On Display-Camera Synchronization for Visible Light Communication

Kaixuan Liu #1, Xiaolin Wu #*2, Xiao Shu *3

Shanghai Jiao Tong University
800 Dongchuan Road, Shanghai, China
1 kaixuan_liu@sjtu.edu.cn
2 xwu510@sjtu.edu.cn

* *McMaster University*
1280 Main Street West, Hamilton, Canada
3 shux@mcmaster.ca

***Abstract*—Recently, Wu and Shu proposed an intriguing technique that transmits data from regular information display devices to smart phone cameras without using any additional hardware or interfering the original content intended to display. This preliminary work reveals the possibility of conducting high-speed data transmission with displays in transparency to human eyes; however, in practice, it could introduce visible artifacts and the display-camera synchronization remains an issue that affects data throughput. In this paper, we examine this technique in more realistic application scenarios and put forward a practical solution that hides the data more effectively and greatly improves the robustness of the transmission. Experimental results show that, using the proposed solution, viewers can hardly notice any artifacts caused by the data embedded in videos, and the bit error rate is acceptable under common use conditions.**

***Index Terms*—Visible light communication, temporal psycho-visual modulation, information display,perceptual image quality, synchronization.**

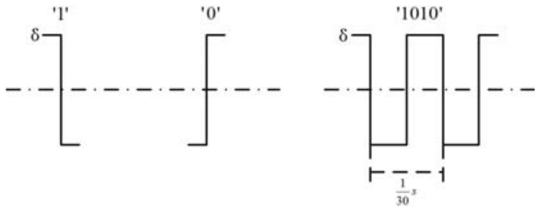
I. Introduction

Over the last few years, visible light communication (VLC) has attracted significant attention from both academia and industry. By virtue of its advantages in bandwidth availability, safety and data security, VLC is considered a better alternative to radio frequency and infrared in many applications, such as, wireless network using lighting system [1], indoor positioning [2], optical interconnects to electronic chips [3], body sensor networks [4], etc. Many of these VLC techniques are designed to exploit existing light-producing devices in order to make the techniques easier to adopt, however, they often still require some specialized receivers to extract data signals from light sources, raising the cost of deployment inevitably.

In a recent study, Wu and Shu explored a novel idea of transmitting and receiving data using commonly available display and camera, respectively [5]. One highlight of the technique is that it does not disrupt the original functionality of the display; during data transmission, the display can still show still images or videos normally without human viewers perceiving the hidden data signals. This capability is achieved by modulating data with a flipping signal added to the original display content. If the added signal flips at a higher frequency than flicker fusion frequency, which is around 60 Hz, then most people cannot detect the flickering caused by the signal and would perceive it as temporally constant [6]; on the other hand, the added signal must be slow enough to be captured by video camera, otherwise, it is not possible to recover the modulated data by Nyquist-Shannon sampling theorem. To show a temporal signal of 60 Hz, the display device should support at least 120 Hz refresh rate, which is quite common for modern display devices, such as liquid crystal displays (LCDs). To catch the signal, the camera must be able to capture video at 120 frames per second (fps). Many available cameras can achieve this speed or even higher. For example, the rear-facing camera on iPhone 6 supports capturing 720p video at 240 fps, which is likely to become the norm of smart phone cameras in the near future. Nowadays, the hardware devices required by Wu and Shu's technique are not only available at relatively low cost but also already widely used by consumers.

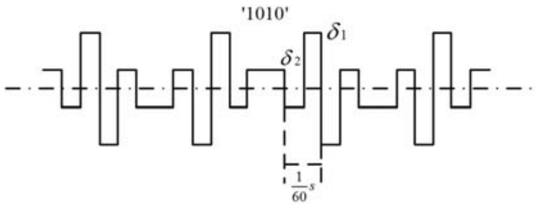
Although Wu and Shu's technique reveals the possibility of conducting high-speed data transmission with displays in transparency to human eyes in theory, to apply the technique in a more realistic environment, several practical issues that could greatly deteriorate the display visual quality and data transfer accuracy must be addressed. In essence, the greatest challenge of designing such communication system is that the data transmission only goes one direction from the display to the camera. If the camera detects error, it could not ask the display to repeat the data or change the coding without involving other communication mechanisms. Therefore, to guarantee an acceptable bit error rate under common use conditions, the display must provide sufficient data redundancy. In this paper, we investigate a method of placing redundant data blocks in each frame so that at least one copy of the data is likely to be detected. If the camera is hold upside down, meaning that

This research is supported by National Natural Science Foundation of China (NSFC grant 61331014).
IEEE VCIP'15, Dec. 13 - Dec. 16, 2015, Singapore
000-0-0000-0000-0/00/$31.00 ©2015 IEEE.



(a)

(b)



(c)

Fig. 1: Signal waveform

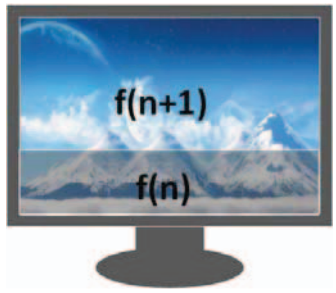


Fig. 2: Image displayed on the screen at a certain time point might be a blend of two consecutive frames.

the camera scans the display from bottom to top, this method can effectively maintain a low bit error rate regardless of the timing of the display and camera. Our study also shows the necessity of employing redundancy in temporal domain. Not only does temporal redundancy improve the reliability of data transfer like the spatial case, but it also helps to conceal the data signal from human eyes better.

The remainder of this paper proceeds as follows. Section II discusses how to more effectively conceal data transmission from human eyes. Section III introduces our methods to make the transmission more robust in realistic environments. The experimental results are presented in Section IV, and Section V concludes.

II. Conceal the Transmission

The fundamental idea of Wu and Shu's technique is to generate two atom frames x_1, x_2,

$$x_1 = y - (-1)^{data}\delta,$$
$$x_2 = y + (-1)^{data}\delta, \quad (1)$$

for each frame y of a normal video and display all the atom frames in sequence at a refresh rate of 120 Hz [7], [5]. Essentially, the embedded data signal is encoded using Manchester code with the waveform as shown in the left part of Fig. 1a. If the data sequence contains all '1's or '0's, the encoded signal is a square wave of frequency 60 Hz, half of the refresh rate of the display. In this case, because the signal flips at the human flicker fusion frequency, it is indeed indiscernible to most human viewers while still detectable using a high-speed camera as suggested by the authors. However, if there exists a series of '10's or '01's in the data as shown in the right part of Fig. 1a, the frequency of the encoded data signal drops to 30 Hz, which is perceivable to viewers. Although this problem does not affect data transmission, the perceived visual quality of the original video is degraded by the visible flickering.

This flickering problem only happens at the data bit transitions either from 0 to 1 or from 1 to 0. Thus, simply to use more atom frames to encode each data bit instead of two would decrease the impact of the transitions and conceal the data transmission better. As demonstrated in Fig. 1b, 6 atom frames are used to transmit one bit. The signal waves have relatively smaller amplitude δ_2 at the beginning and end to smooth the transition between bit 0 and bit 1. The amplitude δ_1 in the middle of the waveform is larger in order to reduce bit error rate. However, the amplitude δ_1 cannot be too large, as that would shift the average of two adjacent atom frames, revealing the concealed signal. From experiment results, we found that $\delta_1 = 3\delta_2$ usually yields a good balance between bit error rate and output visual quality. As this improvement on reducing flickering artifacts is achieved at the cost of data transfer rate, the number of atom frames per bit cannot be arbitrarily large. In most scenarios we tested, encoding each bit with 6 atom frames is sufficient to hide the data transmission from human eyes.

III. Increase Transmission Robustness

To achieve a reasonable and stable transfer rate in a more realistic environment, the hidden high-speed data transmission technique must be able to deal with several practical issues that could greatly affect the transmission. One of the most conspicuous problems is that, since it takes some time in each frame period to refresh pixels sequentially from top to bottom, at the majority of the time, the top part of the screen has received the update and starts to display the new frame while the rest of the screen remains showing the previous frame. As a result, a shot of the display is most likely to be two consecutive frames stitched together, as shown in Fig. 2, rather than a clean image of one frame. In addition, some display devices such as LCDs have slow response times; it would take an LCD pixel a few milliseconds before it gradually transits to a new intensity [8]. Thus, in a shot of the display, the region where the new and previous frames meet each other is not a clear line as in Fig. 2 but a wide belt of two frames blended together. Unfortunately,

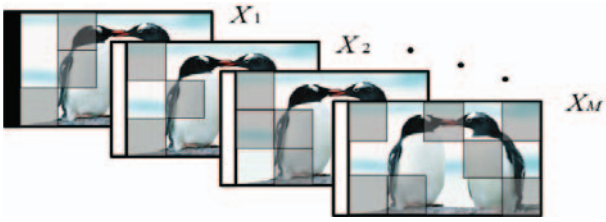


Fig. 4: An indicator signal is added to mark the first atom frame.

any data embedded in the blended region is unrecoverable, as data signal is modulated in the small differences between two consecutive frames while blending removes the differences. Besides these non-ideal properties of common displays, the characteristics of video cameras also contribute to the problem of unrecoverable region. Most cameras use rolling shutters to scan a scene from top to bottom slowly and require a considerable long exposure time to suppress sensor noise. Consequently, the unrecoverable region could often occupy a large portion of the captured display, and in some worst case scenarios, the whole shot is unrecoverable.

Since camera captures at the frame rate roughly the same as the refresh rate of the display, the positions of the unrecoverable regions in successive captured frames are nearly identical, meaning that in a period of time, data embedded by the display at some positions is lost for certain. However, the display is unaware which positions should not be used, because the camera cannot notify this information to the display without involving other communication channel. Thus, to circumvent the problem of unrecoverable region, spatial redundancy is a necessity. To make the spatial redundancy more effective, we also change the scan direction of the camera. For the instance shown in Fig. 2, if we use a camera that scans the display from the bottom to the top, then in the captured image, it is mostly likely that the bottom part of the display shows the last frame and the top part of the display shows the new frame. Both regions are relatively free of the ghosting artifacts from adjacent frames while the middle part is a blend of two frames, which is unrecoverable. Then if we choose to sacrifice data transmission rate and repeat the data embedded in a frame twice, at least one copy of the data in each row is not affected by unrecoverable region. Since almost all mainstream cameras scan from the top to the bottom, to reverse its scan direction, we can simply hold the camera upside down as in Fig. 3.

As illustrated in Fig. 4, to synchronize the data transmission, a thin starting signal block is placed at the left of the display. The starting signal block is black for the first atom frame in a display frame group and white for the rest of the atom frames. The synchronization signal blocks some content of video sequence, but it is thin and does not affect perceptual visual quality too much. Moreover, as shown in Fig. 1c, during the transition from bit 1 to bit 0, there are two consecutive low signals. This special mark can also be used as a supplementary synchronizing signal to indicates the boundary of two bit signals.

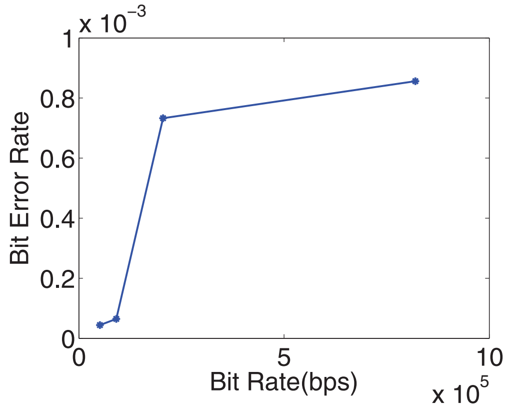


Fig. 5: Smaller superpixel leads to higher bit rate along with higher bit error rate (sizes of superpixel corresponding to the four points in Fig. 5 from left to right are 20×20, 15×15, 10×10 and 8×8 separately).

In some applications, it is acceptable to use every single row of pixels as an independent signal source. Although pixels in the same row do not get refresh at the same time, the time differences are negligible, since the time it takes to activate a horizontal line is extremely short compared to the time to refresh a complete frame. To decode each row, we first traverse rows at the same position from captured frames after affine transformation until we find the starting indicator of this row, then decode the data of this row every 6 frames sequentially. The signal embedded at pixel i in 6 consecutive frames can be decoded as the equation below:

$$data_i = \begin{cases} 0 & \text{if } \vec{b} \cdot (1,1,1)^T < 0 \\ 1 & \text{otherwise,} \end{cases} \tag{2}$$

where $\vec{b} = (b_1, b_2, b_3)$ is the difference vector between adjacent odd atom frame and even atom frame, i.e., $b_m = X_{2m-1} - X_{2m}$ for $m \in \{1, 2, 3\}$. Finally we stitch all rows decoded to get a full data frame.

IV. Experimental Results and Discussions

The proposed technique is evaluated in an indoor environment with a 120 Hz LCD display (ASUS VG278HE) used as the transmitter and a smart phone (Apple iPhone 6) set to capture 120 fps video as the receiver. The smart phone is hand held in landscape mode and oriented so that its image sensor scans from the bottom of a scene to the top.

In theory, every pixel on the display screen can be regarded as an independent signal source, however, due to inevitable camera shake, registration errors and noises, it is difficult to align all pixels accurately, causing high error rate in transmission. To alleviate the problem, we group pixels in each $k \times k$ neighbourhood together to form a superpixel and count the superpixel as a signal source. As plotted in Fig. 5, with smaller size of superpixel, more data bits can be transferred in each frame, however, the bit error rate also increases correspondingly. The bit error rate is affected by the strength δ of the embedded signal as well. As shown in Fig. 6,

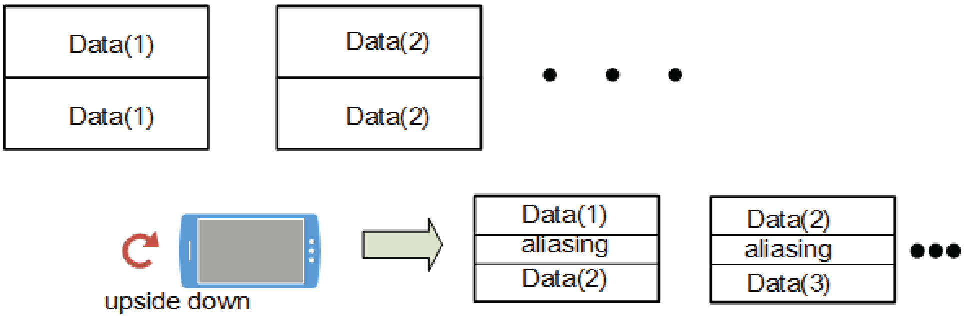


Fig. 3: Reverse the camera to scan each frame from bottom to top

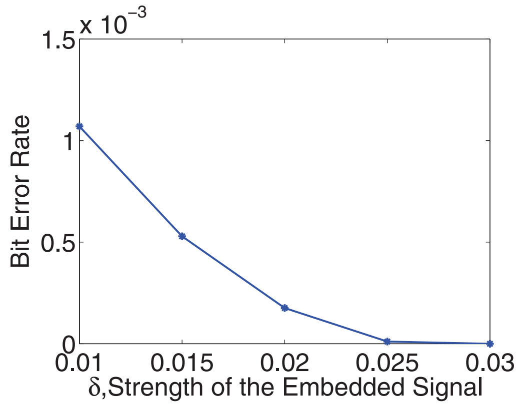


Fig. 6: The higher the strength δ of the embedded signal, the lower the bit error rate.

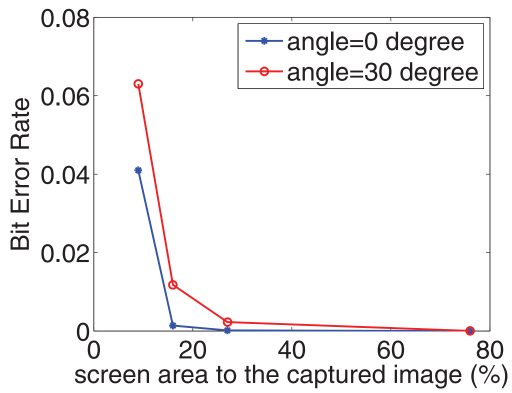


Fig. 7: Bit error rate is lower when the display occupies a large portion of the field of view of the camera and when the camera is right in front of the display

increasing the signal strength makes the embedded data signal easier to detect by the camera hence lowering the bit error rate. However, the drawback is that higher signal strength also makes it easier for human viewers to observe the data signal.

Another practical problem that affects the bit error rate is the distance and angle between the display and camera. If the display and camera are far apart, the display only occupies a small portion of the field of view of the camera, thus the effective spatial resolution of the captured display in the video is low. The angle between display and camera has the similar effect; large angle reduces the effective spatial resolution of the part of the display that is further away from the camera. With the low resolution of the captured display, the embedded signal is more difficult to be detected and recovered hence causing higher bit error rate as demonstrated in Fig 7.

V. Conclusion

In this paper, based on Wu and Shu's work on the topic of hidden wireless data transmission using display and camera, we proposed an improved technique that conceals the data signal more effectively and makes the transmission more robust in realistic environment. Experimental results show that the proposed technique works well in most practical application scenarios.

References

[1] T. Komine and M. Nakagawa, "Integrated system of white LED visible-light communication and power-line communication," *Consumer Electronics, IEEE Transactions on*, vol. 49, no. 1, pp. 71–79, 2003.

[2] X. Liu, H. Makino, and Y. Maeda, "Basic study on indoor location estimation using visible light communication platform," in *Engineering in Medicine and Biology Society*, Aug 2008, pp. 2377–2380.

[3] D. V. Plant and A. G. Kirk, "Optical interconnects at the chip and board level: challenges and solutions," *Proc. IEEE*, vol. 88, no. 6, pp. 806–818, 2000.

[4] X. Huang, B. Chen, and A. Roscoe, "Multi-channel key distribution protocols using visible light communications in body sensor networks," in *Computer Science Student Conference 2012*, 2012, p. 15.

[5] X. Wu and X. Shu, "Combining information display and visible light wireless communication," in *40th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Apr 2015.

[6] M. Kalloniatis and C. Luu, "Temporal resolution," *Webvision: The Organization of the Retina and Visual System*, 2007. [Online]. Available: http://webvision.med.utah.edu

[7] X. Wu and G. Zhai, "Temporal psychovisual modulation: a new paradigm of information display," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 136–141, 2013.

[8] L. Jiao, X. Shu, Y. Cheng, X. Wu, and N. Burini, "Optimal backlight modulation with crosstalk control in stereoscopic display," *Display Technology, Journal of*, vol. 11, no. 2, pp. 157–164, Feb 2015.