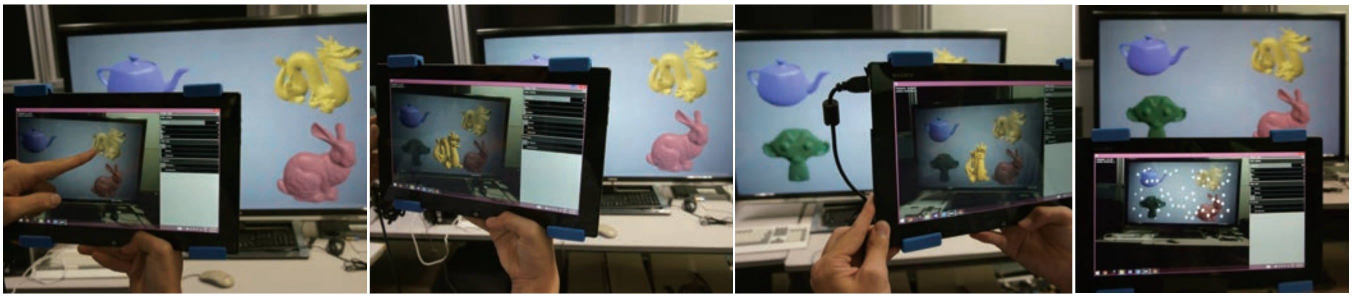
Imperceptible On-Screen Markers for Arbitrary Background Images

Goshiro Yamamoto* Luiz G. M. Sampaio† Takafumi Taketomi‡ Christian Sandor§ Hirokazu Kato¶

Graduate School of Information Science
Nara Institute of Science and Technology



(a) (b) (c) (d)

Figure 1: A user can experience Augmented Reality (AR) objects on a handheld device, which are displayed on top of an external display (a)-(c). Our method has all following properties: computation of full pose, the camera can move, and the external display images can change. Furthermore, the external display content don't need to be known in the handheld device in advance. Our algorithm can embed these markers imperceptibly for users (b).

Abstract

In this demo, we present a system that can realize a spatial interaction between a handheld device and external displays via imperceptible on-screen markers. Our system can compute a full pose of the camera installed on the handheld device by extracting a marker from the external display, while a user does not notice where the marker is. This technology enables users to see Augmented Reality (AR) objects on top of the external display through their handheld devices. Additionally, we can implement a remote interface to access to the content on the external displays by touching the handheld device screen. Our method consists of two parts: marker embedding on external displays and marker detection. Compared to previous research, our system allows that the external display content can change and don't need to be known in advance. Our method is working well with small movements, while large movements can lead to loss of tracking. We consider our work to be the first step towards large scale deployment of handheld AR content for a large number of users with unmodified devices.

Keywords: Imperceptible marker, human visual perception, unsynchronized capture

Index Terms: H.5.2 [Information Interfaces and Presentation]: User Interfaces—Interaction style; I.2 [Artificial Intelligence]: Vision and Scene Understanding—Representations, data structures, and transforms

*e-mail: goshiro@is.naist.jp
†e-mail: luiz.sampaio.ll6@is.naist.jp
‡e-mail: takafumi-t@is.naist.jp
§e-mail: sandor@is.naist.jp
¶e-mail: kato@is.naist.jp

1 Background

Handheld devices such as smartphones and tablet computers have become common devices that people use in their daily life. These devices have small, portable screens which are ideal for personal content. Besides handheld devices, displays tend to vary in sizes from desktop displays in office spaces to wall-sized displays. These make external devices necessary for collaboration and displaying public content. Recently researchers propose to leverage on the strength of these two types of devices by creating user interactions between them. One of such interactions is using handheld devices for extracting some secondary content (QR codes, augmented reality marker, etc.) from some primary content on the external display. Handheld devices are suitable for handling personal information, whereas the large displays show information anyone in public places. Considering these characteristics, the interaction using handheld devices with external displays can be seen as a good combination.

In this demo, we show a novel marker that is imperceptible foreground for arbitrary background images on the screen. We accomplish this by leveraging on the property of the human eyes to average colors when presented with different colors at a rate faster than the flicker fusion threshold of the human visual system. We treat the two colors as a complementary pair. This allows us to modify the arbitrary images to communicate with a computer vision system, without obstructing the human's view.

Our work contributes to a wide-area of use cases for any type of interaction between handheld devices with cameras and digital displays. Our method is the first with all the following properties: computation of full pose, the camera can move, and the external display content can change without synchronization. Using our system, arbitrary images can be backgrounds of markers. This means that a camera pose can be calculated anytime when the user holds the handheld device as whose camera faces to the digital displays. We consider our work to be the first step towards a large scale deployment of handheld AR content for a large number of users with unmodified handheld devices.

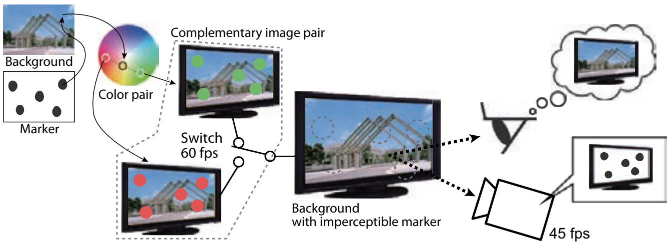


Figure 2: Complementary image pair is created using a background and a marker as a mask. These images are switched faster than Critical Flicker Frequency on a display. Humans perceive only the background, whereas a computer vision system can detect the embedded marker.

2 Related Work

Tablets and smartphones have been widely used as tools for interaction with external objects including distant displays. Content creation, edition, and transfer between devices are some of the activities that can be provided to engage users and increase collaboration [5]. For tasks that require precise manipulation, Touch Projector [2] makes use of zooming and temporary freezing of the camera image. Augmented TV proposes an AR application that synchronizes to the broadcasting movies [4].

In order to implement an imperceptible marker system with ordinary equipment such as consumer-grade cameras and off-the-shelf projectors, unsynchronized technology is suitable for multiple users. VRCodes [8] is the most similar to our work. This system also uses color mixture effects and provides tagging system with digital displays with rolling shutter cameras. However, the tagging process requires the location information of the code in advance. Accordingly, it is difficult to apply this technology to arbitrary backgrounds for embedding codes. Although several similar works that used spatial coding or watermark techniques has been reported [7, 10, 9], they assume that the camera always captures obvious references such as frames of a display device. In our method, we don't need to know background on the screen, that is, our system can work for arbitrary background images. Additionally, we use only the embedded markers without the fiducial points in the captured image.

Natural feature tracking can be implemented to handheld interaction with external displays [1], but it needs feature points on the screen. By contrast, our system works for even monochrome images that have few natural feature points.

3 Method

Our method consists of two parts: marker embedding on external displays and marker detection. Our concept is to treat the original image as a background image on external displays. We then put the marker on the foreground imperceptibly. Figure 2 shows the overview of the method with these two processes. To embed the marker onto the screen, it is important to design the marker to be distinguishable for a computer vision system, but unobtrusive for the human visual system.

We can embed a marker to any images by representing it as a complementary image pair. This pair of images is generated by modifying colors on the background image. We do this modification such that the human visual system will perceive only the background image when we continuously switch between the two images. This method takes advantage of the human visual system's temporal integration of light. That is, when switched faster than the Critical Flicker Frequency (CCF), two alternating colors will be perceived to be the average of these two colors [3]. Although imperceptible to humans, this type of marker can be extracted using computer vision techniques. We use the marker as a mask to determine the area of the background image within which we will calculate complementary pixel values (color pairs). We then create the image pair by replacing the corresponding background pixel with these color pairs. This image pair are switched alternatively on the screen of a display device at 60Hz. This speed is chosen so that the human eyes can only see the background image. Theoretically, the two colors works as one color for human eyes when the refresh rate is higher than CFF which is around 60Hz.

Our method allows a computer vision system to detect the imperceptible on-screen makers. In order to detect a marker through captured images, we have set a camera at a capture rate of 45fps. We then extract an embedded marker by accumulating each difference between three sequentially captured images while the display device switches the complementary image pair on the screen at 60fps. In this demo, we apply the algorithm of random dot markers to make marker mask images [6].

4 Demo Description

Our demo system consists of a tablet device which has a camera and an external large display. Additionally, we use a server that computes image processing instead of the tablet device through wireless network. As a user looks through the tablet screen at the screen of external display which shows an images, augmented reality objects appear on top of the external screen. The user perceives only the background image, whereas the computer can detect an embedded marker and estimate the camera pose and the external display ID. The background image on the external display can be switched with other images. Since the estimated pose and ID gives information where the camera looks towards which external display, we can develop variety of applications based on a spatial interaction using our marker system. In this demo [1], we show an application that enables a user to experience augmented reality objects on top of the external display with simple input interface as shown in Figure 1.

References

[1] D. Baur, S. Boring, and S. Feiner. Virtual projection: Exploring optical projection as a metaphor for multi-device interaction. In *CHI 2012*, pages 1693–1702, May 2012.

[2] S. Boring, D. Baur, A. Butz, S. Gustafson, and P. Baudisch. Touch projector: Mobile interaction through video. In *CHI 2010*, pages 2287–2296, Apr. 2010.

[3] M. D. Fairchild. *Color Appearance Models*. Addison-Wesley, Reading, MA, first edition, 1998.

[4] H. Kawakita, T. Nakagawa, and M. Sato. Augmented tv: an augmented reality system for tv pictures beyond the tv screen. *Trans. Virtual Real. Soc. Japan*, 19(3):319–328, 2014.

[5] M. D. B. Machuca, W. Chinthammit, Y. Yang, and H. Duh. 3d mobile interactions for public displays. In *SIGGRAPH Asia 2014 Mob. Graph. Interact. Appl.*, page 17, 2014.

[6] H. Uchiyama and H. Saito. Random dot markers. In *IEEE Virtual Reality*, pages 35–38, Mar. 2011.

[7] A. Wang, C. Peng, O. Zhang, G. Shen, and B. Zeng. Inframe : Multiflexing full-frame visible communication channel for humans and devices. In *HotNets 2014*, pages 1–7, 2014.

[8] G. Woo, A. Lippman, and R. Raskar. Vrcodes: Unobtrusive and active visual codes for interaction by exploiting rolling shutter. In *ISMAR 2012*, pages 59–64, 2012.

[9] S. Yamaguchi, H. Tanaka, S. Ando, A. Katayama, and K. Tsutsuguchi. Visual syncar: Augmented reality which synchronizes video and overlaid information. *J. Inst. Image Electron. Eng. Japan*, 43(3):397–403, 2014.

[10] W. Yuan, K. Dana, A. Ashok, M. Gruteser, and N. Mandayam. Dynamic and invisible messaging for visual mimo. In *Proc. IEEE Work. Appl. Comput. Vis.*, pages 345–352, 2012.

[1] Video: *http://imd.naist.jp/~goshiro/video/ISMAR2015-demo.mp4*