A Demodulation Method Using a Gaussian Mixture Model For Unsynchronous Optical Camera Communication With on-off Keying

Yoshifumi Shiraki, *Member, IEEE*, Takashi G. Sato, *Member, IEEE*, Yutaka Kamamoto, *Senior Member, IEEE*, Tomonori Izumi, *Member, IEEE*, Yu Nakahara, Kyosuke Kondo, and Takehiro Moriya, *Fellow, IEEE*

***Abstract*—We consider optical camera communication (OCC) between a camera receiver with optical lenses and distributed transmitters. This article investigates the features of an OCC system when the periods of reception and transmission are slightly different from each other. We describe a received light signal model for the OCC system with on-off keying and regard the received signals generated from a probability distribution of a Gaussian mixture model. We obtain the parameters of the probability distributions by applying a variational Bayesian inference method and utilize them for channel estimation. In addition, we define cost functions and minimize them to demodulate the transmitted bit sequences. The demodulation procedure uses a maximum-likelihood sequence detection method, which can be implemented by the Viterbi algorithm and estimates a synchronization parameter by minimizing the cost functions. Our new demodulation method requires neither synchronization devices nor training sequences for estimating the parameters. Moreover, the receiver does not need the precise transmission period, which is difficult to know in advance in practical situations because of the frequency tolerance of the clock generator in the transmitter. To validate our developed method, we conducted numerical simulations and compared the results with those from an oracle estimator that knows the parameters other than the bit sequence in advance. We also experimented in a real setup situation, and the results show the efficiency of our developed method.**

***Index Terms*—Optical camera communication, Gaussian mixture model, clock synchronization, maximum-likelihood sequence detection method.**

I. Introduction

VISIBLE light communication (VLC) is a research area that has emerged mainly due to the widespread use of light-emitting diodes (LEDs). Indoor VLC, which uses room lights from LEDs as a transmitter, can provide a fast and secure communications system because of the massive number of LEDs and controllability of the illumination area [1]–[4]. In addition, great effort has been made to apply VLC to intelligent traffic systems (ITSs) for safer driving and more efficient traffic navigation [5]–[9]. Since both of the above applications will utilize LEDs that are already available for room and automotive lighting and traffic lights, installation costs are expected to be low compared to conventional radiofrequency (RF) wireless communications systems. Furthermore, as visible light currently has no legal restrictions involving bandwidth allocation in many countries, we can freely use VLC in many situations.

One type of VLC is optical camera communication (OCC). OCC employs a camera as a receiver and gathers data from distributed devices that have LEDs or displays as transmitters [10]–[12]. Recent advances in cameras with image sensors (ISs) have opened up wide possibilities for OCC systems. Thanks to the receiver's optical lenses, emitted lights are spatially separated on the IS and the massive number of pixels enables the camera receiver to detect the positions of objects. This enables the receiver to trace moving transmitters and communicate with multiple transmitters simultaneously. By taking advantage of these features, OCC systems have been applied to vehicle-to-infrastructure and vehicle-to-vehicle communications in ITSs [5]–[9], positioning systems [13], [14], multichannel acoustic measurement for beamforming via optical wireless microphones [15], [16] and simultaneous biosignal observation of audiences in live concerts [17].

One major issue facing OCC systems is the limited sampling period, i.e., the reception period of ISs compared to that of photodetectors. This limitation complicates synchronization between the receiver camera and transmitters [18]. Undersampled frequency shift on-off keying [19] and undersampled phase shift on-off keying [20] are sophisticated methods to overcome this synchronization issue; however, these methods utilize very short-time exposure time and thus need sufficiently bright LEDs as transmitters, which is unfavorable in terms of power consumption. Pablo *et al.* adopted an infrared interface that broadcasts the master clock to the transmitters to synchronize the shutter timing of the camera receiver and the bit transition timing of the LED transmitters [15]; however, the additional interface is not

Manuscript received February 26, 2020; revised June 24, 2020; accepted July 18, 2020. Date of publication December 8, 2020; date of current version March 16, 2021. *(Corresponding author: Yoshifumi Shiraki.)*

Yoshifumi Shiraki, Takashi G. Sato, Yutaka Kamamoto, and Takehiro Moriya are with NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation, Atsugi 243-0198 Japan (e-mail: yoshifumi.shiraki.ya@hco.ntt.co.jp; takashi_goto_sato@ieee.org; ytk@ieee.org; t.moriya@m.ieice.org).

Tomonori Izumi is with the Department of Electronic and Computer Engineering, Ritsumeikan University, Kusatsu-shi 525-8577, Japan (e-mail: t-izumi@se.ritsumei.ac.jp).

Yu Nakahara and Kyosuke Kondo are with the Department of Electronic and Computer Engineering, Ritsumeikan University, Kusatsu-shi 525-8577, Japan (e-mail: ml-rsl17nakahara@ml.ritsumei.ac.jp; ml-rsl17kondo@ml.ritsumei.ac.jp).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JLT.2020.3043046.

Digital Object Identifier 10.1109/JLT.2020.3043046

desirable in terms of the fabrication cost and power consumption of transmitter devices. Ohshima *et al.* [21] developed an effective synchronization method for OCC with rolling shutter ISs. Their method enables very fast communication using a conventional rolling shutter IS embedded in a smartphone; however, the simultaneous communication between multiple transmitters is difficult to implement with this synchronization method because it assumes that a transmitter is mapped on the wide area of the IS. Mao *et al.* [22] proposed a method that can demodulate received signals in OCC without any synchronization devices even when the sampling and transmission periods are different. However, to estimate the background noise and the synchronization parameter, the method needs a training sequence. This additional non-informational sequence reduces the bit transmission efficiency. Taking into account these issues, we suggested a method that requires neither synchronization devices nor training sequences [23]. However, this simple method does not work well in environments with non-negligible background noise.

In this study, we developed a demodulation method for an OCC system that can estimate the background noise level and synchronization parameters without any synchronization devices and training sequences. To achieve this, we employed the variational Bayesian inference method for a Gaussian-mixture model [24] for the channel estimation of the OCC system. By applying the inference method to received signals, we can estimate both the background noise level and the gain coefficient for demodulation. In addition, as in [23], the developed method does not require precise knowledge of the bit transition periods of the transmitters. This aspect is practically important because it is sometimes difficult to know them in advance because of the frequency tolerance of the clock generators in the receiver and transmitters.

This paper is organized as follows. In Section II, we describe notations, assumptions, and the modeling of the light signal at the receiver camera in an OCC system. Next, we provide a parameter estimation and demodulation procedure in Section III. In Sections IV and V, to confirm the efficiency of our demodulation method, we show the setups for and the results of numerical simulations and experiments in a real environment. In Section VI, we present our conclusions.

Note that Figs 1 to 5 were originally presented in our conference paper [23]. This work is an expansion of that study.

II. Optical Camera Communication Model With On-Off Keying

Before focusing on the demodulation problem, we describe the channel model discussed throughout this paper. In general, one can model an OCC system with a camera receiver and LED transmitters as a multiple-input and multiple-output (MIMO) communications system, where multiple pixels on an IS of the camera receive lights emitted from the multiple LED transmitters. However, we can simplify the model by stating some assumptions. First, we assume that there is a sufficiently large number of pixels in the IS compared to the number of transmitters in the field of view of the camera. In such a situation, the images of the different transmitters fall on different pixels in the IS of the camera receiver when the transmitters are almost in focus. Second, we assume the camera drives in global shutter mode. Thus all pixel on the IS are charged for the same time range. This enables us to deal with the pixels indexed by Ω_l where the image of the l th transmitter falls on as one set of pixels (see Fig. 1). With these two assumptions, one can consider each transmitter independently and regard the OCC system as a collection of single-input and single-output (SISO) communication systems. This spatial separability allows the OCC system to accommodate multiple transmitters and this type of multiplexing is called spatial-division multiplexing [25].

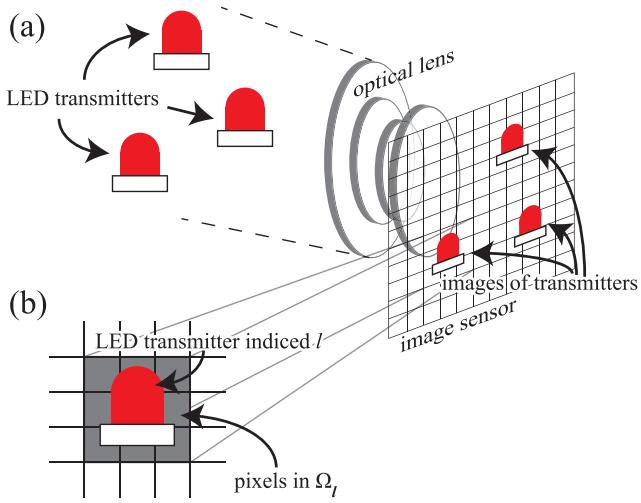


Fig. 1. (a) Diagram of an optical camera communication system: the images of the different transmitters fall on different pixels in the image sensor thanks to the optical lens of the camera receiver. (b) Enlarged view around the pixels on which the image of the lth sensor falls. The gray pixels are in the pixel indices set Ω_l.

Here we describe a model of one SISO communication system in the OCC system. Let $y(t)$ be instantaneous light power emitted from the transmitter at continuous time t and let t_0, T_{RX} and τ be the time offset, sampling period, and exposure time of the camera, respectively. Then the received signal value $s[i]$ obtained by summing over a pixel set at discrete time index i is described as,

$$s[i] = \int_{t_0+T_{\mathrm{RX}}(i-1)}^{t_0+T_{\mathrm{RX}}(i-1)+\tau} h[i]y(t)\mathrm{d}t + d[i] + n[i], \qquad (1)$$

where $h[i]$ represents an attenuation coefficient between the transmitter and a corresponding pixel set. The $d[i]$ denotes a background noise level which corresponds to the signal value when none of the transmitters emit light and $n[i]$ denotes the noise, including shot noise from ambient light and thermal noise at the pixel amplifier.

While many sophisticated modulation methods, such as undersampled frequency shift on-off keying [19] and undersampled phase shift on-off keying [20], have been applied to OCC, here, we assume simple on-off keying (OOK) as a basic example. In OOK, the instantaneous light power $y(t)$ is represented as,

$$y(t) = \sum_j Ax[j]g\left(t-(j-1)T_{\mathrm{TX}}\right), \qquad (2)$$

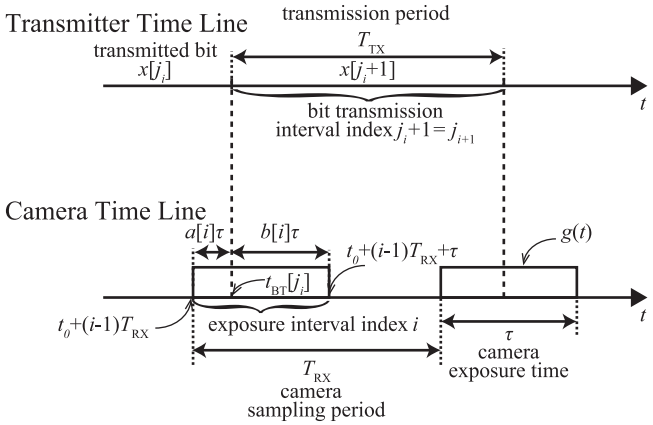


Fig. 2. Diagram of a transmitter and a camera receiver time lines. The bit transition from $x[j_i]$ to $x[j_i+1]$ occurs during the exposure interval i at time T_{BT}.

where A, $x[j] \in \{0,\ 1\}$ and T_{TX} are the maximum optical power, the transmitted bit information and the transmission period, respectively. The function $g(t)$ represents the pulse amplitude of the transmitter, and hereafter we assume that it is the rectangular function

$$g(t) = \begin{cases} 1, & 0 \le t < T_{\mathrm{TX}} \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

Combining (2) and (3), we can rewrite (1) as,

$$s[i] = R[i]\left(a[i]x[j_i] + b[i]x[j_i+1]\right) + d[i] + n[i], \tag{4}$$

where $R[i] = Ah[i]\tau$ and j_i are a gain coefficient and the index of the first transmitted bit in the ith exposure interval, respectively. The terms $a[i]$ and $b[i]$ are linear combination coefficients representing a bit transition that possibly occurs during the camera exposure time as illustrated in Fig. 2. In concrete terms, let $t_{\mathrm{BT}}[j_i]$ be the time of the bit transition from $x[j_i]$ to $x[j_i+1]$. If $t_{\mathrm{BT}}[j_i] < t_0 + (i-1)T_{\mathrm{RX}} + \tau$ is satisfied, the exposure time $t_{\mathrm{BT}}[j_i] - (t_0 + (i-1)T_{\mathrm{RX}})$ is devoted to obtaining the light signal of the first bit $x[j_i]$ and the remaining exposure time $t_0 + (i-1)T_{\mathrm{RX}} + \tau - t_{\mathrm{BT}}[j_i]$ is devoted to obtaining that of the second bit $x[j_i+1]$. By normalizing these two exposure times by τ, we have $a[i]$, $b[i]$ and $a[i] + b[i] = 1$. The linear combination coefficients $a[i]$ and $b[i]$ reflect the synchronization status between the receiver and transmitter.

The quality of demodulation in the OCC system using OOK depends on the synchronization status. Thus the relationship between periods and phases of exposure and bit transmission is important. First of all, as Fig. 3 shows, $T_{\mathrm{RX}} - \tau < T_{\mathrm{TX}}$ should hold to ensure that the camera receiver samples every transmission bit $x[j]$. If the periods and the phases are precisely locked, namely $T_{\mathrm{TX}} = T_{\mathrm{RX}}$ and $a[i] = 1$ hold, the OOK signals can be easily demodulated by, for example, thresholding the received signals $s[i]$ sequentially. In our application, however, the phases are not locked while T_{TX} and T_{RX} are nominally the same. Thus, $a[i]$ and $b[i]$ gradually vary with time. We call this synchronization status the plesiochronous mode as in [22].

III. A Flexible Demodulation Method for Plesiochronous OCC System

In this section, we introduce a new demodulation method for the plesiochronous OCC system. In Section III-A, we provide a basic strategy to develop the demodulation method. After that, we describe algorithms to estimate the parameters in Sections III-B and III-C.

A. Strategy to Estimate Parameters

We clarify which parameters of the current signal model in (4) should be estimated to obtain the bit sequence. Under the assumption that noise $n[i]$ is generated from a Gaussian distribution with a zero mean and σ_{n}^2 variance, the conditional probability density of the received signal $\mathbf{s} = (s[i], s[i+1], \ldots, s[i+l_{\mathrm{p}}-1])$, given the transmission bit sequence $\mathbf{x} = (x[j_i], x[j_i+1], \ldots, x[j_{i+l_{\mathrm{p}}-1}+1])$, is described as,

$$p(\mathbf{s}|\mathbf{x}) = \frac{1}{(\sqrt{2\pi\sigma_{\mathrm{n}}})^{l_{\mathrm{p}}}} \prod_{i\in\Omega} \exp\Bigg[-\frac{1}{2\sigma_{\mathrm{n}}^2}\{s[i] - d[i] - R[i]\left(a[i]x[j_i] + b[i]x[j_i+1]\right)\}^2 \Bigg], \tag{5}$$

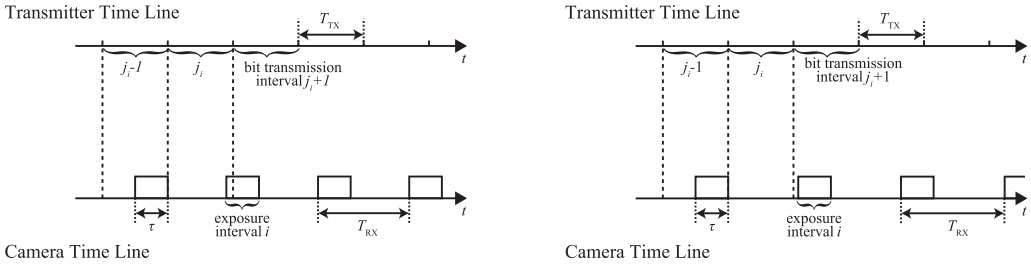
where Ω denotes a set of indices $i, i+1, \ldots, i+l_{\mathrm{p}}-1$. Given a received signal $\mathbf{s}$, we estimate the transmitted bit sequence by finding $\mathbf{x}$ that minimizes the cost function

$$\Gamma = \sum_{i\in\Omega} \left(s[i] - d[i] - R[i]\left(a[i]x[j_i] + b[i]x[j_i+1]\right)\right)^2. \tag{6}$$

Maximum-likelihood sequence detection (MLSD) implemented using the Viterbi algorithm allows us to perform this estimation [22], [26], [27]. To do this, we should estimate $d[i]$, $R[i]$, $a[i]$ and $b[i]$ in advance. Note that the order of computational complexity of the current Viterbi algorithm for solving the MLSD is linear to the size of Ω. That is because the trellis diagram that the Viterbi algorithm aims to solve has no loop but just correlations between two consecutive lattices.

In our application, the channel state between the camera receiver and the LED transmitters changes gradually. In such a case, $d[i]$ and $R[i]$ are supposed to be constant for a large number of consecutive frames, and that allows us to use a large number of received signals to estimate the coefficients. Taking advantage of this, we introduce in III-B a method which utilizes statistical properties of the received signal without explicitly determining the bit sequence to estimate the channel state. We call this estimation method long-term processing.

In the plesiochronous mode, coefficients $a[i]$ and $b[i]$ also change gradually, and we can use several received signals to estimate them. If the rate of the change is similar to that of the channel state, we might also estimate $a[i]$ and $b[i]$ by a statistical method. However, the clock differences between the receiver and transmitters, which determine the stability of $a[i]$ and $b[i]$, are not always sufficiently small to employ a statistical method. To combat this, in III-C, we employ suboptimal per-survivor processing (PSP) [28] to estimate the coefficients. We refer to this estimation method as short-term processing.



(a) $T_{\mathrm{RX}} - \tau < T_{\mathrm{TX}}$ (b) $T_{\mathrm{RX}} - \tau \geq T_{\mathrm{TX}}$

Fig. 3. When $T_{\mathrm{RX}} - \tau < T_{\mathrm{TX}}$ holds, all bits are sampled by the receiver at least once. On the other hand, when the inequality is not satisfied, some bits are not sampled [e.g. bit $x[j_i]$ is not sampled in (b)].

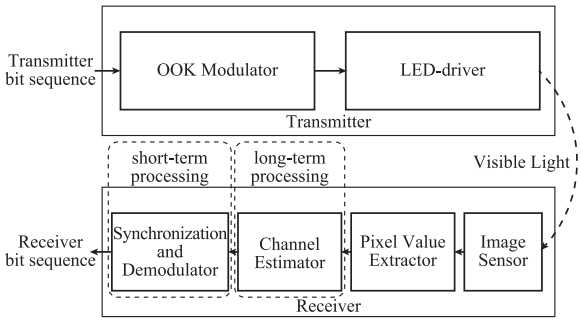


Fig. 4. A block diagram of our optical camera communication system.

Fig. 4 shows a block diagram of our OCC system. The transmitters send bit sequences with OOK modulation and the receiver extracts the signals from its image sensor. After that, the receiver synchronizes and demodulates the received signal with the channel state estimated by the long-term processing. Note that we only illustrate one of each unit in the receiver; however, in actual situations, we can introduce multiple units other than the image sensor to treat multiple transmitters.

B. Estimating Channel State by Variational Bayesian Inference Method

Here we consider how the receiver can estimate the background noise level $d[i]$ and the gain coefficient $R[i]$. If $d[i]$ and $R[i]$ are constant for frames with a long-term processing length l_{long}, we can jointly utilize received signals $\{s[i]\}_{i\in\Omega_{\mathrm{long}}}$ to estimate the coefficients, where Ω_{long} is a set of consequential received signal indices and $|\Omega_{\mathrm{long}}| = l_{\mathrm{long}}$. When l_{long} is a sufficiently large, statistical properties of the received signals could help in the estimation.

Let d and R be values of $\{d[i]\}_{i\in\Omega_{\mathrm{long}}}$ and $\{R[i]\}_{i\in\Omega_{\mathrm{long}}}$. Then, the values of the received signals are written as

$$s[i] = \begin{cases} s_{00}[i] = d + n[i], \text{ if } x[j_i] = x[j_i + 1] = 0 \\ s_{10}[i] = Ra[i] + d + n[i], \text{ if } x[j_i] = 1,\ x[j_i + 1] = 0 \\ s_{01}[i] = Rb[i] + d + n[i], \text{ if } x[j_i] = 0,\ x[j_i + 1] = 1 \\ s_{11}[i] = R + d + n[i], \text{ if } x[j_i] = x[j_i + 1] = 1. \end{cases}$$

We assumed that noise $n[i]$ is generated by a zero-mean Gaussian distribution. Thus $s_{00}[i]$s and $s_{11}[i]$s follow Gaussian distributions $\mathcal{N}(d, \sigma_{\mathrm{n}}^2)$ and $\mathcal{N}(R + d, \sigma_{\mathrm{n}}^2)$. On the other hand, the distributions of $s_{01}[i]$s and $s_{10}[i]$s depend on the properties of $a[i]$ and $b[i]$. Let the distribution of $s_{10}[i]$ be a superposition of the Gaussian distributions of means $a[i]$ and the distribution of $s_{01}[i]$ be a superposition of the Gaussian distributions of means $b[i]$. When the changes in $a[i]$ and thus $b[i]$ for $i \in \Omega_{\mathrm{long}}$ are negligibly small, we can assume that $s_{01}[i]$s and thus $s_{10}[i]$s for $i \in \Omega_{\mathrm{long}}$ are generated from the Gaussian distribution of mean $\bar{a} \simeq a[i], i \in \Omega_{\mathrm{long}}$ and $\bar{b} \simeq b[i], i \in \Omega_{\mathrm{long}}$. On the other hand, when the changes in $a[i]$ and $b[i]$ aren't so small, the distribution of $s_{10}[i]$s resembles a continuous uniform distribution between the maximum value $a_{\max} = \max_{i\in\Omega_{\mathrm{long}}} a[i]$ and the minimum value $a_{\min} = \min_{i\in\Omega_{\mathrm{long}}} a[i]$ with additional tails on both sides due to the Gaussian noise. Similarly, the distribution of $s_{01}[i]$s resembles a continuous uniform distribution between the maximum value $b_{\max} = \max_{i\in\Omega_{\mathrm{long}}} b[i]$ and the minimum value $b_{\min} = \min_{i\in\Omega_{\mathrm{long}}} b[i]$ with additional tails on both sides due to the Gaussian noise. Although it might be difficult to approximate the distribution of the signals in the latter case by GMM, in this paper we assume that the distribution of $\{s[i]\}_{i\in\Omega_{\mathrm{long}}}$ follows a linear superposition of Gaussian distributions of means d, $R + d$, $R\bar{a} + d$ and $R\bar{b} + d$. This is because the role of GMM is not just an approximation of the true generation distribution of the received signals but also has the capability to cluster them and this enables us to remove the negative effect of s_{01}s and s_{10}s to estimate $R + d$ and d using s_{11} and s_{00}, respectively.

One can estimate d and R by determining the means of the Gaussian distributions. To do this, we apply the variational Bayesian inference algorithm for the Gaussian mixture model (VB-GMM) [24] to the received signals. In this inference algorithm, one aims to find approximations for the posterior distribution $p(Z, \pi, \mu, \mathbf{\Lambda}|S)$ and the model evidence $p(S)$ by considering the joint distribution $p(S, Z, \pi, \mu, \mathbf{\Lambda})$, where $S = \{s_1, s_2, \ldots, s_N\}$, $Z = \{z_1, z_2, \ldots, z_N\}$, $\pi = \{\pi_k\}_{k=1}^K$, $\mu = \{\mu_k\}_{k=1}^K$, and $\mathbf{\Lambda} = \{\Lambda_k\}_{k=1}^K$ are the set of observed variables, the set of latent variables, the mixture coefficients, the set of means and the set of inverses of variances of the K-component Gaussian distribution, respectively. All of the variable are in $\mathbb{R}^{1\times 1}$. In our case, the observed variables are the values of received signals minus their mean. The latent variables have no physical meaning, but they are used for convenience. The

mixture coefficients $\pi = \{\pi_k\}_{k=1}^K$ are expected to represent the ratio of s_{00}s, s_{01}s, s_{10}s and s_{11}s in the received signals. $\mu = \{\mu_k\}_{k=1}^K$ and $\mathbf{\Lambda} = \{\Lambda_k\}_{k=1}^K$ are expected to be means and inverses of variances of s_{00}s, s_{01}s, s_{10}s and s_{11}s, respectively. We can decompose the log marginal probability as

$$\ln p(S) = \mathcal{L}(q) + \mathrm{KL}(q||p), \tag{7}$$

where we defined

$$\mathcal{L}(q) = \int q(Z, \pi, \mu, \mathbf{\Lambda}) \ln\left\{\frac{p(S, Z, \pi, \mu, \mathbf{\Lambda})}{q(Z, \pi, \mu, \mathbf{\Lambda})}\right\} \mathrm{d}Z\mathrm{d}\pi\mathrm{d}\mu\mathrm{d}\mathbf{\Lambda}, \tag{8}$$

$$\mathrm{KL}(q||p) = -\int q(Z, \pi, \mu, \mathbf{\Lambda}) \ln\left\{\frac{p(Z, \pi, \mu, \mathbf{\Lambda}|S)}{q(Z, \pi, \mu, \mathbf{\Lambda})}\right\} \mathrm{d}Z\mathrm{d}\pi\mathrm{d}\mu\mathrm{d}\mathbf{\Lambda}, \tag{9}$$

and $q(Z, \pi, \mu, \mathbf{\Lambda})$ are inferred probability distributions. Instead of directly minimizing the KL divergence in (9), we maximize the lower bound (8) with respect to the restricted family of distributions $q(Z, \pi, \mu, \mathbf{\Lambda})$ in a step-by-step manner to obtain estimations of parameters π, μ, and $\mathbf{\Lambda}$. By applying the VB-GMM algorithm to the received signal $\{s[i]\}_{i\in\Omega_{\mathrm{long}}}$, we can estimate the means d, $R + d$, $R\bar{a} + d$ and $R\bar{b} + d$.

To estimate the parameters precisely, we utilize properties of $\bar{a}$ and $\bar{b}$ and determine the number of components K in the generative distribution of $\{s[i]\}_{i\in\Omega_{\mathrm{long}}}$. There can be three cases: In one, when the values of $\bar{a}$ and $\bar{b}$ are different, there are four Gaussian components whose means are d, $R + d$, $R\bar{a} + d$ and $R\bar{b} + d$. In another, when $\bar{a}$ and $\bar{b}$ take nominally the same value and but not zero, three Gaussian components are enough to express the generative distribution. In the other, when $\bar{a} = \bar{b} = 0$ approximately holds, we need only two Gaussian components to fit the generative distribution. Unfortunately, we do not know the values of $\bar{a}$ and $\bar{b}$ in advance of the estimation and can not determine the number of components without any additional information. o combat this, we evaluate lower bound (8) with additional factor $\mathcal{L}(q^*) + \ln K!$ [24] for $K = 2, 3, 4$ cases, where q^* represents the inference probability estimated by VB-GMM, and we adopt the number of components that minimizes this value. Then we obtain the minimum and maximum means as the estimations of the background noise $\tilde{d}$ and the gain coefficient $\tilde{R}$, respectively. We summarized the estimation algorithm in algorithm III.1, where $\psi(\cdot)$ is the digamma function.

Here we explain the computational cost of VB-GMM. As described in [29], the order of the computational complexity of an iteration of the VB-GMM is $\mathcal{O}(lKd_{\mathrm{data}}^2 + KD_{\mathrm{data}}^3)$, where l, K and D_{data} are the number of input data, the number of the Gaussian components and the dimension of the data vector, respectively. In the current case, this becomes $\mathcal{O}(l_{\mathrm{long}}K + K)$ and thus the number of input data l_{long} is a leading factor. Furthermore, we should be careful about the number of iterations for convergence of the estimation. When these factors are considered, a concern might be that the computational time of the estimation could be somewhat too long for practical applications. In Section V, we review the computational time of the whole estimation process to dispel this concern.

Algorithm III.1: Long-Term Processing: VB-GMM Based Algorithm for Channel Estimation.

Require: $\{s[i]\}_{i\in\Omega_{\mathrm{long}}}$, α_0, β_0, m_0 and W_0
Set $N = |\Omega_{\mathrm{long}}|$ and $\mathbf{s} = \{s[i]\}_{i\in\Omega_{\mathrm{long}}} - \Sigma_{i=1}^N s[i]/N$,
for $K = 2, \ldots, 4$ **do**
  Set initial values $\alpha = \{\alpha_k\}_{k=1}^{k=K}$, $\beta = \{\beta_k\}_{k=1}^{k=K}$, $\mathbf{m} = \{m_k\}_{k=1}^{k=K}$ and $\mathbf{W} = \{W_k\}_{k=1}^{k=K}$
  while convergence criteria is not satisfied **do**
    for $k = 1, \ldots, K$ **do**
      $\ln\tilde{\Lambda}_k = \psi(\nu_k/2) + \ln 2 + \ln|W_k|$
      $\ln\tilde{\pi}_k = \psi(\alpha_k) - \psi(\hat{\alpha})$
      for n = 1, 2, ..., N **do**
        $r_{nk} = \tilde{\pi}_k\tilde{\Lambda}_k^{-1/2} \times \exp\{-1/2\beta_k - \nu_k W_k/2(s[n] - m_k)\}$
      end for
      for $n = 1, 2, \ldots, N$ **do**
        $r_{nk} = \frac{r_{nk}}{\sum_{m=1}^N r_{mk}}$
      end for
    end for
    for $k = 1, \ldots, K$ **do**
      $N_k = \sum_{n=1}^N r_{nk}$
      $\bar{s}_k = \frac{1}{N_k}\sum_{n=1}^N r_{nk}s[n]$
      $S_k = \frac{1}{N_k}\sum_{n=1}^N r_{nk}(s[n] - \hat{s}_k)^2$
      $\alpha_k = \alpha_0 + N_k$
      $\beta_k = \beta_0 + N_k$
      $m_k = (\beta_0 m_0 + N_k\bar{\mathbf{s}})/\beta_k$
      $W_k^{-1} = W_0^{-1} + N_K + \frac{\beta_0 N_k}{\beta_0 + N_k}(\bar{s}_k - m_0)^2$
      $\nu_k = \nu_0 + N_k$
    end for
    Estimate the parameters with
    Sort the index k as μ_k be in descending order
  end while
  Evaluate the lower bound
  Store the estimations as $\hat{\mu}_k(K) = m_k$ for every k
  Store the lower bound with additional factor $\mathcal{L}(q^*) + \ln K!$
end for
Determine $\hat{K}$ corresponding to the minimum of the stored lower bound with additional factor
Output the estimations as $\tilde{d} = \hat{\mu}_1(\hat{K})$ and $\tilde{R} = \hat{\mu}_{\hat{K}}(\hat{K}) - \hat{\mu}_1(\hat{K})$

C. Estimating the Synchronization Parameter and Bit Sequence

Here we consider how the receiver can estimate synchronization parameter $a[i]$, $b[i]$ and the bit sequence $x[i]$'s. To do this, we firstly formulate a cost function to be minimized for the estimation, especially when $T_{\mathrm{RX}} < T_{\mathrm{TX}}$ holds. Since we assume that the relational difference between the periods is small, we can expect that $a[i]$ and $b[i] = 1 - a[i]$ vary gradually as Fig. 5 shows. Let Ω_{short} be a range of time indices in which

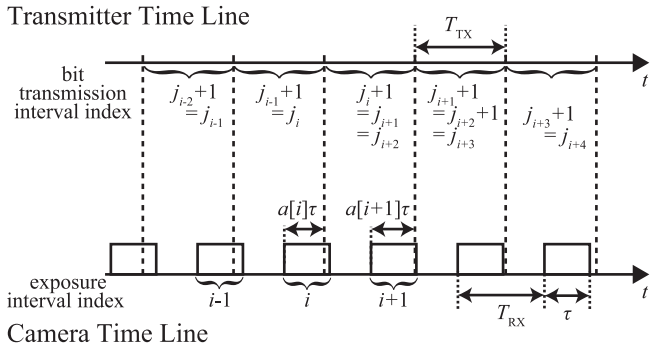


Fig. 5. When $T_{\mathrm{RX}} < T_{\mathrm{TX}}$ holds, $a[i]$ as well as a_{c} continue to increase. On the other hand, when $T_{\mathrm{RX}} > T_{\mathrm{TX}}$ holds, $b[i]$ as well as a_{c} continue to increase. In these figures, one can find that a_{c} and the bit transmission index are reset in the exposure interval indexed $i+2$.

the variation is infinitesimally small and a_{c} be the average value of $\{a[i]\}_{i\in\Omega_{\mathrm{short}}}$. Then we obtain the cost function as,

$$\Gamma_+(\mathbf{x}, a_{\mathrm{c}}) = \sum_{i\in\Omega_{\mathrm{short}}} \left[s[i] - \left\{\tilde{R}\left(a_{\mathrm{c}}x[j_i] + (1-a_{\mathrm{c}})x[j_i+1]\right) + \tilde{d}\right\}\right]^2, \tag{10}$$

where $0 \le a_{\mathrm{c}} < 1$ denotes the fraction of the exposure time spent to obtain the light signal of the first bit and $\tilde{R}$ and $\tilde{d}$ are estimated values of R and d, respectively. When $a_{\mathrm{c}} = 0$ holds, the sampling timing is nominally synchronous with the bit transmission timing in the range Ω_{short}, while they are not in the synchronization state when $a_{\mathrm{c}} > 0$.

It is not easy to exactly minimize the cost function to estimate the above parameters. Therefore, we suboptimally minimize it in a stepwise manner instead. First, we estimate $\mathbf{x}$ by the MLSD method with tentative parameter a_{c}. After that, a_{c} is estimated in the PSP manner Thus, we solve $\mathrm{d}\Gamma_+/\mathrm{d}a_{\mathrm{c}} = 0$ using the tentative bit sequence $\hat{\mathbf{x}}$ and obtain,

$$a_{\mathrm{c}} = \frac{\sum_{i\in\Omega_{\mathrm{short}}}(\hat{x}[j_i+1] - \hat{x}[j_i])(\hat{x}[j_i+1] - (s[i] - \tilde{d})/\tilde{R})}{\sum_{i\in\Omega_{\mathrm{short}}}(\hat{x}[j_i+1] - \hat{x}[j_i])^2}. \tag{11}$$

In practice, we initially use 0 as the tentative a_{c} or try various candidate values and choose one that minimizes (10) as in [22].

Algorithm III.2: Short-Term Processing: Synchronization and Demodulation

Require: $\tilde{R}$, $\tilde{d}$, $\mathbf{s} = \{s[i]\}_{i\in\Omega_{\mathrm{short}}}$, a_{th}, and the number of iterations to minimize the cost function N_{iter}

```
Set â_c be the current estimate a_c
for n = 1, ..., N_iter do
  if T_TX > T_RX then
    x̂ = arg min_x Γ_+(x, â_c, s, R̃, d̃)
    â_c = arg min_{a_c} Γ_+(x̂, a_c, s, R̃, d̃)
  else
    x̂ = arg min_x Γ_−(x, â_c, s, R̃, d̃)
    â_c = arg min_{a_c} Γ_−(x̂, a_c, s, R̃, d̃)
  end if
end for
if â_c > a_th then
  â_c = 0
  Reset index
end if
```

Usually, we iteratively apply the above procedure to improve the estimation quality.

When $T_{\mathrm{RX}} \ge T_{\mathrm{TX}}$ holds, we consider the following cost function:

$$\Gamma_-(\mathbf{x}, a_{\mathrm{c}}) = \sum_{i\in\Omega_{\mathrm{short}}} \left[s[i] - \left\{\tilde{R}\left((1-a_{\mathrm{c}})x[j_i] + a_{\mathrm{c}}x[j_i+1]\right) + \tilde{d}\right\}\right]^2, \tag{12}$$

In this case, $b[i]$ and thus a_{c} gradually increases in the range of 0 to 1, as Fig. 5 shows. As in the $T_{\mathrm{RX}} < T_{\mathrm{TX}}$ case, we can estimate the parameters by,

$$a_{\mathrm{c}} = \frac{\sum_{i\in\Omega_{\mathrm{short}}}(\hat{x}[j_i] - \hat{x}[j_i+1])(\hat{x}[j_i] - (s[i] - \tilde{d})/\tilde{R})}{\sum_{i\in\Omega_{\mathrm{short}}}(\hat{x}[j_i+1] - \hat{x}[j_i])^2}, \tag{13}$$

and the transmitted bit sequence $\mathbf{x}$ by individually minimizing the cost function for each of them.

Thanks to the small relational difference between the two periods, one can expect that $j_{i+1} = j_i + 1$ holds in most cases; however, the fraction a_{c} continues to increase and sometimes reaches one, and the equation does not hold. In such a case, we have to reset as $a_{\mathrm{c}} = 0$ and substitute the index $j_i \leftarrow j_i - 1$ in the $T_{\mathrm{RX}} < T_{\mathrm{TX}}$ case and $j_i \leftarrow j_i + 1$ in the opposite case (see Fig. 5) to maintain bit synchronization. In practice, one cannot expect a_{c} to be exactly one; hence, we introduce threshold a_{th} and execute the above reset procedure when the estimated fraction a_{c} exceeds it.

We summarize the above procedure in algorithm III.2. Note that we do not know whether T_{RX} is larger or smaller than T_{TX} in advance. In addition, when a_{c} is 0, Γ_+ and Γ_- are equal. Therefore, we need to execute the algorithm in both the $T_{\mathrm{RX}} < T_{\mathrm{TX}}$ and in the opposite case for a sufficiently large number of frames N_{det} in which there might be frames with nonzero a_{c}. After that, we choose one in which the sum of the values of the minimized cost functions is smaller.

TABLE I
PARAMTERS USED FOR THE NUMERICAL SIMULATION

parameters	value
l_{short}: the number of frames to estimate a_c and bit sequence	60
l_{margin}: size of overlapping margin for the short-term estimation	10
l_{long}: the number of frames to identify channel state R and d	800
a_{th}: threshold for the reset of bit sequence index	0.9
T_{TX}: period of bit transmission relative to period of exposure T_{RX}	$0.9995T_{\mathrm{RX}}$, $0.9999T_{\mathrm{RX}}$, $1.0001T_{\mathrm{RX}}$, $1.0005T_{\mathrm{RX}}$
N_{iter}: the number of iterations for minimizing the cost function	2
$\alpha_0, \beta_0, m_0, \nu_0$ and W_0: initial parameters in the VB-GMM inference	$10, 1, 0, 10$ and the inverse of the variance of input signal s
SNR: signal to noise ration in dB	16, 17, 18, 19, 20

IV. EXPERIMENTAL SETUPS

We conducted numerical simulations and experiments in a real environment. In both setups, the receiver does not know the gain coefficients, background noise magnitudes, or accurate transmission clock in advance. Our purpose is to verify successful estimation of the unknown parameters by our method and the method's demodulation performance in such a situation.

A. Numerical Simulations

In numerical simulations, we investigated the performance of our demodulation method under different signal-to-noise ratios (SNRs) and parameters. The camera receiver sampling period is T_{RX}, and the exposure time is half of the sampling period. The transmitter sends pseudorandom bit sequences in periods of $0.9995T_{\mathrm{RX}}$, $0.9999T_{\mathrm{RX}}$, $1.0001T_{\mathrm{RX}}$, and $1.0005T_{\mathrm{RX}}$; thus, the relative differences in the clock generators are -500, -100, $+100$, and $+500$ ppm. These relative differences are within the frequency tolerance of conventional crystal oscillator units [30]. The frame size for estimating the linear combination coefficients and the bit sequence, i.e., the size of the set Ω_{short} is 60 which includes 20 additional overlapping samples. The frame size for the channel estimation, i.e, the size of the set Ω_{long} is 800. We set the threshold a_{th} to 0.9, the number of frames to estimate whether $T_{\mathrm{TX}} > T_{\mathrm{RX}}$ or not $N_{\mathrm{det}} = 1000$, and the number of iterations to minimize the cost functions (10) or (12) $N_{\mathrm{iter}} = 2$. The R and d relatively vary with the ratio generated by zero-mean and $\sigma_{\mathrm{c}} = 10^{-6}$ variance Gaussian distributions. The noise is generated by a Gaussian distribution with zero-mean and σ_{n}^2 variance, and the SNRs of the received signals range from 16 to 20 dB, where the SNR is defined as,

$$\mathrm{SNR} = \frac{R^2}{2\sigma_{\mathrm{n}}^2}. \tag{14}$$

The number of the bit sequences is 1 Mbits per simulation, and we performed 1000 simulations for each condition. Note that only T_{RX} and $\tau = 0.5T_{\mathrm{RX}}$ are known in advance. In addition, to confirm the potential demodulation performance for each condition, we also employed an oracle estimator. The oracle estimator knows the exact values of the gain factor R, the background noise level d, and the percentage coefficient a_{c} and estimates the bit sequences by the MLSD method with the parameters. In our problem, one can get the global minimum solution by the MLSD method because there does not exist any correlation, except that the two consecutive grids on the trellis diagram should be solved. Therefore, the estimated bit sequences are expected to be the most accurate compared to those obtained by any other demodulation methods. The parameters used in the numerical simulations are summarized in Table I.

To confirm the robustness of our method against the parameter settings for demodulation, we also conducted numerical simulations with various processing lengths. We employed the short-term processing length $l_{\mathrm{short}} = 30, 40, 60, 100, 180, 340, 660,$ and 1300 which include 20 additional overlapping samples to estimate the bit sequences and a_{c}. Because the relative clock difference should affect the demodulation performance with differences of short-term processing lengths, we conducted the simulations for $+100$ and $+500$ ppm cases. We also performed numerical simulations for long-term processing lengths $l_{\mathrm{long}} = 100, 200, 400, 800, 1600, 3200, 6400,$ and 12800. In the simulations for long-term processing lengths, we employed the ratio of variation of R and d as $\sigma_{\mathrm{c}} = 10^{-4}, 10^{-6}, 10^{-8}$ to confirm the relationship between the channel state variation and the long-term processing length. We set $\mathrm{SNR} = 16$, the number of trials to be 100 for each condition and other parameters except for those mentioned here to be the same as in Table I.

Since our method uses iterative convergence technique, the computation time required for achieving sufficient accuracy is unknown and may differ with different SNR. We therefore conducted test to show that our method can process data in a practical period od time with satisfactory accuracy. The iteration for the VB-GMM was limited to 100 in one condition and 10,000 in the other. The other parameter settings were the same as in Table I. We performed the experiment 100 times for each condition.

B. Real Environment Experiment

We experimented to validate our demodulation method in practical situations. As Fig. 6, we conducted tests with a high-speed camera as a receiver and a single LED as a transmitter at various distance conditions. The LED transmitter transmitted pseudo-random bit sequences with OOK modulation at about 10 kilobits per second (kbps). The high-speed camera receiver drove at 10000 frames per second with exposure time 0.5×10^{-4} second and sent each captured image to a field-programmable gate array (FPGA) for further processing. The FPGA extracted values of pixels on which the image of the LED transmitter falls as in Fig. 1 and saved the sum of the values in its local storage. After that, we applied our demodulation method to obtain the transmitted bit sequences and evaluated their bit error rate (BER). For comparative purposes, we employed a conventional demodulation method [23] that cannot estimate the background

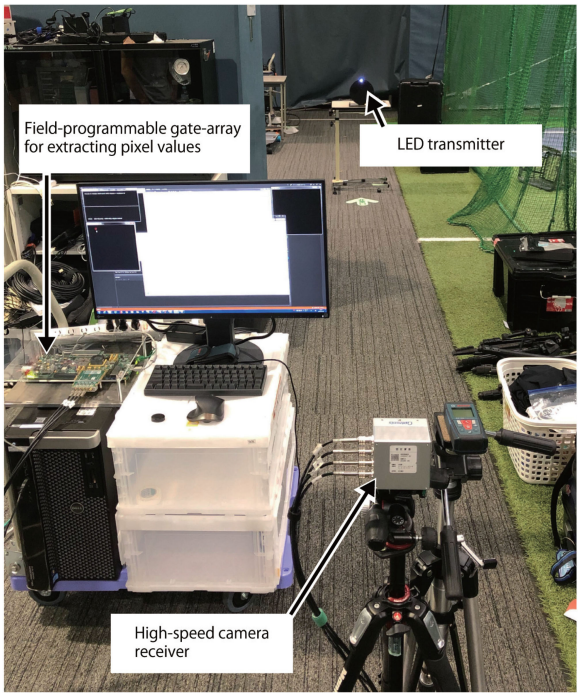


Fig. 6. Photograph of the experimental setup. The signal from the LED transmitter is observed with a high-speed camera receiver. The observed signals are pre-processed by an FPGA and sent to the PC for storage.

TABLE II
PARAMTERS AND EQUIPMENT USED FOR THE REAL ENVIRONMENT EXPERIMENT

parameters and equipment	value
T_{RX}: period of exposure	1/10000 second
T_{TX}: period of bit transmission	about 1/10000 second
High-speed camera model	Optronis CP80-3-M-540
LED model	Linkman BL304W1CA1B02
distances in the experiment (m)	$5, 7.5, 10, 12.5, 15, 17.5, 20$

noise level. The distance between the transmitter and receiver in the experiment was set from 5 m to 20 m in 2.5 m increments. In order to show the effectiveness of our demodulation method in a more practical situation, we also conducted experiments in which the lens f-value was varied from 0.8 to 4 to change the effective luminance of the transmitter. The parameter settings for the demodulation and equipment used in the experiment are shown in Table II. The processing lengths and other parameter settings for demodulation were the same as in the numerical simulations of performance versus SNR.

V. RESULTS AND DISCUSSIONS

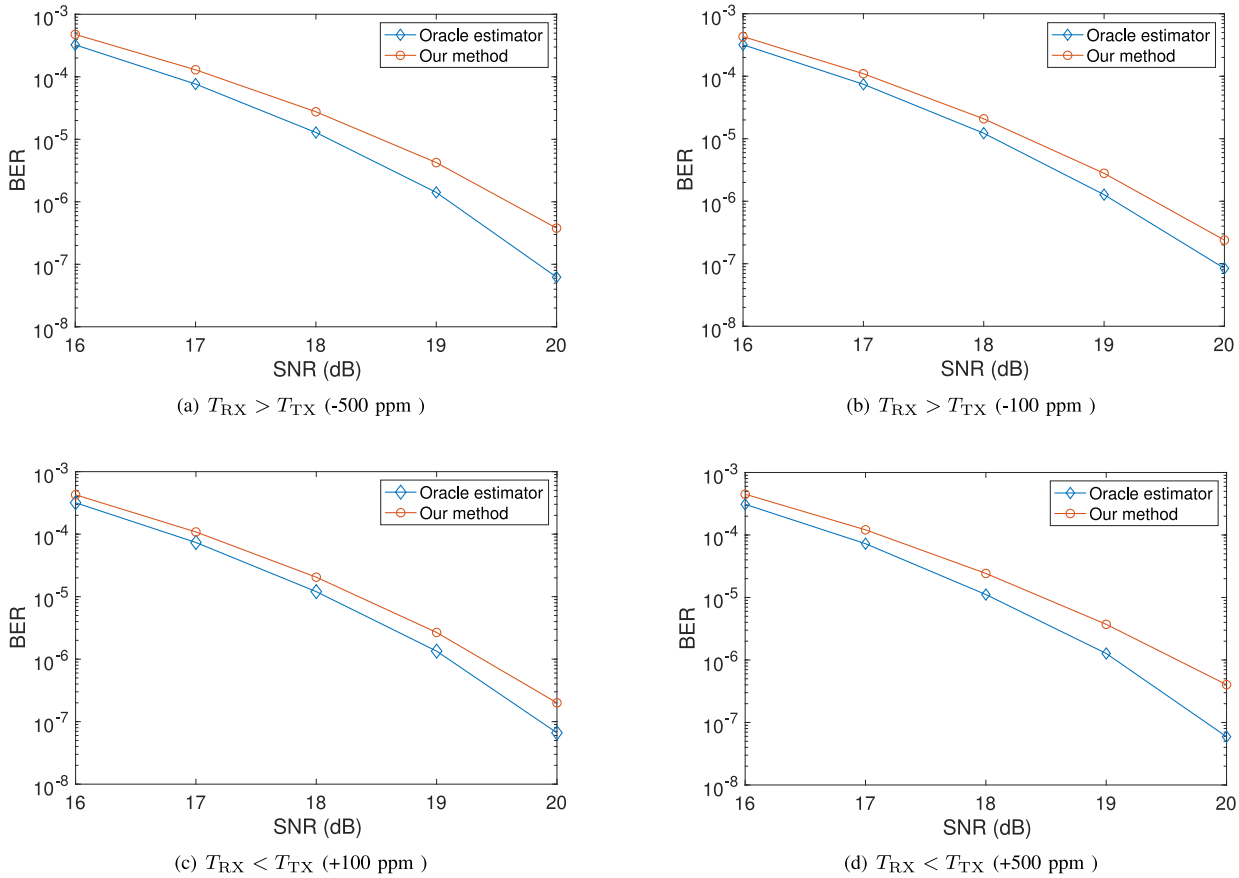
A. Numerical Experiment

Fig. 7 shows BERs versus SNR of our demodulation method and the oracle estimator. As the SNR increases, the bit sequences demodulated by both methods become more accurate. BERs resulted for the oracle estimator are lower than those for our method as expected, but they are almost on the same order. Fig. 8 shows BERs versus relational clock differences for our demodulation method and the oracle estimator for SNR $= 16$. A large relational clock difference decreases the estimation accuracy of our method. This is because it makes the parameter estimation of a_{c} difficult. On the other hand, BERs of the oracle estimator are not sensitive to changes in clock difference. This is simply that it knows the exact values of the parameters. Note that, whether the sampling period T_{RX} is longer than the transmission period T_{TX} or not, the BERs stay in the low range, although signals in the $T_{\mathrm{RX}} > T_{\mathrm{TX}}$ cases resemble undersampling situation and should be more difficult to demodulate than that in the opposite case.

Fig. 9(a) shows BERs of our demodulation method for various short-term processing lengths. The BERs of demodulated bit sequences do not dramatically vary with short-term processing length for $l_{\mathrm{short}} = 20, 40$, and 80 in both $+100$ and $+500$ ppm cases. When the short-term processing length is too short, estimating the synchronization parameter is supposed to be difficult. In estimating a_{c} based on (11) or (13), the bit sequence should contain 0 and 1 in a well-balanced ratio because a biased bit sequence degrades the estimation accuracy. Unfortunately, in the $l_{\mathrm{short}} = 10$ case, the bit sequence for the estimation is likely to be such a biased one, and the estimated parameter has large error. This error fails to reset the synchronization parameter to zero and thus the bit index synchronization corrupts. In $+500$ case, the BERs become nearly 0.5 when the short-term processing length exceeds 160. In these situations, the large change in the true synchronization parameters in the long-term processing lengths decreases their estimation accuracy. This also fails the reset of the synchronization parameter.

The BERs of demodulated bit sequences are stable for long-term processing lengths except for extreme cases as Fig. 9(b) shows. When l_{long} is too small, the statistical approach algorithm III.1 could fail because of a lack of samples for estimating the parameters. The estimation accuracy could also decrease when l_{long} is too long compared to the channel state variation cases. This is because the variations of $R[i]$ and $d[i]$ in such a large number of frames are large and thus the estimations $\tilde{R}$ and $\tilde{d}$ are not accurate enough to estimate the synchronization parameter and demodulate the bit sequences. When the channel state variation is large, as in the $\sigma_{\mathrm{s}} = 10^{-4}$ case, our demodulation method does not work well for the above reasons.

Fig. 10 shows the experimental computation time of our demodulation method. We applied the VB-GMM with the number of iterations limited to 100 and with that limited to 10,000 for the same input signal. The left and right axes represent the BER and computation time, respectively. The orange line indicates the upper limit of the computation time allowed for a 10 kbps signal. The limit of 10,000 iterations is meant to prevent the program from ever ending, so the VB-GMM algorithm is expected to converge in this condition. When we execute the VB-GMM until the algorithm converges, it is not able to process in real time in some situations. On the other hand, this problem can be avoided by limiting the number of iterations to 100. These two BER results are almost identical. Therefore, we can say that our method is practically feasible for transmissions of about 10 kbps.



(a) $T_{\mathrm{RX}} > T_{\mathrm{TX}}$ (-500 ppm)

(b) $T_{\mathrm{RX}} > T_{\mathrm{TX}}$ (-100 ppm)

(c) $T_{\mathrm{RX}} < T_{\mathrm{TX}}$ (+100 ppm)

(d) $T_{\mathrm{RX}} < T_{\mathrm{TX}}$ (+500 ppm)

Fig. 7. BER v.s. SNR in dB of our method and the oracle estimator are shown for each relative clock difference. The relative clock differences are -500, -100, $+100$ and $+500$ ppm.

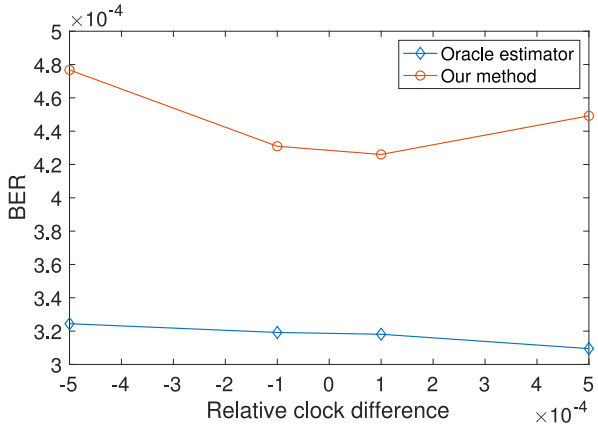


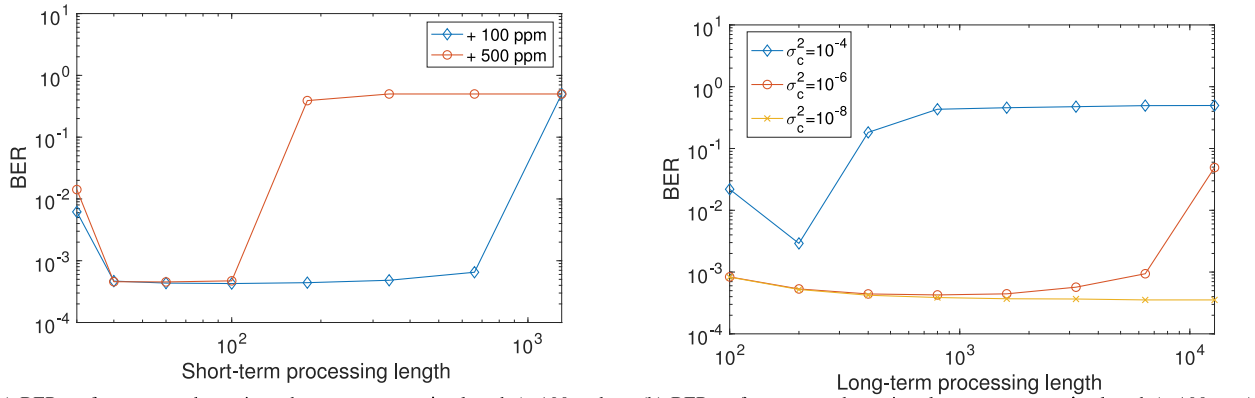
Fig. 8. BER performance of our method and the oracle estimator under different relational clock differences. We set SNR to be 16.

To show that VB-GMM accurately estimates the distribution of received signals, we plot the probability density functions of estimated results and the received signals in the Fig. 11 as examples. As described in Section III-B, the purpose of VB-GMM estimation is to identify the gain coefficient and the background noise. Thus, it is not necessary to accurately estimate the distribution of received signals classified as s_{01} and s_{10}. The figure shows that VB-GMM correctly clusters s_{01} and s_{10} and effectively identifies the gain coefficient and the background noise from the estimated distributions of s_{11} and s_{00}.

B. Real Environment Experiment

Fig. 12 shows the results of the real environment experiment. Note that the plot in Fig. 12 is in a logarithmic scale, and the results plotted on 10^{-7} mean BERs $= 0$. Our method demodulated the bit sequence without any error within a 15 m distance and with low BERs in the 17.5 m and 20 m conditions. The conventional method [23] also demodulated the bit sequences accurately in up to 10 m conditions, while its performance degraded significantly at distances of more than 15 m. In such long-distance conditions, the gain coefficient becomes substantially low, so the effect of the background noise is not negligible for the demodulation. Fig. 13 shows the computational times of our method for the real-setup experiment. Within 20 m, our method can demodulate the received signals in a real-time manner.

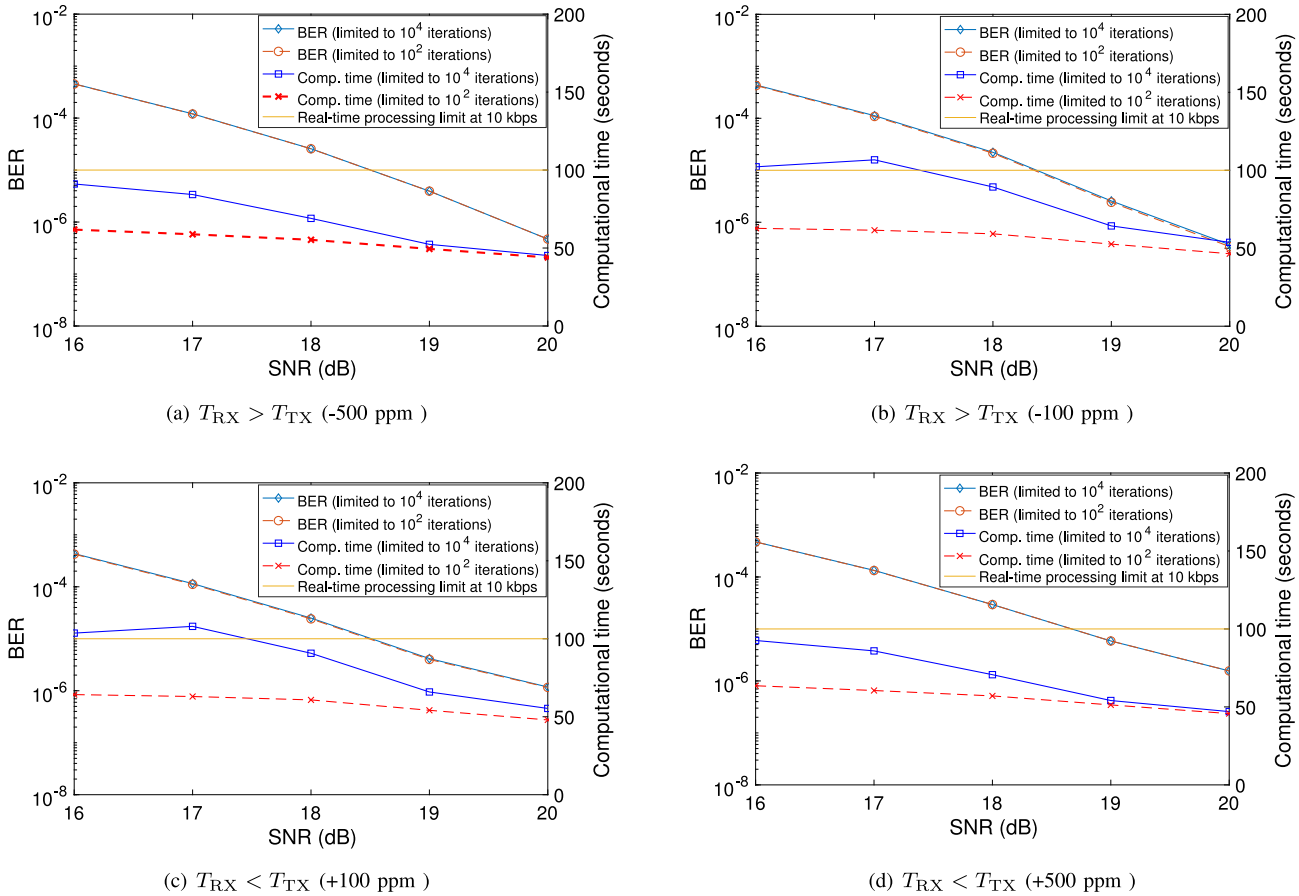
To check if our demodulation method is working effectively, we show a part of the received signal and the results of demodulation for the 5 and 20 m cases in Fig. 14. Because of the



(a) BER performance under various short-term processing length (+100 and +500 ppm)

(b) BER performance under various long-term processing length (+100 ppm)

Fig. 9. BERs performance of our method are shown for various processing lengths.



(a) $T_{\mathrm{RX}} > T_{\mathrm{TX}}$ (-500 ppm)

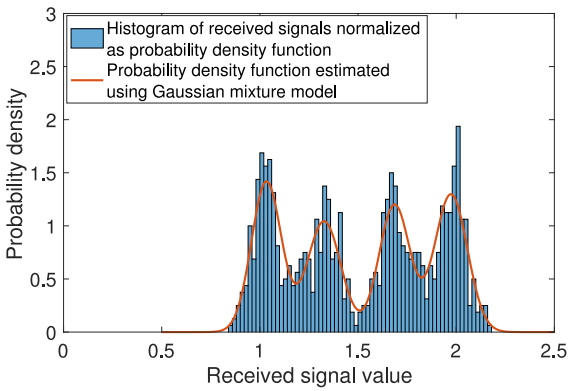
(b) $T_{\mathrm{RX}} > T_{\mathrm{TX}}$ (-100 ppm)

(c) $T_{\mathrm{RX}} < T_{\mathrm{TX}}$ (+100 ppm)

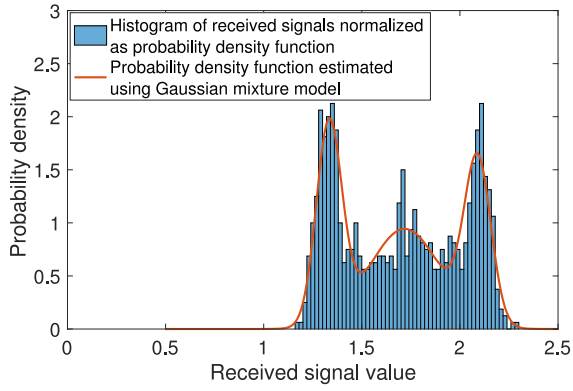
(d) $T_{\mathrm{RX}} < T_{\mathrm{TX}}$ (+500 ppm)

Fig. 10. Computational time of our demodulation method in various situations. We also plot the BERs of them. The relative clock differences are -500, -100, $+100$ and $+500$ ppm.

relatively low background noise level of the received signal at 5 m, both the present method and our previous one in [23] is able to estimate the coefficients. However, due to the relatively large background noise at 20 m, the conventional method which lacks an accurate estimation methods to accurately estimate background noise levels, cannot demodulate the received signal. The present demodulation method was able to perform demodulation stably in all cases. Note that the normalized received signals



(a) $T_{\mathrm{RX}} < T_{\mathrm{TX}}$ (+100 ppm)



(b) $T_{\mathrm{RX}} < T_{\mathrm{TX}}$ (+500 ppm)

Fig. 11. Histograms of the received signals normalized as the probability density functions and the estimated probability density function by VB-GMM are shown. Relative clock differences are $+100$ and $+500$ ppm.

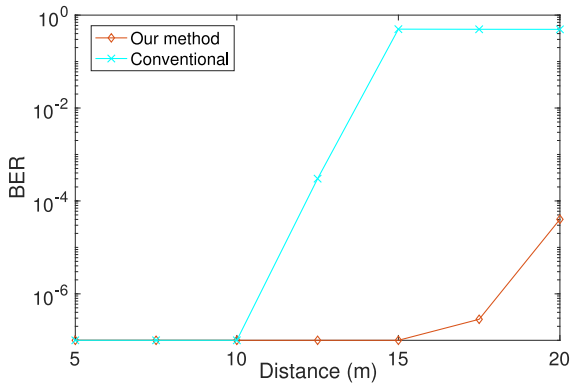


Fig. 12. BER performance of our method under different communication distances in the real environmental experiment.

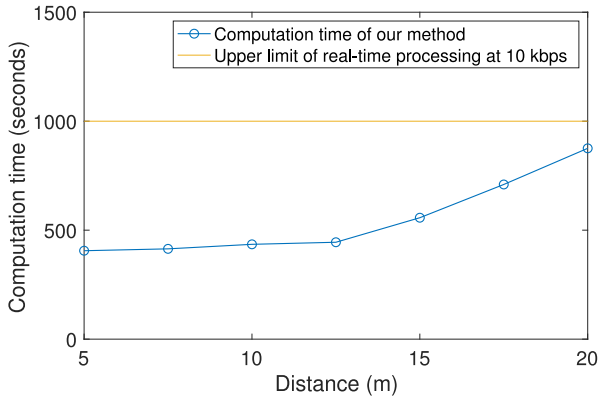


Fig. 13. Computational times of our method at different communication distances in the real-environment experiment.

$s_{\mathrm{normalized}}[i]$s are obtained by

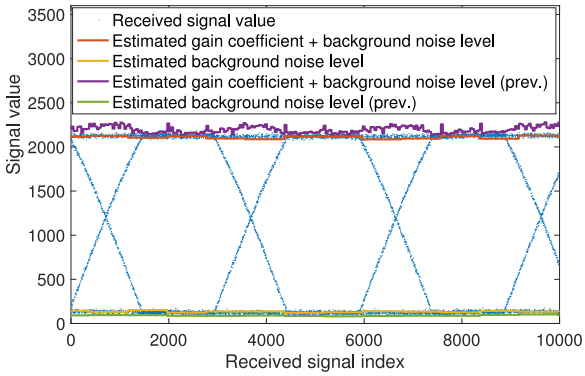
$$s_{\mathrm{normalized}}[i] = \frac{s[i] - \tilde{d}}{\tilde{R}} \tag{15}$$

and the estimated background noise levels in the conventional method are obtained by taking the minimum value of the received signals in the current long-processing frame. Fig. 15 shows the probability density functions of the received signals and those estimated by the VB-GMM method. The algorithm estimated the distributions of the received signal even when the distance was long and thus the relative noise level was high.

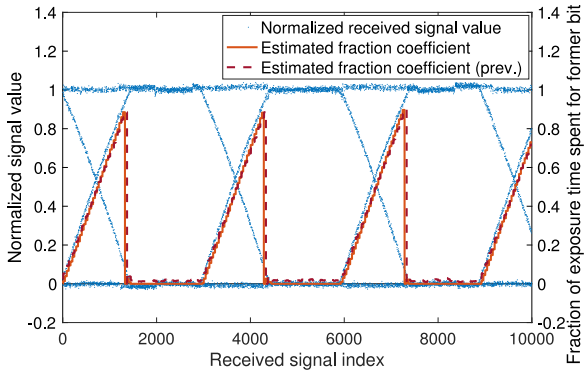
Fig. 16 shows the received signals and coefficients estimated by our method when the f-value of lens is varied. It can be seen that our demodulation method stably performs the channel estimation and synchronization even though the demodulation becomes more difficult due to the luminance change by continuously changing the f-value. This implies that it is possible to demodulate signals received from a transmitter that varies in distance from the receiver.

C. Short Discussion

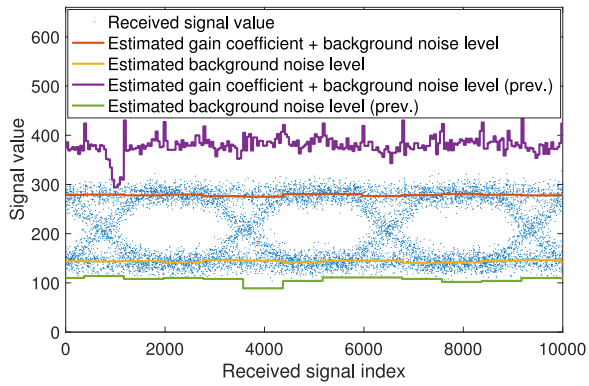
The above results validate that our method successfully demodulates received signals under practical situations. Thus, even if only the sampling period and the exposure time are known in advance, our method can estimate the gain coefficient, background noise level, and the synchronization parameter with enough accuracy to demodulate the bit sequences. Moreover, our demodulation method is sufficiently robust against parameter settings. For a wide range of short-term processing lengths, it can estimate synchronization parameters and bit sequences with almost the same accuracy when the relative clock difference between a receiver and a transmitter is within ordinary frequency tolerance of conventional crystal oscillator units. Except for the extreme cases, the channel-state estimation by VB-GMM works accurately for some ranges of long-term processing lengths. In the real-environmental experience, our demodulation method showed its superiority to the conventional method especially when the communication distance is long. In such situations, the estimation of background-noise becomes important for demodulation because the gain coefficient is effectively low compared to the noise.



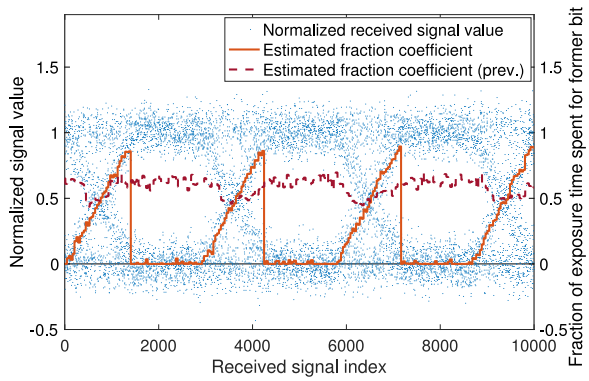
(a) Received signal and estimated coefficients at 5 m



(b) Normalized received signal and estimated fraction coefficient a_c at 5 m

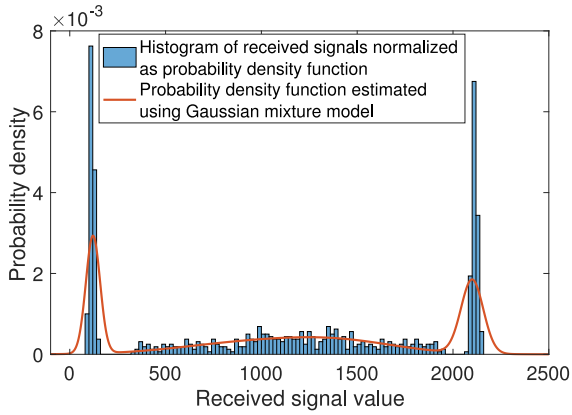


(c) Received signal and estimated coefficients at 20m

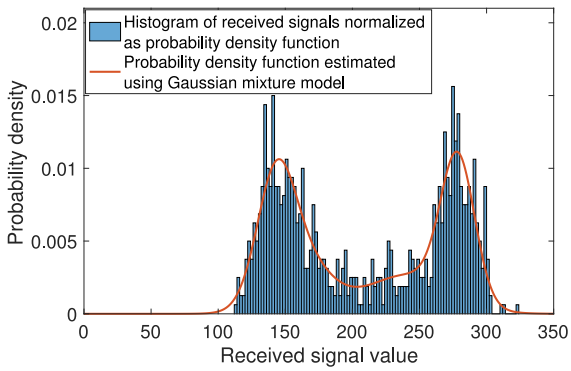


(d) Normalized received signal and estimated fraction coefficient a_c at 20 m

Fig. 14. Received signals and estimated gain coefficients and background noise level. We also plot normalized received signals and the estimated fraction coefficients a_c. We applied two methods: the one in our previous work in [23] and the method in the present work.

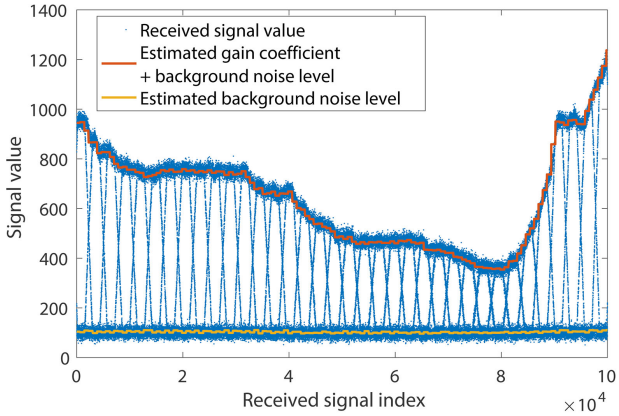


(a) Histogram and the estimated probability density function at 5 m

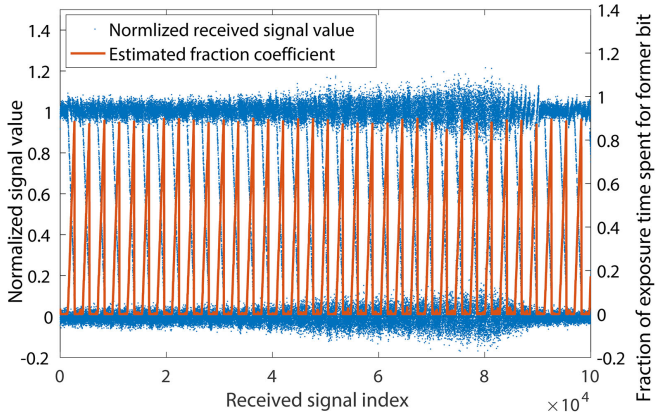


(b) Histogram and the estimated probability density function at 20 m

Fig. 15. Histograms of the received signals normalized as the probability density functions and the estimated probability density function by VB-GMM are shown. The distances are 5 m and 20 m in the real-environment experiment.



(a) Received signal and estimated coefficients



(b) Normalized received signal and estimated fraction coefficient a_c

Fig. 16. Received signals and estimated gain coefficients and background noise levels. We also plot normalized received signals and the estimated fraction coefficients a_c.

VI. Conclusion

OCC systems with distributed transmitter nodes are expected to be used in sensor networks for acoustic measurement, biosignal monitoring and other applications. The sensors in these networks should be fabricated as simply as possible to suppress costs and power consumption. For this purpose, in this paper, we considered an OCC system without any synchronization devices in the transmitters. In such a case, the periods of the sampling at the receiver and the bit transmission generally may differ from each other and the received light signal sometimes becomes a linear combination of two adjacent transmitted bits. To demodulate the bit sequence from the light signals, we introduced a received signal model and a cost function that should be minimized. To estimate unknown channel states, we employ the VB-GMM method to infer the generative probability of the received signals. The demodulation procedure also uses a maximum-likelihood sequence detection method, which can be implemented by the Viterbi algorithm, in combination with suboptimal per-survivor processing to estimate the bit sequence and the synchronization parameter by minimizing the cost function. Our new method can demodulate the signal even when the receiver is not given the precise transmission period, which the conventional method must know in advance. To confirm the efficiency of our method, we conducted numerical simulations and compared the results with those obtained by an oracle estimator that knows parameters other than the bit sequence in advance. The accuracies of bit sequences demodulated by our method are almost on the same order as those estimated by the oracle estimator. We also tested with the demodulation method in a practical environment and showed its superiority over the conventional method, which cannot estimate the channel state precisely when background noise is relatively large.

The results of this study suggest directions for further researches are needed in the future. While we assumed SISO in this study, we expect that our demodulation method could find a wider range of applications when we remodel the channel mode by MIMO, for which one should consider spatial correlations. In addition, although our demodulation method can be performed in a real-time manner at about 10 kbps, we shall aim to develop simpler methods for faster communication.

Acknowledgment

We thank the two anonymous reviewers whose comments and suggestions helped improve and clarify this manuscript. We also would like to thank our colleague Mr. Masahiro Nakano who gave us helpful advice.

References

[1] T. Komine and M. Nakagawa, "Integrated system of white LED visible-light communication and power-line communication," *IEEE Trans. Consum. Electron.*, vol. 49, no. 1, pp. 71–79, Feb. 2003.
[2] Y. Tanaka, T. Komine, S. Haruyama, and M. Nakagawa, "Indoor visible light data transmission system utilizing white LED lights," *IEICE Trans. Commun.*, vol. 86, no. 8, pp. 2440–2454, 2003.
[3] T. Komine and M. Nakagawa, "Fundamental analysis for visible-light communication system using LED lights," *IEEE Trans. Consum. Electron.*, vol. 50, no. 1, pp. 100–107, Feb. 2004.
[4] H. Haas, L. Yin, Y. Wang, and C. Chen, "What is LiFi?" *J. Lightw. Technol.*, vol. 34, no. 6, pp. 1533–1544, Mar. 2016.
[5] T. Yamazato *et al.*, "Image-sensor-based visible light communication for automotive applications," *IEEE Commun. Mag.*, vol. 52, no. 7, pp. 88–97, Jul. 2014.
[6] T. Yamazato *et al.*, "Vehicle motion and pixel illumination modeling for image sensor based visible light communication," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 9, pp. 1793–1805, Sep. 2015.
[7] Y. Goto *et al.*, "A new automotive VLC system using optical communication image sensor," *IEEE Photon. J.*, vol. 8, no. 3, pp. 1–17, Jun. 2016.
[8] T. Nguyen, A. Islam, and Y. M. Jang, "Region-of-interest signaling vehicular system using optical camera communications," *IEEE Photon. J.*, vol. 9, no. 1, pp. 1–20, Feb. 2017.
[9] N. B. Hassan *et al.*, "Interference cancellation in MIMO NLOS optical-camera-communication-based intelligent transport systems," *Appl. Opt.*, vol. 58, no. 34, pp. 9384–9391, Dec. 2019. [Online]. Available: http://ao.osa.org/abstract.cfm?URI=ao-58-34-9384
[10] B. W. Kim, H. Kim, and S. Jung, "Display field communication: Fundamental design and performance analysis," *J. Lightw. Technol.*, vol. 33, no. 24, pp. 5269–5277, Dec. 2015.
[11] R. Boubezari, H. Le Minh, Z. Ghassemlooy, and A. Bouridane, "Smartphone camera based visible light communication," *J. Lightw. Technol.*, vol. 34, no. 17, pp. 4121–4127, Sep. 2016.

[12] S. Pergoloni, M. Biagi, S. Colonnese, R. Cusani, and G. Scarano, "A space-time RLS algorithm for adaptive equalization: The camera communication case," *J. Lightw. Technol.*, vol. 35, no. 10, pp. 1811–1820, May 2017.
[13] Z. Jiao, B. Zhang, M. Liu, and C. Li, "Visible light communication based indoor positioning techniques," *IEEE Netw.*, vol. 31, no. 5, pp. 115–121, 2017.
[14] Y. Li, Z. Ghassemlooy, X. Tang, B. Lin, and Y. Zhang, "A VLC smartphone camera based indoor positioning system," *IEEE Photon. Technol. Lett.*, vol. 30, no. 13, pp. 1171–1174, Jul. 2018.
[15] G. P. Nava *et al.*, "A high-speed camera-based approach to massive sound sensing with optical wireless acoustic sensors," *IEEE Trans. Comput. Imag.*, vol. 1, no. 2, pp. 126–139, Jun. 2015.
[16] G. P. Nava *et al.*, "GPU-based real-time beamforming for large arrays of optical wireless acoustic sensors," *Acoust. Sci. Technol.*, vol. 36, no. 6, pp. 489–499, 2015.
[17] T. G. Sato, Y. Shiraki, and T. Moriya, "Heart rate measurement based on event timing coding observed by video camera," *IEICE Trans. Commun.*, vol. 100, no. 6, pp. 926–931, 2017.
[18] P. Luo, M. Zhang, Z. Ghassemlooy, S. Zvanovec, S. Feng, and P. Zhang, "Undersampled-based modulation schemes for optical camera communications," *IEEE Commun. Mag.*, vol. 56, no. 2, pp. 204–212, Feb. 2018.
[19] R. D. Roberts, "Undersampled frequency shift on-off keying (UFSOOK) for camera communications (CamCom)," in *Proc. 22nd Wireless Opt. Commun. Conf.*, May 2013, pp. 645–648.
[20] P. Luo *et al.*, "Experimental demonstration of RGB LED-based optical camera communications," *IEEE Photon. J.*, vol. 7, no. 5, pp. 1–12, Oct. 2015.
[21] K. Ohshima, T. Naramoto, K. Yamaguchi, and W. Chujo, "Rolling-shutter-based asynchronous optical camera communication by a cycle pattern of received symbols using smartphones," *IEICE Commun. Exp.*, vol. 8, no. 3, pp. 49–54, 2019.
[22] W. Mao and J. M. Kahn, "Free-space heterochronous imaging reception of multiple optical signals," *IEEE Trans. Commun.*, vol. 52, no. 2, pp. 269–279, Feb. 2004.
[23] Y. Shiraki, T. G. Sato, Y. Kamamoto, and T. Moriya, "Flexible synchronization in optical camera communication with on-off keying," in *Proc. IEEE Globecom Workshops*, Dec. 2017, pp. 1–6.
[24] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2006.
[25] Z. Ghassemlooy, P. Luo, and S. Zvanovec, "Optical camera communications," in *Proc. Opt. Wireless Commun.* Springer, 2016, pp. 547–568.
[26] Xiaoming Zhu and J. M. Kahn, "Markov chain model in maximum-likelihood sequence detection for free-space optical communication through atmospheric turbulence channels," *IEEE Trans. Commun.*, vol. 51, no. 3, pp. 509–516, Mar. 2003.
[27] T. Foggi, E. Forestieri, G. Colavolpe, and G. Prati, "Maximum-likelihood sequence detection with closed-form metrics in OOK optical systems impaired by GVD and PMD," *J. Lightw. Technol.*, vol. 24, no. 8, pp. 3073–3087, Aug. 2006.
[28] R. Raheli, A. Polydoros, and Ching-Kae Tzou, "Per-survivor processing: A general approach to MLSE in uncertain environments," *IEEE Trans. Commun.*, vol. 43, nos. 2–4, pp. 354–364, Feb. 1995.
[29] T. Kimura, T. Tokuda, Y. Nakada, T. Nokajima, T. Matsumoto, and A. Doucet, "Expectation-maximization algorithms for inference in dirichlet processes mixture," *Pattern Anal. Appl.*, vol. 16, pp. 55–67, 2013. [Online]. Available: https://link.springer.com/article/10.1007/s10044-011-0256-4
[30] J. Murdock and D. Griffith, "Crystal oscillator and crystal selection for the CC26xx and CC13xx family of wireless mcus," Texas Instruments Incorporated, Tech. Rep. SWRA495E, Nov. 2016.

Yoshifumi Shiraki (Member, IEEE) received the B.A. degree in natural science from the International Christian University, Tokyo, Japan, in 2008, and the M.S. and Ph.D. degrees in computational intelligence and systems science from the Tokyo Institution of Technology, Tokyo, in 2010 and 2015, respectively. Since 2010, he has been studying signal processing, particularly distributed compressed sensing with the NTT Communication Science Laboratories, Atsugi, Japan. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE).

Takashi G. Sato (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in information science and technology from the University of Tokyo in 2003, 2005, and 2008, respectively. He is currently a researcher with NTT Communication Science Laboratories. His research interests include human interface, especially in brain-machine, audio and tactile interfaces, using psychophysiological measurement as feedback information. He is a member of IEICE, the Japan Neuroscience Society and the Society of Instrument and Control Engineers.

Yutaka Kamamoto (Senior Member, IEEE) received the B.S. degree from Keio University in 2003, and the M.S. and Ph.D. degrees from the Graduate School of Information Science and Technology, University of Tokyo, in 2005 and 2012, respectively. Since joining NTT Communication Science Laboratories, in 2005, he has been studying signal processing and information theory, particularly lossless coding of time-domain signals. He additionally joined NTT Network Innovation Laboratories, where he developed the audio-visual codec for ODS (Other Digital Stuff/Online Digital Source), from 2009 to 2011. He has contributed to the standardization of coding schemes for MPEG-4 Audio lossless coding (ALS), ITU-T Recommendation G.711.0 Lossless compression of G.711 pulse code modulation, and 3GPP Enhanced Voice Services (EVS). Dr. Kamamoto was the recipient of the Telecom System Student Award from TAF in 2006, the IPSJ Best Paper Award from the Information Processing Society of Japan (IPSJ) in 2006, the Telecom System Encouragement Award from TAF in 2007, the Awaya Prize Young Researcher's Award from Acoustic Society of Japan (ASJ) in 2011, and the Technical Development Award from ASJ in 2016. He is a member of IPSJ, ASJ, and IEICE.

Tomonori Izumi (Member, IEEE) received the B.E. degree in computer engineering, and the M.E. and Ph.D. degrees in electrical and electronic engineering all from the Tokyo Institute of Technology, Japan, in 1992, 1994, and 1998, respectively. From 1998 to 2005, he was a Research Associate with Kyoto University, Japan. He joined Ritsumeikan University, Japan in 2005 and has been a Professor with the Department of Electronic and Computer Engineering, since 2016. He was also a Senior Research Scientist with Synthesis Corporation from 1998 to 2017. His research interests include systems, applications, and design methodologies of reconfigurable hardware. He is a member of IEICE, IPSJ .

Yu Nakahara received the B.E. and M.E. degrees in electronic and computer engineering from Ritsumeikan University, Japan, in 2015 and 2017, respectively. His research interests include communication, image processing, and reconfigurable hardware. He is a member of IEICE.

Kyosuke Kondo received the B.E. degree in electronic and computer engineering from Ritsumeikan University, Japan, in 2017. His research interests include reconfigurable hardware and embedded systems.

Takehiro Moriya (Fellow, IEEE) received the B.S., M.S., and Ph.D. degrees in mathematical engineering and instrumentation physics from the University of Tokyo, in 1978, 1980, and 1989, respectively. Since joining NTT laboratories in 1980, he has been engaged in research on medium to low bitrate speech and audio coding. In 1989, he worked with AT&T Bell Laboratories as a Visiting Researcher. Since 1990, he has contributed to the standardization of coding schemes for the Japanese Public Digital Cellular system, ITU-T, ISO/IEC MPEG, and 3GPP. He is currently the Head of the Moriya Research Laboratory with NTT Communication Science Laboratories. Dr. Moriya is a recipient of many awards, including IEEE James L. Flanagan Speech and Audio Processing Award in 2016. He is currently an NTT Fellow. He is a member of IPSJ, Acoustic Engineering Society, and ASJ. He is also a member of IEEE Technical Committee on Speech and Language Processing.