Time Synchronization Method Using Visible Light Communication for Smartphone Localization

Takayuki Akiyama
Department of Informatics
SOKENDAI
(The Graduate University for Advanced Studies)
Tokyo, Japan
Email: tak@nii.ac.jp

Masanori Sugimoto
Department of Computer Science
School of Engineering
Hokkaido University
Sapporo, Japan
Email: sugi@ist.hokudai.ac.jp

Hiromichi Hashizume
Information Systems Architecture
Science Research Division
National Institute of Informatics
Tokyo, Japan
Email: has@nii.ac.jp

***Abstract*—We describe a time synchronization technique based on visible light communication (VLC). The precision of time synchronization is the key factor for localization based on time of arrival. Hence, we have proposed the *SyncSync* method using a modulated light-emitting diode (LED) light and a smartphone video camera, which enables time-of-arrival localization by measuring the time of flight of sound waves. This method gives better results than those of time-difference-of-arrival localization. However, we had to use a dedicated light synchronization device for our method. VLC is becoming a popular application of smartphones. If VLC demodulation could be used for time synchronization in acoustic localization, VLC and indoor localization could be integrated into a single application. In this paper, we examine the feasibility of using VLC codes for localization time synchronization. Experimental results show a time synchronization precision equivalent to 1.0 cm for an airborne acoustic wave. This is sufficiently precise for practical applications.**

Keywords–Time synchronization; Visible light communication; Acoustic localization; ToA measurement; Smartphone.

I. INTRODUCTION

Smart devices are being increasingly used for services in which geographical location is an important factor. Numerous location-aware applications have been developed for the popular mobile operating systems and the development costs are low. The key to many of these applications is the rich assortment of smartphone sensors, such as global navigation satellite system or global positioning system (GNSS/GPS) receivers, accelerometers, gravity sensors, and gyroscopes.

In a previous study [1], we introduced an indoor 3D localization system for smartphones. We used the smartphone's microphone to detect acoustic signals and its video camera for time synchronization between sender nodes and the smartphone. We conducted time-of-arrival (ToA) measurements based on the time synchronization, and this offered better performance than did time-difference-of-arrival (TDoA). In our ToA measurements, the standard deviations were less than 10 mm, whereas for TDoA they were 10 mm to 100 mm. In the worst cases, the positioning calculations diverged.

Here, we report our new time synchronization technique based on visible light communication (VLC) [2]. Our experimental results show that the performance of VLC-based synchronization is comparable to that of dedicated systems. As VLC becomes more prevalent for mobile devices, our method will become useful for both communication and localization.

Many of the parameters that are needed for localization, such as transmitter coordinates and the frequencies used for acoustic waves, can be conveyed by VLC. Hence, the application software does not need to refer to an external database for this information, and the system can work without network connectivity. Furthermore, user privacy is protected because no individual user is announced to others. These are some of the advantages of the proposed method.

Our motivation is to realize localization with this scheme. We have verified the concept of the proposed method and found some issues with it. Here, we describe the status of the development. The remainder of the paper is organized as follows. Section II describes our proposed method. Section III presents the results of an experimental investigation of our method. Finally, Section IV gives conclusions and suggestions for future work.

II. PROPOSED METHOD

We describe the proposed method in this section.

A. System Overview

We aim to realize ToA-based acoustic localization that uses VLC for time synchronization. We use a transmitter module that consists of loudspeakers and an LED. The loudspeakers emit short bursts of sound simultaneously, which are registered by a smartphone. The smartphone then calculates the time of flight (ToF) of each burst, and determines its position using ToA-based multilateration. The LED is modulated for VLC, the carrier wave of which is synchronized to the emitted bursts. In the demodulation process, the VLC carrier signal is extracted and used to estimate the emission time; this is the time synchronization of the system. The ToF is the time between burst emission and reception.

B. LED Modulation

Many of the video cameras that are now available use complementary metal-oxide-semiconductor (CMOS) image sensors. An image sensor is a 2D array of photodiodes that record a meshed image as electrical charges. Because most CMOS image sensors read out the horizontal lines of the array one by one, the reconstructed image consists of lines acquired sequentially. This is known as the *rolling shutter* effect, and is exploited by many VLC systems [3][4].

 ISBN: 978-1-61208-514-2

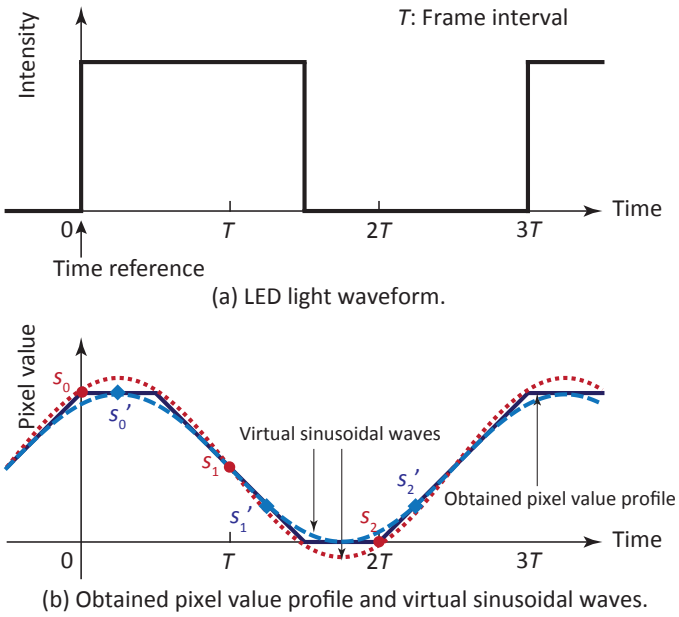

Figure 1. LED light waveform, obtained pixel value profile, and virtual sinusoidal waves.

In a previous study [5], we modulated an LED with a square wave whose frequency was exactly half the frame rate of the video camera. Capturing the modulated light by a CMOS image sensor, we observed a gradation pattern in the obtained image. By assuming that the exposure time was the video-frame period, the profile of the pattern in the shutter-scanning direction becomes part of a triangle wave. We detected either the top or bottom of the profile and used it as a reference point for time synchronization.

We now consider the case in which the modulation frequency of the square wave is one third of the video frame rate (Fig. 1 (a)). The interval of integration of the square wave is the exposure time of the video camera, and the rolling shutter works as a sliding window function. Applying the window function for three consecutive frames (i.e., the period of the LED modulation), we obtain the pixel value profile shown in Fig. 1 (b). This can be approximated as a sinusoidal wave whose phase denotes when the LED is turned on or off.

C. Three-point Demodulation Method

We use the smartphone camera to capture a video of the illuminated LED. In each frame, the group of pixels that contains the LED is regarded as the region of interest (RoI). The *triad* $\boldsymbol{s} = (s_0, s_1, s_2)^{\mathrm{T}}$ is the set of mean RoI pixel values for three consecutive frames. It is used to detect a complex sinusoidal wave

$$f(\theta) = Ae^{j\theta} + Be^{-j\theta} + C, \tag{1}$$

where θ denotes the phase of the sinusoidal wave, A and B express its amplitude, C is the constant component of the wave, and $j = \sqrt{-1}$. Note that the unknown values of A, B, and C are complex. Assigning $\theta = 0$, $2/3\pi$ and $4/3\pi$ in (1), we get a matrix equation

$$\begin{pmatrix} s_0 \\ s_1 \\ s_2 \end{pmatrix} = \begin{pmatrix} f(0) \\ f(\frac{2}{3}\pi) \\ f(\frac{4}{3}\pi) \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ \omega & \omega^2 & 1 \\ \omega^2 & \omega & 1 \end{pmatrix} \begin{pmatrix} A \\ B \\ C \end{pmatrix} = M \begin{pmatrix} A \\ B \\ C \end{pmatrix}, \tag{2}$$

where

$$\omega = \frac{-1 + j\sqrt{3}}{2},$$
$$\omega^2 = \bar{\omega} = \frac{-1 - j\sqrt{3}}{2},$$
$$M = \begin{pmatrix} 1 & 1 & 1 \\ \omega & \omega^2 & 1 \\ \omega^2 & \omega & 1 \end{pmatrix}.$$

Here, $\bar{\cdot}$ denotes the complex conjugate, and ω, ω^2, and $\omega^3 = 1$ are the three cube roots of 1. We can solve (2) as

$$\begin{pmatrix} A \\ B \\ C \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 1 & \omega^2 & \omega \\ 1 & \omega & \omega^2 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \\ s_3 \end{pmatrix} = \frac{1}{3} M^* \begin{pmatrix} s_1 \\ s_2 \\ s_3 \end{pmatrix}. \tag{3}$$

Here, M^* is the conjugate transpose of M.

Because $\boldsymbol{s}$ is a vector whose components are real, $B = \bar{A}$ holds. Then, (1) can be written as

$$f(\theta) = a\cos(\theta + b) + c, \tag{4}$$

where the unknowns a, b, and c are real numbers that satisfy

$$a = 2|A|,$$
$$b = \arg A,$$
$$c = C. \tag{5}$$

The sinusoidal function (4) is not the original square wave that modulated the LED, but it corresponds to it uniquely. Therefore, we refer to (4) as the *virtual sinusoidal wave*.

We set the time reference as the rising edge of the original square wave. Because the LED modulation and the exposure time of the video camera are not synchronized, the time at which the RoI pixel values are sampled varies in each measurement. Two examples of triads, $\boldsymbol{s} = (s_0, s_1, s_2)^{\mathrm{T}}$ and $\boldsymbol{s'} = (s'_0, s'_1, s'_2)^{\mathrm{T}}$, are shown in Fig. 1 (b). Although each triad yields a different sinusoidal wave, we can determine the time reference by using (4) in each case.

D. VLC with Three-point Demodulation Method

We modulate the LED with six square wave patterns. They have the same frequency, amplitude, and mean, but their initial phases are separated by $2/3\pi$. This is a phase-shift keying (PSK) scheme that uses six symbols. The constellation diagram is shown in Fig. 2.

As explained above, a pixel value is the result of integrating the LED output during an exposure. If the exposure time spans the boundary of the LED's lighting pattern, the integration of this term is not suitable for demodulation. To avoid this situation, we introduce a *guard frame* that is similar to the guard interval [6] used in telecommunications. Appended to the original frame, a guard frame is the same lighting pattern as the first frame of the original pattern. Accordingly, the signal length of a symbol becomes four times that of the frame duration. This method ensures that we can select three consecutive frames from a video sequence that have the correct integral values of the original waveform. Fig. 3 shows the waveform for symbol "5", whose phase is $-1/3\pi$.

III. Experiment

We explain an experiment of the proposed method.

 ISBN: 978-1-61208-514-2

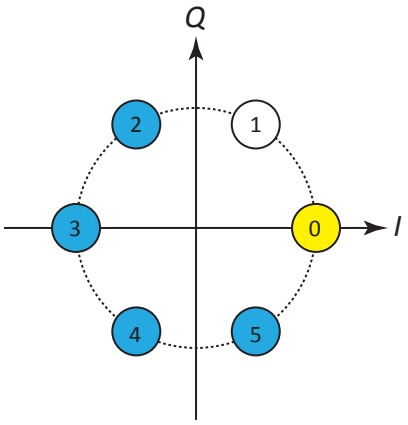

Figure 2. Constellation diagram for 6-PSK.

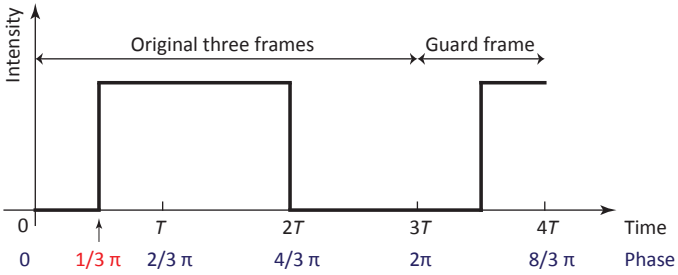

Figure 3. Waveform of symbol "5".

A. Experimental Setup

We conducted an experiment to evaluate the feasibility of the proposed method. Fig. 4 shows a diagram of the experimental system. In the transmitter, an LED floodlight is placed behind a circular translucent window (20 cm in diameter) to emit the modulated light (Fig. 5). The LED floodlight consisted of 56 red LEDs (OS5RKA5111A, Opto-Supply; dominant wavelength: 624 nm). The waveform signal of the LED modulation was generated by an arbitrary function generator (AFG3102, Tektronix). This signal drove a lab-built voltage-to-current converter that supplied electric current to the LED floodlight.

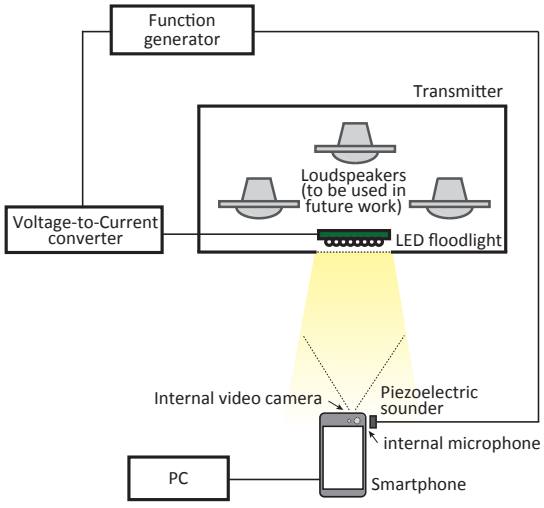

Figure 4. Experimental system.

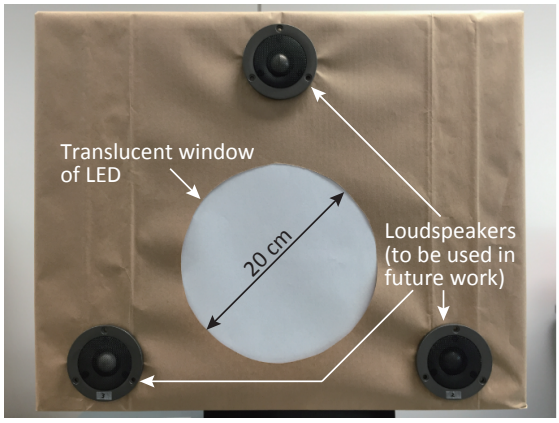

Figure 5. Transmitter.

Three loudspeakers were mounted on the transmitter (Fig. 5) for the purpose of acoustic localization in future work, but they were not used in this experiment. A piezoelectric sounder (PKM13EPYH4000-A0, Murata Manufacturing) was attached directly to the smartphone microphone to evaluate the performance of the smartphone's audio signal processing. Acoustic waves were emitted in 4 ms bursts through the sounder at the beginning of each LED modulation cycle. The phase accordance method [7] was used to determine the precise time of signal reception.

A smartphone (iPhone 6s Plus, Apple) was mounted on a tripod at a distance of 1.0 m, 1.5 m, or 2.0 m from the LED to capture video images. We developed a video-capture application to allow the smartphone to record the pixel values of two RoIs (RoI1 and RoI2). Because they were separated in the shutter-scanning direction, the detected phases of RoI1 and RoI2 were different. The frame rate of the video camera was 60 fps. Although demodulation could be processed with the video-capture application, the recorded pixel values were sent as logging messages and analyzed offline by a personal computer (PC) to evaluate the performance statistically. Timestamps of the video frames were extracted from the sampling buffer of the smartphone through its application programming interface.

Audio signals captured by the smartphone's microphone were recorded alongside the video images as MPEG-4 files that were transferred to the PC. The audio sampling rate was 48.0 ksps. Timestamps of the audio buffer were also extracted and recorded for analysis.

The LED's modulating signal was generated as follows. The signal consisted of a preamble and a payload. The preamble was used to detect the existence of a signal and to find the top frame of a triad. It comprised 12 "0" symbols of the 6-PSK. For the payload, we sent the text message "HELLO, WORLD". Each character in the message was converted to a 6-bit binary code and divided into three 2-bit pieces that were assigned to the symbols "2"–"5". The symbol "1" was not used for data transmission but its existence was used as an integrity check on the receiver side. The payload of the transmission had 36 symbols, meaning that the total length of the signal was 48 symbols. As each symbol occupied four video frames, the duration of the signal was 3.2 s. The signal was transmitted repeatedly, and the video capture duration was 30 s for each measurement.

 ISBN: 978-1-61208-514-2

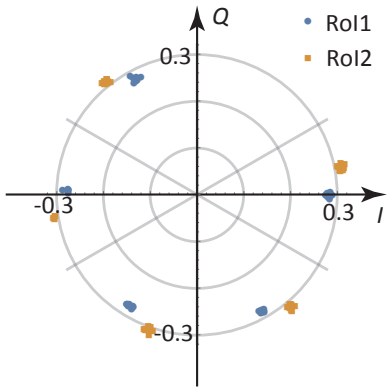

Figure 6. Constellation diagram of the results.

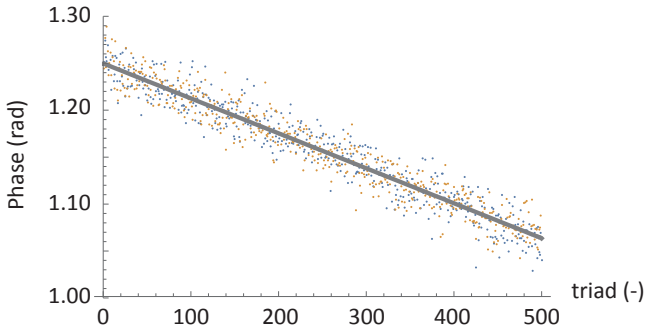

Figure 7. Estimated phase of the triad.

B. Performance of VLC Demodulation and Time Synchronization

The pixel values captured by the smartphone application were demodulated on a PC. We assume a sequence of pixel values $p_i (i = 1, 2, 3, \dots)$ in an RoI. We can make a series of triads as $t_1 = (p_1, p_2, p_3), t_2 = (p_2, p_3, p_4), t_3 = (p_3, p_4, p_5), t_4 = (p_4, p_5, p_6), \dots$. As mentioned before, there is a *right* triad that does not span two symbols in $t_1 \dots t_4$. A right triad appears every four triads. Triads in the preamble are used to select the correct sequence of triads in the payload.

A typical constellation diagram of the results is shown in Fig. 6. Here, the plotting plane is rotated so that the mean phase of RoI1's preamble becomes zero. It can be seen that the symbols are definitely separated, which means that the message was perfectly demodulated. The phase difference between RoI1 and RoI2 is due to the scanning speed of the rolling shutter. Using this speed and the phase of RoI1 or RoI2, we can estimate the phase of the first scanning line of the image buffer (Fig. 7).

When the smartphone was 2.0 m from the LED, the standard deviation of the regression was 0.015 rad, which is equivalent to 4.1 cm for an airborne acoustic wave. We use the standard error here to evaluate the precision of the estimated phase. Within 1 s (i.e., using at least 15 triads), the standard error of the regression becomes 1.1 cm for an airborne acoustic wave. With separations of 1.5 m and 1.0 m, the standard errors are 1.0 cm and 1.7 cm, respectively. The relationship between the separation and the precision, including analysis of the signal-to-noise ratio, should be investigated in detail in future work. Even though our previous study, which used a dedicated modulating waveform, showed a better standard error of 0.17 mm, the proposed system is sufficiently precise and is also applicable to VLC.

The timestamp of the image buffer and the time offset derived from the phase of the first line yield the reference time relative to the buffer's time. The precision of this reference time depends on the timing clock of the video camera. The standard deviation was 1.5 µs in this experiment.

C. Performance of Audio Signal Detection

We extracted the acoustic bursts from the video files and determined their positions on the audio track. The audio buffer size of the smartphone was 4,096 bytes and the timestamp of each buffer was recorded in a logging message. The emission time of each burst was calculated from the timestamp of the buffer to which the burst belonged and from the relative position of the burst in the buffer. The precision of the burst interval was 2 µs to 7 µs.

D. Alignment of Video and Audio Signals

We compared the time references obtained from the video images and audio bursts. Although they should have had the same time, a difference was observed. We intend to investigate this phenomenon in future work.

IV. Conclusion and Future Work

We have confirmed the feasibility of time synchronization using VLC. We successfully transmitted short messages and were able to obtain time references with practicable precision. The audio signals were also detected precisely. However, we were unable to align the video and audio tracks on the time axis. We intend to resolve this issue in order to perform ToA-based acoustic localization using VLC time synchronization.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number 26280036.

References

[1] T. Akiyama, M. Sugimoto, and H. Hashizume, "Syncsync: Time-of-arrival based localization method using light-synchronized acoustic waves for smartphones," in Indoor Positioning and Indoor Navigation (IPIN), 2015 International Conference on, pp. 1–9, Oct 2015.

[2] C. Danakis, M. Afgani, G. Povey, I. Underwood, and H. Haas, "Using a CMOS camera sensor for visible light communication," in 2012 IEEE Globecom Workshops, pp. 1244–1248. IEEE, Dec. 2012.

[3] O. Ait-Aider, N. Andreff, J. Lavest, and P. Martinet, "Exploiting Rolling Shutter Distortions for Simultaneous Object Pose and Velocity Computation Using a Single View," in Fourth IEEE International Conference on Computer Vision Systems (ICVS'06), pp. 35–35. IEEE, 2006.

[4] G. Lepage, J. Bogaerts, and G. Meynants, "Time-Delay-Integration Architectures in CMOS Image Sensors," IEEE Transactions on Electron Devices, vol. 56, no. 11, pp. 2524–2533, Nov. 2009.

[5] T. Akiyama, M. Sugimoto, and H. Hashizume, "Light-synchronized acoustic toa measurement system for mobile smart nodes," in Indoor Positioning and Indoor Navigation (IPIN), 2014 International Conference on, pp. 749–752, Busan, Korea, 2014.

[6] H. Liu and G. Li, OFDM Fundamentals. John Wiley & Sons, Inc., 2006, pp. 13–30. [Online]. Available: http://dx.doi.org/10.1002/0471757195.ch2

[7] H. Hashizume, A. Kaneko, Y. Sugano, K. Yatani, and M. Sugimoto, "Fast and accurate positioning technique using ultrasonic phase accordance method," in Proceedings of IEEE Region 10 Conference (TENCON2005), pp. 1–6, Melbourne, Australia, 2005.

 ISBN: 978-1-61208-514-2