M., Kato, H., Regenbrecht, H., and Tetsutani, N., "On-demand, In-place Help for Augmented Reality Environments," *Proc. Ubicomp*, Atlanta, GA, 2001.

[16] Tversky, B., Morrison, J.B., and Betrancourt, M., "Animation: can it facilitate?," in *Int. J. Human-Comp. Studies*, vol. 57, 2002, 247-262.

[17] White, S., Feiner, S., and Kopylec, J., "Virtual Vouchers: Prototyping a Mobile Augmented Reality User Interface for Botanical Species Identification," *Proc. 3DUI*, Alexandria, VA, March 25-26, 2006, 119-126.