Deploying and Evaluating a Mixed Reality Mobile Treasure Hunt: Snap2Play

Yilun You[1], Tat Jun Chin[1], Joo Hwee Lim[1], Jean-Pierre Chevallet[1], Céline Coutrix[2] & Laurence Nigay[2]

[1] Image Perception, Access and Language Lab, Institute for Infocomm Research, Singapore

{ylyou,tjchin,joohwee,viscjp}@i2r.a-star.edu.sg

[2] Laboratoire d'Informatique de Grenoble

{celine.coutrix,laurence.nigay}@imag.fr

Figure 1. Snap2Play screenshots in sequence during a game play (Clue-Radar-Digital | Physical Card Collection)

ABSTRACT

With the current trend, we can anticipate that future mobile phones will have ever-increasing computational power and be able to embed several captors/effectors including cameras, GPS, orientation sensors, tactile surfaces and vibro-tactile display. Such powerful mobile platforms enable us to deploy mixed reality systems. Many studies on mobile mixed reality focus on games. In this paper, we describe the deployment and a user study of a mixed reality location-based mobile treasure hunt, Snap2Play[1], using technologies such as place recognition, accelerometers and GPS tracking for enhancing the interaction with the game and therefore the game playability. The game that we deployed and tested is running on an off-the-shelf camera phone.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces & Presentation**]: User Interfaces

General Terms

Algorithms, Human Factors

Keywords

Mobile Mixed Reality, Location Awareness, Place Recognition

1. INTRODUCTION

In parallel with game consoles, mobile gaming is catching up. One the key challenges for mobile phones is to enhance the playability of the game by designing innovative interaction modalities that will overcome the current limited interaction resources of mobile phones. Since mobile phone manufacturers

now provide phones that embed several captors/effectors including GPS, orientation sensors and tactile surfaces, there is a high potential for games with innovative interaction modalities based on such captors/effectors. In this prototype we will experiment with GPS & orientation sensors to begin with.

For the design of our game, we focus on augmented virtuality as defined in the continuum of Mixed Reality [4]. Our goal was to base the mobile game on physical playground augmented with computer-generated graphics. "Video see-through MR", as described in [8], enables a user to see the real world with embedded computer-generated graphics as if they are part of the real world. Yet interaction taking place in such augmented virtuality environments often requires complex setups and external wearable devices, such as a see-through head-mounted display (ST-HMD [9]), in order to define a real-time immersive environment. Moreover vision-based complex algorithms requiring high computational power are necessary. Such augmented virtuality techniques are therefore not ready to be fully operational on a mobile phone. In Snap2Play[1], we have engineered a modular mobile game that implements various simple techniques which will scale easily when the mobile devices are powerful enough to implement the complex algorithms. The game designed using Snap2Play is inspired from the card game "Memory", asking the player to match a pair of identical cards. Organized along a trail, like treasure hunting, the Snap2Play player is asked to match a "digital card" with a "physical card" by taking pictures using the phone camera. An example of the game flow is presented in Figure 1. In this paper, after briefly explaining the game design and in particular how we augment the physical scenes with digital objects, we focus on the deployment of the game and the user study that we performed both to test the feasibility of the implementation on the treasure hunt game and to study how players interact with the mixed reality world using portable handsets with limited screen space.

2. RELATED WORK

For mixed reality, vision-based techniques are common such as location tracking [2], digital objects placement onto the real world and handling of self-occlusions [3]. Mobile phones with added

sensors (e.g., GPS, accelerometers) are the minority of the market currently. In [6] a mobile game based on detecting and matching markers, the markers are positioned against the camera viewpoint in order to place the corresponding digital object on the camera view plane with matching coordinates and estimated magnification. The drawback is the deployment of the game and its scalability, since markers must be installed on site. In our method, we need a solution to that and a fast algorithm without burdening the processing power on the mobile phones.

Mediascape[7], in short mscape, is an authoring prototype technology from Hewlett Packard (HP) Lab to design location-based games anywhere. This technology allows designers to blend audio, images and GPS data into the storyboard of a life-size game on HP PDA phones. Mscape is in many ways similar to Snap2Play; mobile, games and the usage of context data. Currently, Mscape platform uses context data (eg. GPS, IR, RF beacons, etc) to augment the virtual reality as demonstrated by Reid at al from their project "Scape-the-Hood" [5]. At the moment, Mscape concentrates on providing a flexible platform for user's creation of games and we planned to experiment with augmented virtuality interaction using computer vision and orientation sensing on a compact mobile device.

3. IMPLEMENTATION
3.1 System Design
The underlying architecture of Snap2Play relies on a storyline with different interaction modalities and on a tracking mechanism of the player. We are currently using GPS for tracking the player but other technologies including place recognition can be integrated in our architecture. In Figure 2, we number the steps of the storyline for the developed memory game. For collecting the cards, Snap2Play supports two modes of mixed reality interaction (steps 4 and 7 in Figure 2). Along the trail, the cards can be collected by taking a picture of (1) a digital card located in the real world (2) a physical place (i.e., physical card) in the real world, using respectively orientation sensing and place recognition techniques to identify the cards.
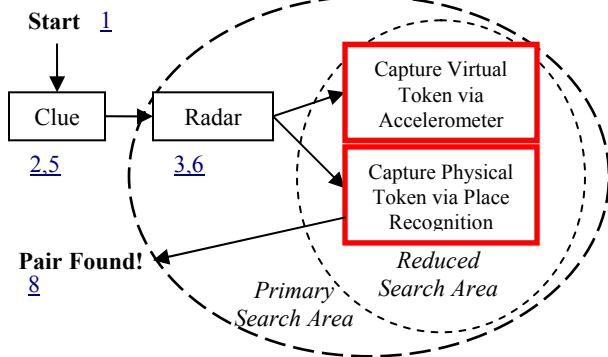

Figure 2. Memory Game in a Trail: Story corresponds to Figure 1

3.2 Virtual Object Vertical Placement
For the case of digital cards, the player can see a superimposed digital object through the video feed from the mobile phone camera lens. For positioning of the digital object, we based the technique on how mobile users are looking at the screen. The mobile phone is usually held at least half an arm length from the eyes. Moreover users usually move their arm along a large

curvature path at the height of the upper body as shown in Figure 3-a. Our technique allows us to see the digital card on screen at different angles on the vertical axis relative to the ground.
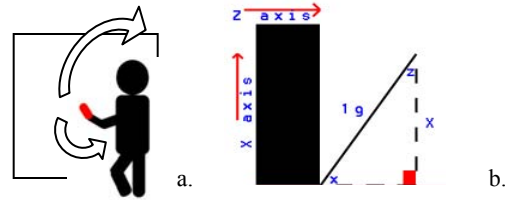

Figure 3. Vertical placement of a digital object

We calculate the position of the digital object according to the position of the mobile phone along this curvature path. To do so, we consider different tilts and angles along the vertical axis (X axis in Figure 3-b) and the gravitational force on the device measured along both X and Z axes, assuming that the acceleration (Y axis, parallel to the ground) is null. For capturing the required sensed data (tri-axis accelerometer reading), we used the SHAKE (Sensing Hardware Accessory for Kinaesthetic Expression) [1] device. We then calculate the offset value of the current position against the target position and paint the card only when it falls within the displayed region of the video stream.

3.3 Game Design: Storyboard
We have designed the game to be played outdoors, when the weather is fine and GPS readings available. The Memory trail is created on a site in the National University Singapore (NUS) campus, where traffic is sparse enough to play the game safely. Players have to find three pairs of cards sequentially positioned as shown on the map in Figure 4, starting with a digital card followed by a physical card and so on.
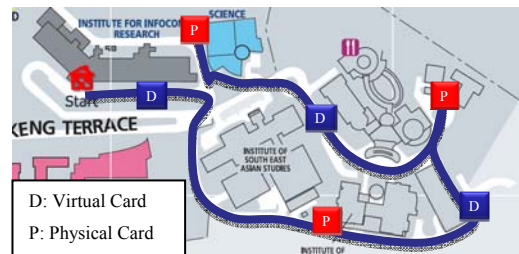

Figure 4. Map of Trail in NUS Campus

At the start the players will be given a map, Nokia N80 attached with the SHAKE device and a GPS receiver. Upon starting the game on the mobile phone, the player is to select a trail to play in. The trail begins with a clue leading to the location of the 1st card (virtual) in the 1st pair. Players are to take a picture of virtual card (Treasure box in Figure 2.4) to get visual clue for the next card (Physical). Players are to collect the physical card to form a matched pair, and they can be any of the places in Figure 5.

Figure 5. Examples of Places in the Trail as Physical Cards

4. EVALUATION METHODOLOGY

The objective of the conducted evaluation is twofold:

- Place Recognition performance. We would like to evaluate the time and recognition performance of our state-of-art and how it compares to users' expectation.
- Study the interaction within the mixed reality environment. Is it possible for players to differentiate underlying technologies used when a picture is taken (collecting physical & digital cards)? Do the players find it intuitive and natural to take pictures with the different interaction modes?

We have planned a 3-step approach to perform an effective user study that answers the questions while trying to avoid the technical issues that might affect the results of the study. Section 4 outlines the activities performed during each phase. The entire evaluation requires about 4 man-months.

4.1 Pilot Test

Before conducting any evaluation, a pilot study is performed to ensure that the application is running correctly without errors that will affect the player's experience. A version of the game is generated with a mock storyboard using GPS coordinates of the card locations and recognition models of the physical cards. This version is to be used on the mock site with the corresponding card locations pre-recorded. The purpose of this short test is to verify that there is no logical error in the game storyboard and application. It was also an effective method for us to learn how to collect the relevant GPS coordinates and pictures of places for training and testing the outdoor game scenario. (E.g. what makes a good card location for reliable GPS readings and the types of scenes that define good recognition models?)

4.2 Alpha Evaluation

The purpose of this phase is to prepare the site, storyboard, survey and interview for the actual evaluation. This step can be considered as the preparation phase of the evaluation materials. A set of players are invited to play the version of the game at this stage. Their goals consist of helping us evaluate the quality of the evaluation setup and reduce the unforeseen problems that could arise during the actual experiment. Furthermore, we can check the quality of the collected to ensure the data collected later in the beta phase will be as clean and accurate as possible.

For creating the storyboard and therefore selecting the digital and physical cards, the site survey is based on the lessons learnt during the pilot study: indeed the selection of the cards has a strong impact on the difficulty level of the game. During this step, we also designed the clues for linking the cards along the designed trail. The clues are to be phrased clearly and easily understandable by players. We also took into consideration the terrain and distances between cards. Different combinations of card pairs and terrain challenges can be designed on the same site to define different levels of difficulty of the game. This flexibility in the storyboard is also one of the reasons why we designed the game by combining a classic Memory game with a treasure hunting game. However during the beta step of the evaluation, we only use one Memory trail. This is to eliminate the clues and terrain challenge as a factor in the players' experience.

Upon completion of the evaluation design, we conducted alpha evaluation with 5 players having the following characteristics:

- **Age Group**: 25 to 40
- **Type**: Working Adults
- **Knowledge of Map**: A mixed of new & familiar
- **Weather**: fine-cloudless, fine-cloudy, cloudy
- **Time of Day**: 2pm to 4.30pm (GMT+8)
- **Mobile Phone Expertise**: Basic to Average

Two sources of data were collected. The first source is from the system logs that contain accurate recorded data including GPS trail, time and recognition results. The second source of data is from the questionnaires and interviews with the players. An evaluator follows the player throughout the game, making observations of where the players entered card proximity areas and interviewing the player during the game. This is to verify that the instructions are clear or to note any issues and difficulties during the game. After each play, the player went through an interview based on a survey questionnaire to be filled and commented with the evaluator. We then reviewed the collected data to be sure that important data are not missing. At the end of the evaluation, we refine the setup and materials of the experiment according to players' feedback and observations.

4.3 Beta Evaluation

Using the improved evaluation materials we began the evaluation on 30 players with the following characteristics:

- **Age Group**: 14 to 33
- **Type**: Students, Working adults
- **Knowledge of Map**: All are new to the map
- **Weather**: drizzling, fine-cloudless, fine-cloudy, cloudy
- **Time of Day**: 11am to 5pm (GMT+8)
- **Mobile Phone Expertise**: Basic to Average
- **Total No. of Teams & Individual Players**: 8T & 16 I
- **No. of Players/Team**: 2 maximum

The general comments from alpha evaluation include that the game might be more entertaining if the players are playing with friends. Thus, we have decided to consider teams of players as we would like to observe if there are differences in the performances and entertainment values between playing alone and playing by teams. To do so, we did not develop a new multi-player game instead they are playing this game together on the same set of mobile device setup.

5. RESULTS & DISCUSSION

The surveys and data logging were designed to collect statistics on the place recognition technology, mixed reality interaction and general game usability on the mobile device. 86% of the players found that the overall duration of the game was appropriate. In the category of the game environment based on safety, terrain difficulty, interesting storyboard and overall distance, players had respectively rated 56%, 73%, 100% and 23% as acceptable.

All players found the game interesting even though the distance was too long. 90% had agreed or strongly agreed that the game was very entertaining while others are reserving their judgment (neutral). However, only 50% of the players are keen to play again. When interviewing the other 50% of the negative players, they said that it would be too easy and boring if they had to play again in the same site but are positive to play again if the game is deployed in another site such as treasure hunting in the zoo. This confirms that players are fascinated by new stories and have no

problem playing again with the same mixed reality interaction techniques. This point reinforces our approach of developing a platform designed for easily defining different game storyboards, and it is as important as to extend for new interaction techniques.

Regarding the performance of our place recognition technique for collecting physical cards, 61% successful physical card queries were made against the total number of attempts. For all the "failed to recognize" results at the first attempt to match a card, players needed to make one more attempt for a successful match. In other words, the maximum number of attempts was 2. Comparing against users' expectation on the maximum acceptable attempts, 80% of the players would not mind if the recognition system fails once or twice for a match in a consecutive run. While 10% of the players were extremely relaxed about the expectation of a recognition system, they would not mind making 5 attempts for a match. Only 10% expected a perfect system. Based on this result, it would seem like the users are easily satisfied but in our opinion this result is a rough gauge to the limit of users' patience or tolerance to a system instead. The average number of attempts for the system to fail is 2; so by the 3rd attempt the player would give up using the system if he still gets a wrong result.
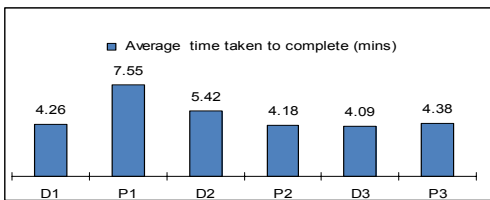

Figure 6. Average Time to Collect Physical & Digital Card

During the survey, 87% of the players found the number of attempts to recognize a match acceptable. 93% of the players found that the system is able to accurately recognize a card. Players are generally satisfied with the system, even when they needed to take 1.69 no. of attempts for the system to recognise the correct scene. This high acceptance rate we found is due to the task assigned: to take the picture as close as possible to the displayed clue. Therefore when the system failed the first time, they wouldn't expect the recognition system to work with photos taken at absurd angles and would willingly move and adjust to get the right angle such that the photo taken is as similar as possible.
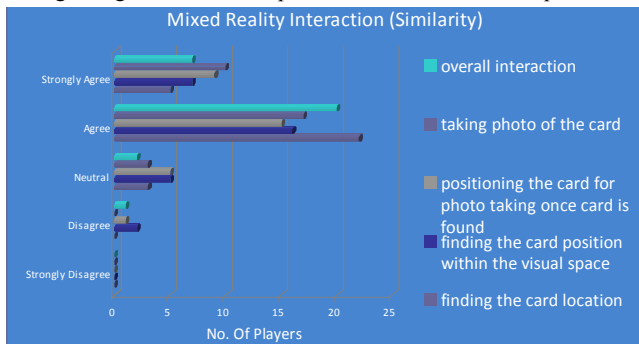

Figure 7. Survey on whether players found collecting digital and physical cards similar.

We can see how players handle the mixed reality interaction by observing them collecting digital and physical cards. Players need to connect to the GPRS/3G network for the first time during the collection of the first physical; it leads to a spike in the roundabout time as shown in Figure 6 at P1. On the average of 4-

7 mins, they managed to complete both tasks at a similar amount of time. Players were asked to evaluate their experiences on the cards collection. Most players found the two processes of taking a picture of a digital card and of a physical card similar, with only some minor differences as shown in Figure 7. The players found major difference when collecting physical and digital cards is the difference in finding the card when reaching the spot and positioning it for photo taking. As finding physical card is simply easier since the visual space available is much larger than digital card, which you need to see through the camera lens.

6. CONCLUSION

We have managed to complete the first prototype for a mixed reality authoring tool with different interaction modalities using sensors on mobile camera phones. This simple game had shown potential to enriching the players' experience with mixed reality techniques despite the limitation on mobile devices. It is therefore important to ensure that our platform is well designed so that it can scale up when the mobile phones will embed new sensors and be able to run mixed reality interaction modalities. The next step is to work to plug & play any storyline easily for experimentation.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] Chin,T.J.,You, Y.L, Coutrix, C., Lim, J.H., Chevallet, J-P. and Nigay, L., 2008 Snap2Play: A Mixed-Reality Game based on Scene Identification, 14th Int. Multimedia Modeling

[2] DiVerdi, S., Höllerer, T., 2007, GroundCam: A Tracking Modality for Mobile Mixed Reality, IEEE Virtual Reality, 75-82

[3] Gay-Bellile, V., Bartoli, A., Sayd , P., 2007, Deformable Surface Augmentation in Spite of Self-Occlusions, ISMAR

[4] Milgram, P., Takemura, H., Utsumi, A., Kishino, F., 1994, Augmented Reality: A Class of Displays on the Reality-Virtuality Continuum, SPIE Vol. 2351, Telemanipulator and Telepresence Technologies

[5] Reid, J., Ben Clayton, T.M., Hull, R., 2006, Scape the Hood: A design case study of a location based digital story mediascape, DIME-ARTS

[6] Sasaki, Y., Kamada, M., Yonekura, T., 2006, Implementation of Outdoor Mixed Reality Gaming on Mobile Devices, Int. Conference on Cyberworlds, 173-176

[7] Stenton, S. P., R. Hull, P. M. Goddi, J. E. Reid, B. J. Clayton, T. J. Melamed and S. Wee, 2007, Mediascapes: Context-Aware Multimedia Experiences, IEEE Multimedia, 98 - 105

[8] Tamura, H. and Yamamoto, H., 1998, Vision and Graphics in Producing Mixed Reality Worlds, IEEE Computer Vision for Virtual Reality Based Human Communications, 78-85

[9] Tamura, H., Yamamoto, H., katayama, A., 2001, Mixed reality: future dreams seen at the border between real and virtual worlds, IEEE Computer Graphics and Applications, 64-70