Data signal demodulation based on machine learning for digital signage and image sensor based visible light communication

Yuki Iyoda\textsuperscript{1a)}, Kentaro Kobayashi\textsuperscript{1b)}, and Wataru Chujo\textsuperscript{1}

\textsuperscript{1}Graduate School of Science and Engineering, Meijo University
1-501 Shiogamaguchi, Tempaku-ku, Nagoya, Aichi 468-8502, Japan
\textsuperscript{a)} 213427004@ccmail.meijo-u.ac.jp \quad \textsuperscript{b)} kkobayashi@meijo-u.ac.jp

Abstract: In this paper, we discuss a visible light communication (VLC) system using a digital signage and an image sensor. We focus on the demodulation part and propose a method that demodulates transmitted data signals from received images of the image sensor using machine learning. The performance of the proposed machine learning-based demodulation method is evaluated for various datasets that simulate noise, blurring, and misalignment that may occur in received images.

Keywords: Visible light communication, Digital signage, Image sensor, Machine learning, Demodulation

Classification: Wireless communication technologies

References

[1] H. Aoyama and M. Oshima, “Visible light communication using a conventional image sensor,” \textit{Proc. 12th Annu. IEEE Consumer Commun. Netw. Conf.}, pp. 103-108, Jan. 2015. DOI: 10.1109/CCNC.2015.7157954

[2] Y. Lin, T. Wada, K. Mukumoto, and H. Okada, “Performance evaluation of information embedding schemes based on wavelet transform for parallel transmission visible light communication systems,” \textit{Proc. 2017 IEEE 6th Global Conference on Consumer Electronics (GCCE)}, pp. 1-2, Oct. 2017. DOI: 10.1109/GCCE.2017.8229348.

[3] R. Mushu, T. Wada, K. Mukumoto, and H. Okada, “A Proposal of Information Embedding Scheme Based on Discrete Cosine Transform in Parallel Transmission Visible Light Communications,” \textit{Proc. 2018 IEEE 7th Global Conference on Consumer Electronics (GCCE)}, pp. 175-176, Oct. 2018. DOI: 10.1109/GCCE.2018.8574698.

[4] H. Okada, S. Sato, T. Wada, K. Kobayashi, and M. Katayama, “Preventing Degradation of the Quality of Visual Information in Digital Signage and Image-Sensor-Based Visible Light Communication Systems,” \textit{IEEE Photonics Journal}, vol. 10, no. 3, July 2018. DOI:10.1109/JPHOT.2018.2829146

[5] T. Sasaki, K. Kobayashi, H. Okada, and M. Katayama, “Data signal modulation
scheme based on perceptually uniform color space for image sensor based visible light communication,” Proc. IEEE Vehicular Technology Society Asia Pacific Wireless Communications Symposium (APWCS), Aug. 2019. DOI:10.1109/VTS-APWCS.2019.8851671

[6] I. Hamamoto and M. Kawamura, “Image Watermarking Technique Using Embedder and Extractor Neural Networks,” IEICE Trans. Inf. & Syst., vol. E102.D, no. 1, pp. 19-30, Jan. 2019. DOI:10.1587/transinf.2018MUP0006

[7] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition”, Proc. International Conference on Learning Representations (ICLR), Apr. 2015.

1 Introduction

Digital signage is a medium that displays advertisement images and various useful information using electronic display devices and has already been equipped in many places. Several VLC schemes have been studied in which wireless communication is realized by superimposing data signals on displayed images of the digital signage and capturing them with an image sensor such as a smartphone camera. Related research, including a method of modulating the backlight of digital signage at high speed [1], modulation methods using frequency domain transformation [2, 3], and modulation methods using color’s chroma components that are difficult for human eyes to perceive [4, 5], have been studied. These studies mainly focus on the processing of the transmitter side. In the processing of the receiver side, all of them detect and extract the data signal part from the captured image and demodulate the data signal by a threshold decision.

We focus on applying machine learning to such digital signage and image sensor based VLC schemes. In digital watermarking, which is one of the techniques to superimpose data signals, the literature [6] utilized compression and decompression of an autoencoder and improved the accuracy of superimposition. In this paper, we apply machine learning to construct a new signal demodulation method for the VLC systems like [4, 5] to achieve more accurate signal demodulation. As a first step, the proposed method demodulates data signals from received images using machine learning in which the communication channel simulates noise, blurring, and misalignment that may occur in received images. We compare the proposed method with the conventional method and clarify its effectiveness.

2 System model

Figure 1 shows a simplified system model that focuses only on data signal demodulation of the VLC systems, in which data signal is superimposed as an image with square cells corresponding to binary data bits. The proposed method demodulates the transmitted data signal from the received image via the communication channel. In the communication channel, we assume that there are misalignment during signal detection, blur caused by light diffusion and lens, and noise caused by illumination and background images, and these are modeled as translation or projection, smoothing, and Gaussian noise, respectively.
2.1 Modulation
The data signal $d$ to be transmitted is represented as $M \times N$ binary symbol matrix and defined as follows.

$$d = \begin{bmatrix} d_{1,1} & \ldots & d_{1,N} \\
\vdots & \ddots & \vdots \\
\vdots & \vdots & \vdots \\
d_{M,1} & \ldots & d_{M,N} \end{bmatrix}$$

$d_{m,n} = \{0, 1\} (m = 1, 2, \ldots M, n = 1, 2, \ldots N)$

The transmitted image $D$ corresponding to $d$ is generated as an image that has width $I$ [pixel] and height $J$ [pixel] and consists of $M \times N$ square cells. Each cell corresponds to each data symbol and contains pixels of width $I/M$ and height $J/N$.

The cell $C_{m,n}$ is defined as follows, where $\alpha$ is the signal intensity.

$$C_{m,n} = \begin{cases} 
\begin{bmatrix} 0 & \ldots & 0 \\
\vdots & \ddots & \vdots \\
0 & \ldots & 0 \\
\alpha & \ldots & \alpha \\
\vdots & \ddots & \vdots \\
\alpha & \ldots & \alpha 
\end{bmatrix} (d_{m,n} = 0) \\
\begin{bmatrix} 0 & \ldots & 0 \\
\vdots & \ddots & \vdots \\
0 & \ldots & 0 \\
\alpha & \ldots & \alpha \\
\vdots & \ddots & \vdots \\
\alpha & \ldots & \alpha 
\end{bmatrix} (d_{m,n} = 1) 
\end{cases}$$

2.2 Channel
To represent the misalignment of detection of the data signal part at the receiver, $D$ is treated as an image with a margin of $\beta$ on the 4 sides. The translation generates an image in which $D$ is centered or translated by $\beta$ to 8-neighborhood directions, and the projection generates an image in which the 4 corners of $D$ are moved by $\beta$ to the inside or outside or non-moving. After the translation or projection, the smoothing generates a Gaussian filtered image with standard deviation $\sigma_f$, and then Gaussian noise with mean 0 and standard deviation $\sigma_n$ is added on each pixel. Finally, by taking the absolute value, the received image $D_{\text{tfn}}$ is generated.

2.3 Demodulation
The received image $D_{\text{tfn}}$ is input to a trained demodulator to obtain the demodulated data signal $\hat{d}$. The demodulator is constructed by learning various images that reproduce the degradation caused by the communication channel.
3 Data signal demodulation method using machine learning

In our implementation, the transmitted data signal $M \times N$ is set to $4 \times 4$; the cell size $I \times J$ is set to $8 \times 8$ [pixels]; the signal intensity $a=10$; $\beta=2$ for the translation and the projection, and therefore the size of the received image is $36 \times 36$ [pixel].

3.1 Machine learning model

The proposed model of machine learning is shown in Table 1(a). We use a convolutional neural network based deep learning of Keras API of TensorFlow 2. The proposed model is based on the well-known VGG16 [7]; however, the convolution is limited to three layers due to the input image size, and because the ratio of noise to the input image is large in the considered system, the kernel of the convolution is set to $5 \times 5$. In addition, because there is a tendency to lose the information of cells with pixel value 0 if with the max pooling, the average pooling is used. The output layer outputs the likelihood of each of the binary data symbols using the Sigmoid function in the range of 0 to 1. The demodulated data signal $\hat{d}$ is obtained by thresholding the outputs with the center value, 0.5.

3.2 Datasets

The datasets are shown in Table 1(b). The training images includes all data signal patterns, all translation and projection patterns, and several smoothing and noise patterns. The training images and the correct data signals are trained in the following three ways.

(A) The dataset includes smoothing and noise.

(A)+(B) The dataset includes smoothing, noise, and translation.

(A)+(B)+(C) The dataset includes smoothing, noise, translation, and projection.

The loss function is mean-square error, and mini-batch learning is performed, where the mini-batch size is 4096 and the number of training sessions is 15. When inputting these images for training and validation, the input image is preprocessed by linear transformation so that the minimum and maximum of pixel values become 0 and 255, respectively.

Table 1 Machine learning model

| Layer | Output size |
|-------|-------------|
| Input | Received image $36 \times 36 \times 1$ |
| Layer1 | Convolution $(5 \times 5 \times 3)$ $32 \times 32 \times 3$ |
| Layer2 | Average pooling $(2 \times 2)$ $16 \times 16 \times 3$ |
| Layer3 | Convolution $(5 \times 5 \times 128)$ $12 \times 12 \times 128$ |
| Layer4 | Convolution $(5 \times 5 \times 256)$ $8 \times 8 \times 256$ |
| Layer5 | Average pooling $(2 \times 2)$ $4 \times 4 \times 256$ |
| Layer6 | Dropout (0.5) $4 \times 4 \times 256$ |
| Layer7 | Flatten $4096$ |
| Layer8 | Dense (ReLU) $128$ |
| Layer9 | Dropout (0.25) $128$ |
| Layer10 (Output) | Dense (Sigmoid) $16$ |
4 Performance evaluation

4.1 Evaluation methods
We evaluate the performance of the proposed demodulation method by using the bit error rate and comparing it with the threshold decision method used in the previous studies. The validation images are processed in the same way as the training images and used for the evaluation. Different from the training images, these are unknown images for the demodulator because of the random Gaussian noise. The demodulator does not know the correct data signal, the translation, projection, smoothing, and noise parameters of the validation images, and its output is decided only by the weights obtained through training.

To clarify the effect of the training contents on the demodulation performance in the three trained neural networks with (A), (A)+(B), and (A)+(B)+(C) shown in Section 3.2, the evaluation are performed for each of the validation data (A), (B), and (C). The cell average method, which is the target of comparison, thresholds the average pixel values in each cell without considering the translation and projection for the validation images. Here, the threshold value is set to 5.0 because the signal intensity $\alpha = 10$.

4.2 Evaluation Results
Figure 2 shows the bit error rates of the cell average method and the proposed method. The performance of the proposed method is shown for training the three different datasets. In each method, the left bar (blue), the center bar (orange), and the right bar (black) shows the bit error rates for the evaluation of the validation data (A), (B), and (C), respectively.
Compared with the cell average method, the proposed method significantly improves the bit error rate for the learned image in both cases. In the cell average method, because the average value of each data cell is highly affected by random noise and interference among neighboring data cells due to misalignment and blur, a fixed threshold value cannot make a correct decision. In contrast, the proposed method acquires positional universality by using weights acquired by learning images with various type of misalignment, blur, and noise in advance, and can correctly decide data cells even if with ambiguous boundaries and noise.

5 Conclusion

This paper proposed a signal demodulation method using machine learning for digital signage and image sensor VLC. The proposed demodulator with a convolutional neural network was trained by various datasets including smoothing, noise, translation, and projection that may occur in received images. Through the evaluation of the bit error rate and the effect of training datasets, it was shown that the proposed method is superior to the conventional method.

Acknowledgments

A part of this work was supported by The Naito Research Grant, JSPS KAKENHI Grant Number 21K04047, and Joint Research Program of IMaSS, Nagoya Univ.