Exploring User Experience in "Blended Reality": Moving Interactions Out of the Screen

David F. Huynh
MIT CSAIL, the Stata Center
32 Vassar St., 32-350
Cambridge, MA 02139, USA
dfhuynh@mit.edu

Yan Xu
Renmin University of China
Zhixing 1, 1203#
Beijing 100872, P.R.China
vivi.xy@gmail.com

Shuo Wang
Microsoft Research Asia
5F, Beijing Sigma Center
No.49, Zhichun Road
Beijing 100080, P. R. China
shuowang@microsoft.com

CHI 2006, April 22–27, 2006, Montréal, Québec, Canada.
ACM 1-59593-298-4/06/0004.

Abstract
Video game players often learn to map their physical actions (e.g., pressing buttons) onto their on-screen avatars' actions (e.g., wielding swords) in order to play. We explored the experience resulted from eliminating this mapping by modeling the screen as a "window" through which virtual objects enter the player's physical space, and the player interacts with them directly without the mediation of an avatar. We define this interaction as "Blended Reality" (BR). We designed, developed, and evaluated a BR game prototype called "Apple Yard" in which the player was to use a wand to hit apples flying out of the screen. A camera was used to track the positions of the player's eyes and wand, and the 3D game scene was rendered accordingly to create the illusion of looking through a window. A user experiment conducted on this prototype indicated BR's potential in camera-based entertainment.

Keywords
Blended Reality, augmented reality, physical interaction, camera-based game.

ACM Classification Keywords
H.5.2 [User Interfaces]: Interaction styles, Prototyping, Evaluation/methodology.

Introduction
While physical interaction in video games and software user interfaces is being promoted by ubiquitous computing

and tangible user interfaces, much of such interaction still requires cognitive mapping from the user's physical movements to those of an on-screen avatar or a mirrored image. In QuiQui's Giant Bounce game, children's gestures were amplified, extrapolated, or re-interpreted when applied on the virtual dragon [5]. Such translations from the player's movements to the avatar's actions depend on the limited number of behavior patterns that the system has previously learned. In the arcade shooting game that support IR guns, the player only has to aim at the pixels on the screen that represent a target. The depth information of the target is ignored and, hence, the player is shooting at targets on a 2D plane rather than in a 3D world resembling physical reality.

Although these mappings have their merits, we wish to explore more direct and natural physical interactions that require no mapping. In this paper, we demonstrate a new concept called Blended Reality (BR) through a camera-based game prototype called Apple Yard (Figure 1). In this game, the user is led to perceive the display screen as a window through which virtual game objects can cross. The player uses a real wand to hit virtual apples that "fly out of" the screen into the player's physical space. User experiment results on Apple Yard show that BR systems can deliver rich physical experiences and more preferred interactions.

Blended Reality

Blended Reality is the realm where the real and virtual worlds blend together as one space, letting users and real objects interact with virtual objects in a direct and physically natural manner.

In Milgram's "virtuality continuum" spectrum [6], Blended Reality is closer to Augmented Reality (AR) than Virtual Environment (VE). While users in VE [2] are almost disconnected from the physical world that they inhabit, users of both AR and BR remain very much connected to the real world.

At the present, most AR research involves processing live video feeds and augmenting them with computer-generated graphics. The physical world remains dominant in the AR experience. In contrast, BR presents the virtual world on screen all by itself and thus keeps the two worlds mostly apart and neither dominant over the other. BR only processes sensory data from a subspace of the physical world in order to map the user's actions into the virtual world, thus making the two worlds seemingly blend into one another in this subspace. Finally, while AR and VE usually require on-body displays and sensors, BR strives for non-intrusive interaction.

Prototype Design: Apple Yard—a Camera-based Blended Reality Game System

In "Apple Yard" (Figure 1), virtual apples in a virtual yard "fly out" of the screen and the player is to hit them with a hand-held wand. Because we do not render the apples

Figure 1. "Apple Yard" and its elements

after they enter the physical space, the player must guess their trajectories. When an apple is hit, it bounces back and is rendered as an apple core after crossing into the virtual world behind the screen.

"Apple Yard" differs from AR systems (such as TouchSpace, which recaptures human touch and physical interaction as essential elements of game play [2], and AR Bowling, which enhances game realism by an integrated real-time kinematics multi-body system simulation [4]) through the following unique features:

- The player directly interacts with virtual objects in the physical world;
- The display screen is rendered as the "window" that connects the physical and virtual world;
- The interaction is less cumbersome due to the use of computer vision technologies [7] rather than on-body sensors and displays. However, since our computer vision technologies were still lacking, the player was still required to wear a position marker on his/her forehead for more accurate head tracking in our current prototype.

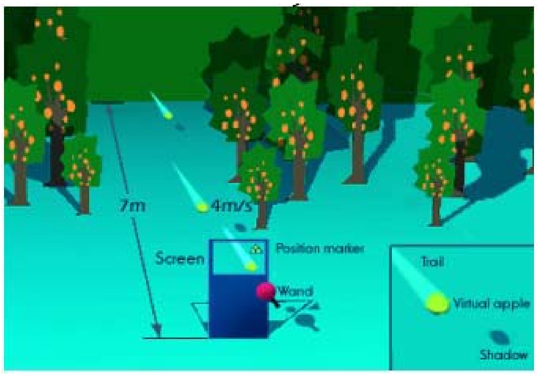

Figure 2. 3D model of Apple Yard

Implementation

"Apple Yard" modeled the virtual world on a 1:1 scale compared to the real world. That is, when sensory data on the physical objects (e.g., the wand) is brought into the virtual world model to analyze for interaction (e.g., collision with apples), dimensions in the real world and the virtual world are unified to the same unit: 1 meter in the real world is modeled as 1 "meter" in the virtual world. This 1:1 scale "aligns" the two worlds, creating the illusion of a blended subspace.

We used Microsoft DirectShow to capture images from a wide-angled CCD security camera with 320 x 240 pixels resolution, 24-bit color depth, and Direct3D to render the virtual world on screen.

- 3D game world

Figure 2 shows the 3D virtual world model, which included purely virtual elements (the apple yard and flying apples) and modeled real elements (the display screen, the wand, and the position marker).

- 3D object tracking

We tracked the position marker worn on the player's forehead to approximate the 3D position of the player's eyes. The marker was a uniform-colored triangle with three small circles of another color at the corners. The center of the triangle was located by hue analysis. Then, the centers of the three circles were located by k-means clustering algorithm. Finally, we used the largest pixel distance between the circles to infer the z-displacement of the marker from the tracking camera.

The wand was a uniform-colored sphere on a handle. The center of the sphere was located by hue analysis. Then the width of the hue blob was used to derive the sphere's z-displacement from the camera.

We adopted simple image analysis strategies to reduce computational costs. Our hue-based detection method traded off accuracy for speed. We also applied a low-pass filter to dampen jitter at the cost of some lag.

- 3D rendering

To give the player the illusion of looking through a window, we positioned the Direct3D scene rendering camera exactly at the coordinates of the player's eyes (tracked and then mapped into the virtual world model). As the player moved around, the rendering perspective shifted naturally. This matched the player's everyday experiences of looking through real windows while requiring no use of additional controls.

To project the virtual world onto the screen from the location of the player's eyes, we first formulated a standard perspective projection matrix such that the viewing frustum's front face covered the same area as the screen. Then, we shifted the viewpoint by multiplying this matrix with a matrix that performed a translation by the displacement between the user's eyes and the z-axis through the center of the screen. The direction of the user's gaze did not (and should not) influence this rendering.

User Experiment

We wished to determine: (1) how players perform in games of "interacting with virtual objects in the physical world"; (2) how the rendering camera perspective influences user experience; and (3) whether users can readily accept interactions with invisible virtual objects in the physical world.

Participants

18 volunteers (1 female, 17 males) from our lab participated in this experiment. They ranged in ages from 22 to 28 (average 25) and were all right-handed. None had any prior experience with the system.

Experiment Design

We tested two factors:

- *First/third-person's view*: whether the player viewed the virtual world in the first person's view or in the third person's view.
- *View-dependent/independent*: when in first-person's view, whether the virtual world was projected on the screen according to the player's eye position or not.

Of the four combinations of these two factors, the view-dependent + third-person's view combination was nonsensical. The other three modes we tested included:

(a) **IT:** view-**I**ndependent & **T**hird-person, in which the player was shown the whole virtual world, including the virtual models of the display screen and the wand, from a static perspective;

(b) **IF:** view-**I**ndependent and **F**irst-person;

(c) **DF:** view-**D**ependent and **F**irst-person.

The test followed a within-subjects design. Each of the participants finished three tasks, one task for each of the three modes (Figure 3). The order effect was counterbalanced by a 3 x 3 Latin square.

Quantitative measures were: accuracy of hitting apples and wand scope. The 3D coordinates of the wand were logged when it hit a virtual object. The scope of the wand was (Max(X)-Min(X), Max(Y)-Min(Y), Max(Z)-Min(Z)). Accuracy was the number of apples hit by the player's wand divided by the total number of apples flying out during the task. We also gathered qualitative information from a questionnaire, through interviews and observations, to analyze and understand the subjects' behaviors.

Figure 3. Experiment setup of Apple Yard. (a) **IT**, (b) **IF**, (c) **DF**. Camera (1), Plasma display (112 x 63 cm^2) (2), Wand (3), Position marker (4), and participant (1.5 to 2 m from screen) (5).

Results

Three major results are reported below and qualitative reasons are presented:

(1) The majority of the participants understood and accepted the concept of "interacting with virtual object in physical world."

Through the interview, four of the users (22.22%) said "I understood the interaction happened in the real world space immediately I started playing the game"; 10 of them (55.56%) said that "I accepted this concept when learning to play"; four (22.22%) did NOT accept this concept at all.

In interviews, the large display (112cm x 63cm) and the high speed of flying apples (4 m/s) were two frequently mentioned factors that enhanced immersion and helped to convince people of the BR concept.

(2) There is no correlation between the measured accuracy and the subjective preference ranking of the three modes.

Eight (44.44%), Five (22.78%), Five (22.78%) participants chose DF, IF, and IT respectively as their most preferred mode. There was no correlation between their rankings and their measured accuracies.

Our interviews also found that there were three major reasons why people preferred DF to IF: first, it was more similar to real-life experience. Second, players enjoyed enriched interactivity brought by changing perspectives. Finally, players could more easily estimate trajectories of the virtual apples by naturally turning to better observation angles. Participants who preferred the IT mode more than the other two modes claimed that they preferred easier game play and would like the whole process of interaction visible by seeing the wand's avatar in the virtual world.

According to our ANOVA analysis, the factor of first/third-person's view significantly influenced the accuracy ($F=14.168$, $p<0.001$, Accuracy: IT=52.63%, IF=37.23%, DF=29.84%). This significant difference in accuracy was expected because in IF and DF, users needed to guess the trajectories after the virtual apples exited the screen and became invisible. Two major reasons from the subjects' complaints about IF and DF were: (1) estimations of trajectories were different from actuality, and (2) no feedback was given when the player missed an apple.

(3) The players' physical movements were more intense and diverse in DF than IT.

3D movements (measured by wand scopes) were significantly larger in DF than in IT (paired-samples T-test, two tailed x: $t=4.832$, $p<0.001$; y: $t=6.225$, $p<0.001$; z: $t=3.576$, $p=0.002$). Also, it was observed that the players' experience was much more strenuous (perspiration) in DF than in IT.

Finally, in IT, participants exhibited more calm and deliberate gestures, bordering on boredom. In contrast, in

DF, they showed more playfulness, including jumping aside to hit, switching the wand from hand to hand, etc. They also differed in their leverage of the adaptive perspective rendering: some tried to front-face the approaching apples while some preferred bigger deflective angles.

Conclusions

As found in the user experiment, the rendering camera in 3D game world does significantly influence a player's experience. While the third-person view assures higher accuracy, first-person view could encourage more intense and diverse movements.

But the factor of view dependent/independent rendering did not show significant advantages in either accuracy or movement scope. From the interviews, we found it was the system's performance (jittery and delay) that harmed the user experience of the view-dependent feature.

There are three major problems discovered in the current prototype: (1) system performance: jittery and lag; (2) lack of feedback when the player misses a target; (3) lack of true stereoscopic vision needed to more accurately estimate trajectories.

Summary and Future work

Besides solving the problems found in the user experiment, we plan to research: (1) how to adopt the concept of Blended Reality in multi-player video games and (2) how other parameters, such as screen size, feedback, and virtual objects' attributes, influence user experience in BR systems.

We hope that "Blended Reality" can inspire the game research and practice communities to incorporate more physical and natural interaction into video games. BR itself is one such exploration. As indicated by our user experiment, BR has great potential in camera-based entertainment for offering rich physical interactivity. Physical movements encouraged in BR can bring health benefits.

References

1. Microsoft DirectX, http://www.microsoft.com/windows/directx/.
2. Adrian David Cheok, Xubo Yang, Zhou Zhi Ying, Mark Billinghurst, Hirokazu Kato, Touch-Space: Mixed Reality Game Space Based on Ubiquitous, Tangible, and Social Computing, *Personal and Ubiquitous Computing*(2002), Volume 6 Issue 5-6.
3. Bimber, O., Wetzstein, G., Emmerling, A., and Nitschke, C., Enabling View-Dependent Stereoscopic Projection in Real Environments, *Proc. ISMAR'05.*
4. Carsten Matysczok, Rafael Radkowski and Jan Berssenbruegge, "AR-Bowling: Immersive and Realistic Game Play in Real Environments Using Augmented Reality", *ACM SIGCHI, Con. A CET 2004.*
5. Hämäläinen, P., and Höysniemi, J. A, Computer Vision and Hearing Based User Interface for a Computer Game for Children. *Pro. ERCIM 2002*, 23-25.
6. Milgram, P. and Kishino, F., A Taxonomy of Mixed Reality Visual Displays, IEICE Transactions on Information Systems, vol. E77-D (12), Dec. 1994.
7. Pattie Maes, Trevor Darrell, Bruce Blumberg, Alex Pentland, The ALIVE system: wireless, full-body interaction with autonomous agents, MIT Media Laboratory, 20 Ames Street, Cambridge, MA 02139, USA, Multimedia Systems (1997) 5: 105–112.