Second Surface: Multi-user Spatial Collaboration System based on Augmented Reality

Shunichi Kasahara
Sony Corporation
MIT Media Lab
Cambridge, MA, USA
shunichi.kasahara@jp.sony.com

Valentin Heun
MIT Media Lab
Cambridge, MA, USA
heun@mit.edu

Austin S. Lee
MIT Media Lab
Cambridge, MA, USA
aslee@media.mit.edu

Hiroshi Ishii
MIT Media Lab
Cambridge, MA, USA
ishii@media.mit.edu

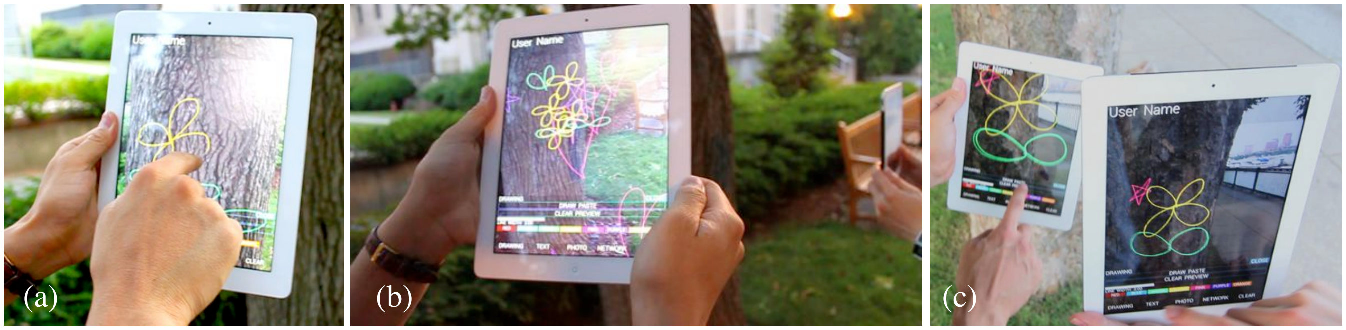

Figure 1. (a) *Create contents such as drawings on a mobile device*. (b) *Situate the content on the augmented reality spatial canvas*. (c) *Collaborative content creation in the shared spatial canvas with co-located users in real-time*.

1. Abstract

An environment for creative collaboration is significant for enhancing human communication and expressive activities, and many researchers have explored different collaborative spatial interaction technologies. However, most of these systems require special equipment and cannot adapt to everyday environment. We introduce *Second Surface*, a novel multi-user Augmented reality system that fosters a real-time interaction for user-generated contents on top of the physical environment. This interaction takes place in the physical surroundings of everyday objects such as trees or houses. Our system allows users to place three dimensional drawings, texts, and photos relative to such objects and share this expression with any other person who uses the same software at the same spot. Second Surface explores a vision that integrates collaborative virtual spaces into the physical space. Our system can provide an alternate reality that generates a playful and natural interaction in an everyday setup.

2. Introduction

It is in human nature to share and shape the world that is surrounding us. This is reflected in architecture, cars and parks as well as in art and media. All of these expressive artifacts adapt throughout long periods of time and almost never change instantly in a real-time creative way. Diverse sub cultures and ideas about the world coexist today and the modern culture is changing from a general consuming culture to a highly creative culture using online platforms such as YouTube, blogs and Facebook as tools for expression [Foad et al. 2012]. Such forms of expression have not been legally possible in the offline world before, as problems with graffiti illustrate very well. Illegal forms of expressions such as love mark engraved in a tree, a territorial gang sign on a building or street art itself should find their way into reality without being illegal.

We believe a virtual world system that creates a second surface on top of the reality, invisible to the naked eyes, could generate a real-time spatial canvas on which everyone could express themselves without the risk of being accused of vandalism. In designing such a canvas that can enhance meaningful spatial expressions, it is important that the system can be used in a collaborative real-time way and that the created content is connected to objects in the real world.

There have been many projects that introduce different techniques for collaborative spatial interaction. However, these systems rely on complex settings such as optical motion capture system or proprietary markers. This spatial constraint prevents the interactive system from becoming a flexible platform for intuitive spatial expressions.

We propose Second Surface which allows the user to create and situate contents on the augmented spatial canvas located in the surroundings of everyday objects, to share with co-located users in real-time collaboration. In our system, the pose estimation of the user's device allows the user to place the generated contents on the virtual space. Our system also provides natural interactive experiences for multiple users in the space. Each set of pose data of the user's devices and the generated contents are shared in real-time via the server. We implemented a mobile application and server application.

3. Related Work

Augmented Reality applications have been widely explored and applied in many domains [Krevelen et al. 2010]. More recently, AR applications, particularly based on mobile devices have become commercially available to the general public with the advancement of smart phone technology [Vuforia][Layer][Metaio]. These mobile-based applications allow users to apply AR in advertisement, entertainment, art, and Human Computer Interaction (HCI) research [Wagner 2009].

The AR companies Layar and Metaio allow users to virtually display interactive digital content on top of print media [Layar][Metaio]. By tracking the visual elements of print surface and using sensors in mobile phone, the applications create AR

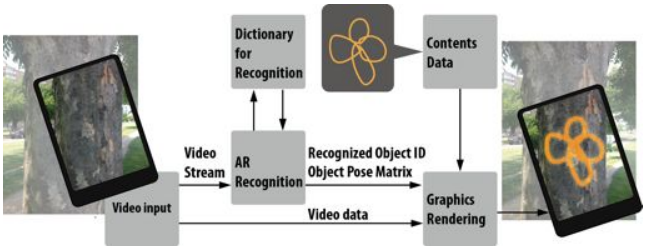

Figure 2. *AR recognition procedure*

contents that blend the online and offline materials together. Stiktu allows users to scan everyday objects and virtually add AR stickers and images on surfaces [Stiktu]. In Stiktu, the generated content can be captured as a digital photo and published on social networking platforms such as Facebook and Twitter. Scrawl is a marker-based three-dimensional AR drawing mobile application by String [String]. Scrawl requires users to print markers to create interactive AR applications. One of the limitations of such products is that they do not support shared workspace for multiple users to generate digital content in real time.

There are projects that introduce spatially aware display which supports collaborative virtual environment in real space [George 1993][Poupyrev et al. 1998] [Michael et al. 2003] [Schmalstieg et al. 2000]. However, the platform is designed to work in a system spatial operating environment based on special settings or devices such as VICON system or external video camera, and does not support interaction in everyday settings.

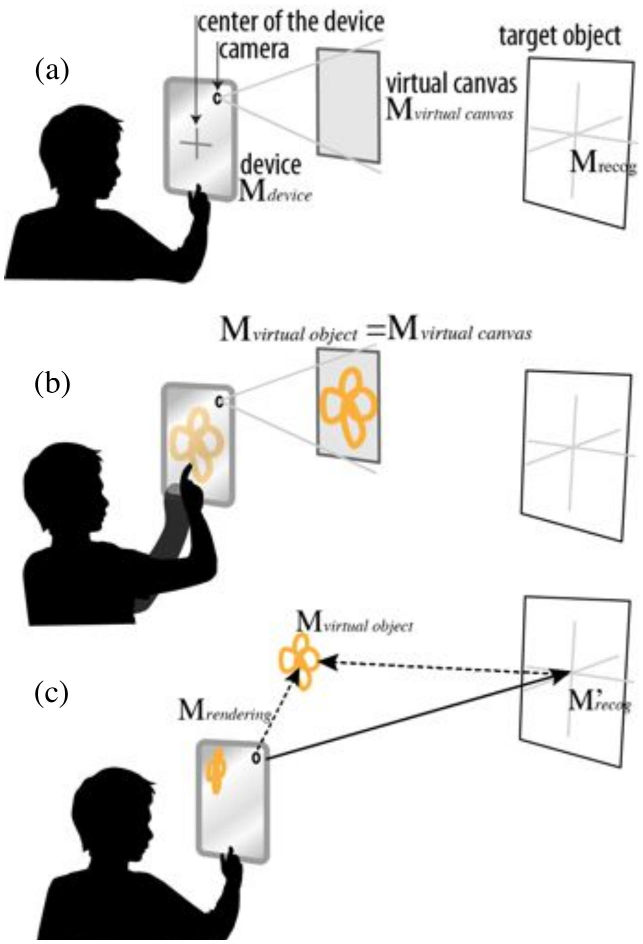

Figure 3. *Matrix definition:* (a) *the pose of the device and the pose of the user's virtual canvas are defined based on the estimated pose of the target object*. (b) *A generated content is placed on the virtual canvas*. (c) *These generated contents are rendered based on on the pose of the target object*

There have been many prior works using image based simultaneous localization and mapping technology (SLAM) for the pose estimation of the camera and the environment recognition [Georg et al. 2009]. Microsoft's Photosynth also show the potential of gathering three-dimensional information of physical space using 2D images of the user's surroundings [N. Snavely et al. 2006]. These emerging technologies are going to be applied into mobile devices.

4. Implementation

Second Surface uses image based AR recognition technology (Vuforia™ [Vuforia]), which recognizes a natural image as a target object with advanced registered dictionary data. We can create a set of dictionary data from the picture of the surface from the everyday object. These dictionary data contains the image feature data for recognition. This feature is like a fingerprint that indicates every surface as a unique object. Dictionary data means a database of trackable objects.
Our current implementation allows multiple tablet devices to share the same AR dictionaries, and allows content from both devices to be placed in the same virtual coordinate system.

A real-time video stream from the built-in camera is used for AR recognition and the system estimates the pose M_{recog} of the target object in relation to the camera focusing on it (Fig 2). Here, M represents a 16-dimensional matrix consisting of the pose includes a translation and a rotation. We can define the coordinates of the shared space by taking M_{recog} as a reference frame.

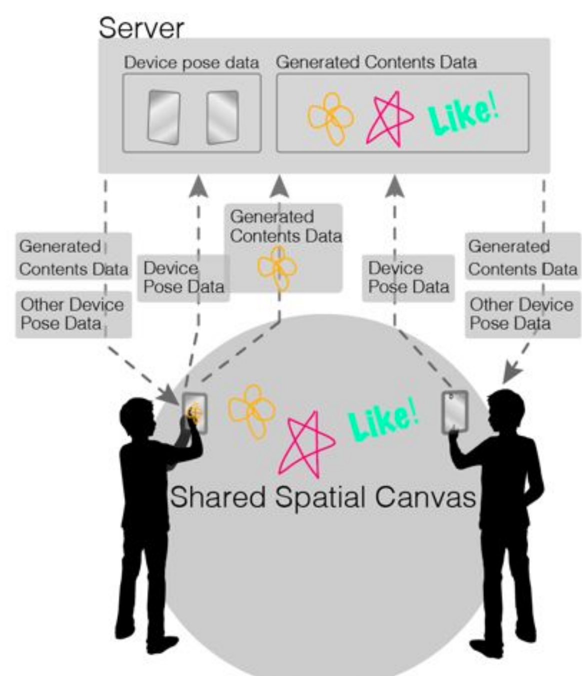

Figure 4. *Server system: Generated Contents Data includes the contents itself and it's pose matrix* $M_{virtual\,object}$.
Device Pose Data includes the device pose matrix M_{device}.

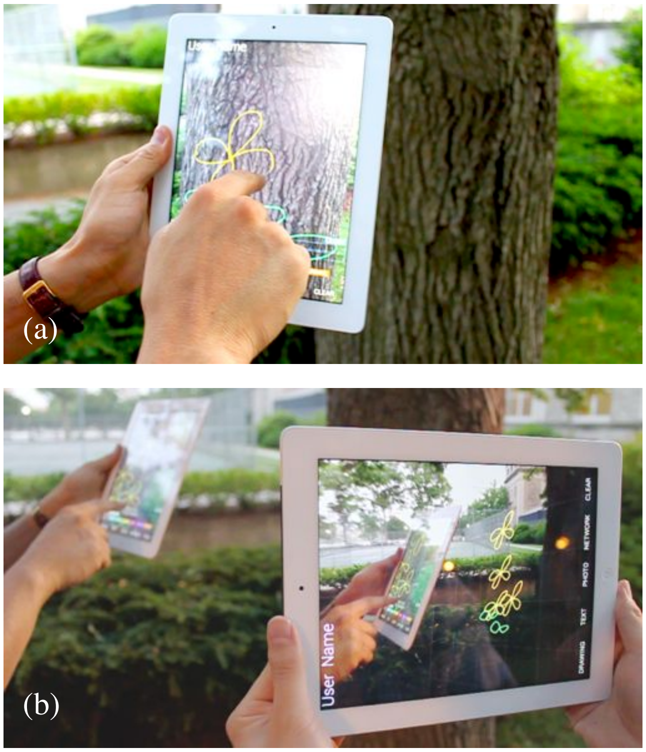

Figure 5. *Content generation and visualization of contents.* (a) *User can create content and situate it on the augmented reality spatial canvas.* (b) *Co-located users can see the created contents through their tablet devices in real-time.*

We calculate the pose M_{device} of the device and the pose $M_{virtual\ canvas}$ of the user's virtual canvas from the real world environment based on the estimated pose M_{recog} of the target object (Fig 3).

$$M_{device} = M_{recog}^{-1} \cdot T_{device\ center}$$

$$M_{virtual\ canvas} = M_{recog}^{-1} \cdot T_{offset}$$

Here, T_{offset} represents a translation matrix from camera to virtual canvas, which is defined based on a device's camera parameter. $T_{device\ center}$ represents a translation matrix from camera to center of the device, which is defined based on the physical placement of device's camera. Once the content is located at the virtual canvas position, the pose $M_{virtual\ object}$ of the virtual object is defined as the pose $M_{virtual\ canvas}$ of the virtual canvas. Afterward, these virtual objects are rendered on top of the video stream based on the pose M'_{recog} of the target object.

The multiple tablet devices are connected through a server application with each other. When a new content is generated, the server handles the data once it is sent from a device, and pushes it to the other devices that are facing the same object. Also all devices' pose data are shared via this server. Since the object acts as a fingerprint or ID, the server needs to handle only those devices that look at the same object. The server stores all generated content data as well, so once a user looks at an object though his/her device, it gets provided with all information that is linked with the object (Fig 4).

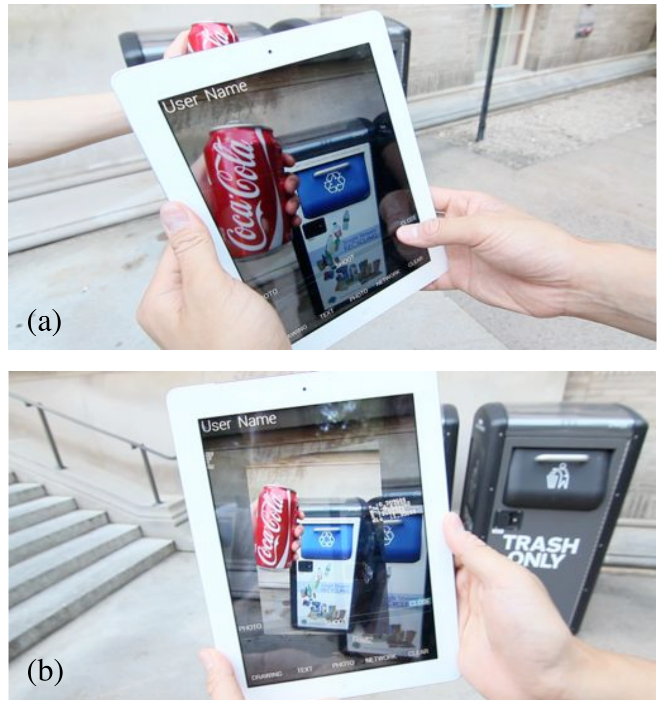

Figure 6. (a) *User can also shoot a photo and place it on the spatial canvas. The pose data and the generated content are saved in the server.* (b) *Later on user can access the archived contents embedded in the spatial canvas.*

Using this infrastructure, the user can create simple drawings, text messages and photo shoots in space using his/her tablet devices. These contents are automatically placed in user's physically located virtual canvas position ($M_{virtual\ Object}$) which is highly relative to where it has been created in the tablet. Other co-located users can see the created content with their tablet devices at the same position in the virtual canvas where the AR content was originally placed (Fig 5). All of these contents are instantly shared with every other user facing the same virtual object through his/her device (Fig 4). Since the content is stored in the server, all created content is continually present in the virtual space.

Some usage case scenarios can include explaining the use of trashcans and visualizing what can be thrown away in them (Fig 6.). For example, user can create photos of the bottles that fit in a

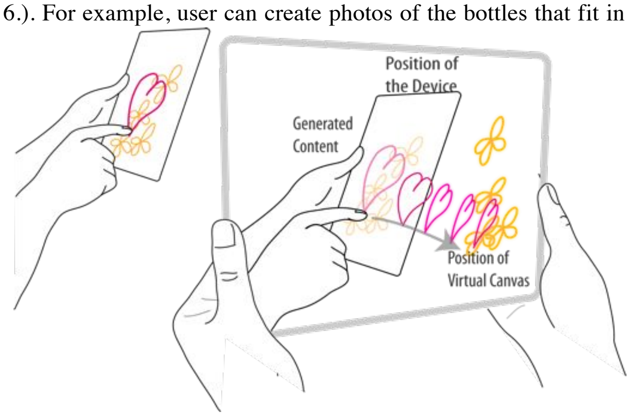

Figure 7. *The contents pop up from the position of the device and move into the place where the virtual canvas is physically located.*

certain type of can and virtually place the generated content near the trashcan. We also tested the system on trees, where we can leave notes for other people who pass by in order to explain the environment to them or show them things we find interesting.

To generate a consistent shared space, it is important that the virtually generated content is in spatial synchronization with the device that has created it. Second Surface uses matrix calculation procedure to provide a very natural feeling relative to the physical scale of the real world and the AR content. For example, in Second Surface, the generated AR content pops out from the position of the device and moves into the place where the virtual canvas is physically located. All of this happens in natural three-dimensional animation (Fig 7). The location of the virtual canvas where the AR content places itself is relative to the device's position and the generated pose matrix data is shared with other co-located devices.

5. Conclusions and Future Work

Second Surface is a novel way to interact with everyday surfaces using spatial collaboration. We believe that our system can create new ways of communicating within cities, schools and households. We hope our implementation will create an interesting and new user experience that feels natural and encourages playful content generation. In our implementation, we utilize image based object recognition using the dictionary data for device pose estimation from the real-world environment. We envision that a future system will be based on a cloud service that collects dictionary data and 3D reconstruction data from all its users and provide this data to anyone in the form of a map. With such a system the real world can be filled with endless virtual interactions, without the charge of vandalism. The full potential of such a system would be massive user and massive space applications. Multiple users could draw on this virtual canvas through the lens of computational devices such as head mounted displays, phones and tablets.

References

FOAD HAMIDI AND MELANIE BALJKO. 2012. Using social networks for multicultural creative collaboration. In *Proceedings of the 4th international conference on Intercultural Collaboration (ICIC '12)*. ACM, 39-46.

D. W. F. VAN KREVELEN, R. POELMAN. 2010. A Survey of Augmented Reality Technologies, Applications and Limitations, *The International Journal of Virtual Reality,* Vol. 9, No. 2, pp. 1-20.

WANGER, D. 2009. History of Mobile Augmented Reality, *Communications*. Retrieved from https://www.icg.tugraz.at/~daniel/HistoryOfMobileAR

Qualcomm , Vuforia™ : http://www.qualcomm.com/solutions/augmented-reality

Layar. Layar. http://www.layar.com

Metaio. Junaio. http://www.junaio.com

Stiktu. http://www.stiktu.com

String. Scrawl. http://www.poweredbystring.com/showcase

GEORGE W. FITZMAURICE. 1993. Situated information spaces and spatially aware palmtop computers. *Commun*. ACM 36, 7, 39-49.

POUPYREV, I., TOMOKAZU, N., WEGHORST, S. *1998*. Virtual Notepad: handwriting in immersive VR, *Virtual Reality Annual International Symposium,* Proceedings., IEEE 1998 , vol., no., pp.126-132, 18-18 1998

MICHAEL TSANG, GEORGE W. FITZMZURICE, GORDON KURTENBACH, AZAM KHAN, AND BILL BUXTON. 2003. Boom chameleon: simultaneous capture of 3D viewpoint, voice and gesture annotations on a spatially-aware display. In ACM *SIGGRAPH 2003 Papers (SIGGRAPH '03)*. ACM, New York, NY, USA, 698-698.

SCHMALSTIEG, D., FUHRMANN, A., HESINA, G. 2000. Bridging multiple user interface dimensions with augmented reality , *Augmented Reality, 2000. (ISAR 2000)*. Proceedings. IEEE and ACM International Symposium on , vol., no., pp.20-29, 2000.

GEORG KLEIN AND DAVID MURRAY. 2009. Parallel Tracking and Mapping on a Camera Phone. In Proc. *International Symposium on Mixed and Augmented Reality (ISMAR'09, Orlando)*. 83-86.

N. SNAVELY, S. M. SEITZ, AND R. SZELISKI. 2006. Photo tourism:Exploring photo collections in 3d. In *SIGGRAPH Conference Proceedings,* pages 835–846, New York, NY, USA, 2006. ACM Press.