Detection of Imperceptible On-Screen Markers with Unsynchronized Cameras

Luiz Sampaio[*1] Yoshio Yamada[*2] Goshiro Yamamoto[*3]
Takafumi Taketomi[*4] Christian Sandor[*5] Hirokazu Kato[*6]

Abstract – We present a method for detecting on-screen markers to identify physical displays and to estimate camera poses against these displays. It is difficult for humans to perceive the variation of two alternating colors when these colors switch in a frequency higher than the flicker fusion threshold. On the other hand, the variation is detectable by of-the-shelf rolling shutter cameras which have certain frame rate setting. Leveraging on the difference between the human visual characteristics and the camera characteristics, imperceptible markers can be embedded onto a display screen. In this paper, we introduce our technique in mobile device based user interfaces for manipulating digital content on large distributed displays. Furthermore, we show that our prototype system enables an augmented reality application on a mobile device using the marker embedded displays. Lastly, we discuss potential scenarios and applications.

Keywords : information embedding, visual performance, sensitively behavior of camera, multi-display environment.

1 Introduction

Public displays and digital billboards have become present in most of urban centers over the last years. However, most of public displays still lack of interactivity, remaining simply electronic versions of their ancestry. The ones that allow some kind of interaction are not designed for simultaneous access, blocking all interested people while one is using. With the proliferation of mobile devices, it is clear that an opportunity to expand this interaction is still poorly explored; they can work as an expansion of the public display, allowing each user to have their own private version of the remote content.

The first step to achieve this goal is the identification of the remote display by the user's mobile device. Other researchers have used visible tags inside and outside the screen, but these markers are obtrusive and degrade the content visualization. Recent works explored the detection of visible salient features (edges) on the displayed image [1], however, it is impractical when considering the identification of one among various displays showing the same content, or when the image has only few or even no points of interest.

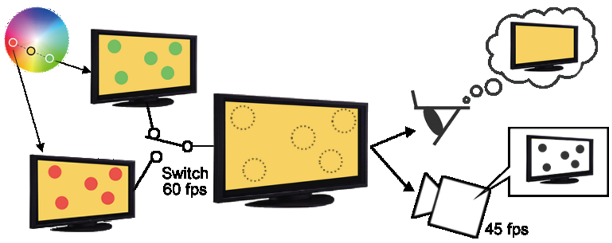

Fig. 1 The difference of perception between the human eye and a camera.

The main contribution of this paper is the method for detecting visual codes embedded in textured images, using off-the-shelf cameras, remaining imperceptible for the human eye. Different from other works, our method does not require synchronization between the screen frames and camera capture, what makes it especially interesting when considering unmanaged environments, such as public places. Additionally, as the computation of the target color is made separately for each pixel, the visual codes can be inserted in any region of the image, regardless what type of texture the image contains.

*1 Nara Institute of Science and Technology
*2 Nara Institute of Science and Technology
*3 Nara Institute of Science and Technology
*4 Nara Institute of Science and Technology
*5 Nara Institute of Science and Technology
*6 Nara Institute of Science and Technology

2 Related Work

The idea of exploiting the difference between how camera sensors and the Human Visual System perceive the colors is not new, and have been explored for different purposes over the years. VRCodes [1] employs the rolling-shutter camera built in smartphones to detect binary codes embed in active displays. Switching complementary hues at 60Hz, the human eye see just a flat gray-colored screen, while the mobile camera see the decomposed aliasing effects of the changing hues. While the use of a rolling-shutter camera seems to be a strict constraint, our approach proved to be efficient in both rolling-shutter and global shutter cameras. Moreover, there is no information regarding the use of this technology on the whole screen, when displaying content simultaneously. When a newspaper headline in shown, the binary code is applied only on the top of the screen.

Grundhöfer et al. [2] propose a method of embedding binary codes integrated with the projected image. The embedded code is dynamically adjusted to guarantee its imperceptibility and to adapt it to actual camera pose. However, in order to achieve the marker tracking method, the camera, projector and a LED illumination module must be interconnected with a synchronization unit, what makes this approach not applicable for our target audience.

While our approach is based on the detection of imperceptible markers, most of other researchers has focus on the detection of visible feature points on the screen. Touch Projector [3] and Virtual Projection [4] rely on the matching of features detected in the camera image and the image displayed on the remote screen. A drawback of this technique is that it is content-dependent, it means, the system is not able to distinguish different screens showing the same content. Visual SyncAR [5] applies digital watermark on videos, that when captured by the camera's smartphone, can show computer graphics animations synchronized with the video. However, in order to compute the camera position, the four corners of the screen must be within the camera image. It can be difficult to achieve when considering large public displays, as we are aiming.

Alternative ways to track the camera position have also been tested. Bergé et al. [6] described the application of Polhemus Patriot Wireless trackers, using the magnetic field created between the transmitter and receptor to determine the position relative to display.

3 Design and Detection of On-Screen Markers

3.1 Imperceptible On-Screen Markers

Flicker fusion threshold is the frequency at which an intermittent light appears to be steady to the human observer. In the same way, complementary colors, when alternated faster than this frequency, are perceived as a combination of both. For instance, at 120Hz it is possible to use colors with higher amplitude and still keep the same effect. In our system, as most of standard displays run with refresh rate of 60Hz, we work with lower contrast colors.

3.2 3.2 Detection and Tracking

The concept of "beat" is well known in acoustics, and it is defined as the interference that occur between two sound waves of slightly different frequency.

Based in this idea, we defined our approach to capture the video frames without relying in signal synchronization. Considering the available displays, the standard refresh rate is 60 frames per second. By setting the camera at 45 frames per second, we found that the 3 frames are necessary in order to extract the marker.

4 Experimental Results

4.1 Implementation

Our system is organized in a distributed architecture as show in Fig. 2. The client application is running in a Sony Vaio Tap 11 Tablet PC. As we need to capture the video at 45fps, so far is not possible to use the built-in camera. Instead, we are using a PlayStation Eye, which offers a wide range of modes for controlling exposure, gain and so forth. The main function of the client application is continuously capture "rounds" of 3 frames and extract the binary pattern by subtracting the color difference of one frame from other.

The binary image is then sent to the server application, which runs in Intel Core i7 3.7GHz with 32 GB of RAM. This application receives the binary image from the client, performs the pattern recognition using the Locally Likely Arrangement Hashing (LLAH), and from the computed homography matrix, performs the camera pose estimation. Additionally this application manage the running instances of

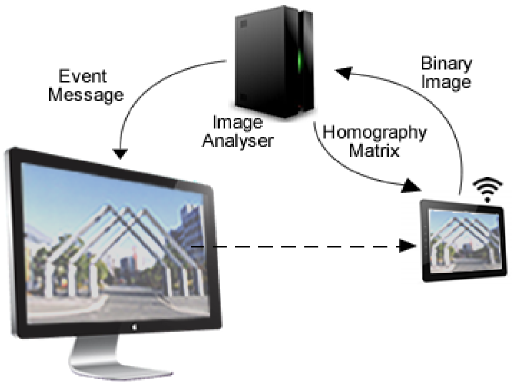

Fig. 2 System configuration.

Fig. 3 Client application showing the dot pattern overlaid on the camera image.

Fig. 4 Example: annotations based in context.

client applications and pattern generators (which run on the displays). Once the marker is detected and the projection matrix is computed, this is sent back to the client application, for rendering an augmented content in the correct position. In this our case, we show some annotation regarding the image that is displayed on the screen.

The "pattern generator" is the application which runs on the remote displays, and for that we are using an Intel Core i7 3.20GHz with 8GB of RAM. To embed the marker in the image we want to show, we generate a Random Dot Pattern [7] beforehand. Then, the system computes the complementary colors on target image, in accordance with the marker's position. As result, we get two images, each of them with the embedded markers in one of the complementary colors. When these two images are flickered at 60 fps, the human eye can perceive only the resulting color. This application also receives events from the server application. For instance, when the user swipes a finger on the tablet screen, the coordinates are sent to the pattern generator through the server application, what results in a line drawing on the remote screen. For displays, we used a 50" Panasonic TH-50PF Plasma Screen, but also conducted tests with diverse models of standard PC screens.

4.2 Results

Figure 4 shows the result of embedding a dot marker in a picture of Tokyo. At the remote display, no dots can be seeing, because we are switching the images with the dots in complementary colors in a frequency faster than the human eye can perceive. On the tablet display, we highlighted the dot pattern for illustration purposes. As the light from the environment and the brightness from the remote display can affect the detection of the pattern, we created a slider to control the camera exposure manually.

From this basic framework, we can develop many applications. Figure 4 shows annotations about the buildings on the image. In a real world application, the annotations could be presented according to user ' s language, or according some specific setting. We believe this kind of application can be very useful in environments such as museums, airports, shopping malls and any other place where it is providing tailored information for diverse and large amount of people.

Due the feature subtraction of three consecutive frames, when we move the camera quickly, the resulting image presents excessive quantity of noise, what makes the marker detection more challenging. We are currently studying ways to identify and separate the blobs generated by noise from those that are actually part of the markers.

For the same reason, our technology is yet not feasible for displaying dynamic content, such as videos and animations. In this case, as the most of colors and contours tend to change drastically at every frame, the marker detection becomes impossible.

5 Conclusion and Future Work

With the proliferation of public displays, and considering that mobile devices are nowadays ubiquitous, there is a huge opportunity to develop new applications and studies in human-computer interaction. However, most of current approaches for tracking the remote display are obtrusive or strongly dependent of the content shown on the screen. In this paper, we presented an innovative method detection of visual codes embedded in textured images, using commodity cameras with 45 fps capture rate. The codes are embedded in the images computing complementary colors, which due the effect of metamerism, when switched at speed higher than the fusion-frequency of the human eye, are perceived as just one resulting color.

As future work, we want to provide deeper understanding of the Human Visual System as well as more detailed user study for a real evaluation of our technique. Tests with newer off-the-shelf displays, with refresh rate of 120Hz will allow us to use higher contrast colors, what can make the marker detection easier and still remaining unobtrusive for the human eye. Additionally, we intend to adapt this technology to work with movies. Then, a new range of applications can be develop on the top of our framework.

References

[1] G. Woo, A. Lippman, R. Raskar: VRCodes: Unobtrusive and Active Visual Codes for Interaction by Exploiting Rolling Shutter. ISMAR 2012. pp. 59–64, 2012.

[2] A. Grundhöfer, M. Seeger, F. Hantsch, O. Bimber: Dynamic Adaptation of Projected Imperceptible Codes; ISMAR 2007, pp. 1–10, 2007.

[3] S. Boring, D. Baur, A. Butz, S. Gustafson, P. Baudisch: Touch Projector: Mobile Interaction Through Video. SIGCHI, pp. 2287–2296, 2010.

[4] D. Baur, S. Boring, S. Feiner. Virtual projection: exploring optical projection as a metaphor for multi-device interaction. SIGCHI. pp. 1693–1702, 2012.

[5] S. Yamamoto, H. Tanaka, S. Ando, A. Katayama, K. Tsutsuguchi. Visual SyncAR: Augmented Reality which Synchronizes Video and Overlaid Information. The Journal of the Institute of Image Electronics Engineers of Japan, Vol. 43, pp. 397–403, 2014.

[6] L. P. Bergé, M. Serrano, G. Perelman, E. Dubois. Exploring Smartphone-based Interaction with Overview+Detail Interfaces on 3D Public Displays. MobileHCI. pp. 125–134, 2014.

[7] H. Uchiyama, H. Saito. Random Dot Markers. IEEE VR. pp. 35–38, 2011.