2022 International Symposium on Nonlinear Theory and Its Applications, NOLTA2022, Virtual, December 12-15, 2022

Recognition Using YOLO for Degraded Images on Visible Light Communication

Hiroko Matsuda, Haruna Matsushita† and Shintaro Arai‡

†Dept. of Electronics and Information Engineering, Kagawa University
‡Dept. of Electrical and Electronic Engineering, Okayama University of Science
Email: haruna@ieee.org, arai@ee.ous.ac.jp

Abstract—This paper focuses on the visible light communication (VLC) and proposes a signal demodulation method using a You Only Look Once (YOLO). Especially, VLC used in intelligent transport systems (ITS), we assume that a transmitter is a LED traffic light and a receiver is a high-speed on-vehicle camera. The VLC requires a high image recognition accuracy from captured images and a high-speed decoding processing. In addition, it is necessary to be able to accurately recognize images even if the captured image is degraded by noise. This study proposes a YOLO-based demodulation method for the VLC. Simulation results show that our method was able to exactly demodulate the transmitted images with almost 100% probability. The confidence score of the detected lighting patterns was above 80%.

1. Introduction

A visible light communication (VLC), which communicates by the lighting pattern of Light Emitting Diode (LED), has been attracted attention [1]. The VLC does not affect other machines with electromagnetic waves, and a transmission source is easily recognized. In this paper, we study the VLC especially in intelligent transport systems (ITS). The VLC requires a high image recognition accuracy from the captured image and a high-speed decoding processing. In addition, it is necessary to be able to accurately recognize images even if the captured images are degraded by noise because vehicles using the VLC receive traffic information in real time in the ITS.

In a previous study, a decoding method by using Gaussian filter estimates the degree of images degradation [2]. This previous method requires finding out exactly the position and shape of the LED array in the captured image. On the other hand, some previous studies based on a deep learning (DL) were also proposed [3]. However, the DL-based decoding method takes time to identify the transmitted lighting pattern from the captured image.

This study proposes a VLC decoding method using a You Only Look Once (YOLO) [4]. The YOLO is a machine learning for the object detection, and it is capable of high-speed processing and detection with high accuracy. In the proposed method, the input data of the YOLO is images captured by a high-speed camera. Simulation results show that the proposed method can exactly demodulate the transmitted images including noise or blur with almost 100% probability, without information on the position or shape of the LED array in the captured image. Furthermore, we confirm that the confidence score of the detected lighting patterns is above 80%.

ORCID iDs Hiroko Matsuda: 0000-0002-4358-5964, Haruna Matsushita: 0000-0002-7850-5119, Shintaro Arai: 0000-0002-4562-9385

2. You Only Look Once (YOLO)

This study uses the YOLO for the recognition of the LED lightning pattern on the ITS because of its versatility and speed in detecting and identifying. The YOLO is a machine learning that performs “detection” and “identification” at the same time, thus it can detect objects faster than other machine learning techniques. The YOLO performs detection and identification by following three steps.

1. The input image is divided into $S \times S$ regions, called grid cell.
2. Do “detection” and “identification” at the same time.
   (a) For each grid cell, estimate “bounding box” (BBox) and “Confidence”.
   (b) Each grid cell has a conditional class for each of the object class
3. Combining BBox’s confidence and each gridcell’s predicted establishment, respectively. We calculate “Confidence score”.

By performing detection and identification simultaneously, the YOLO can perform high-speed image detection and improve recognition accuracy. Furthermore, the YOLO is capable of highly accurate object detection, even when many objects have been captured in the image. These features of the YOLO are advantageous in the ITS using the VLC, thus we use the YOLO in this study.

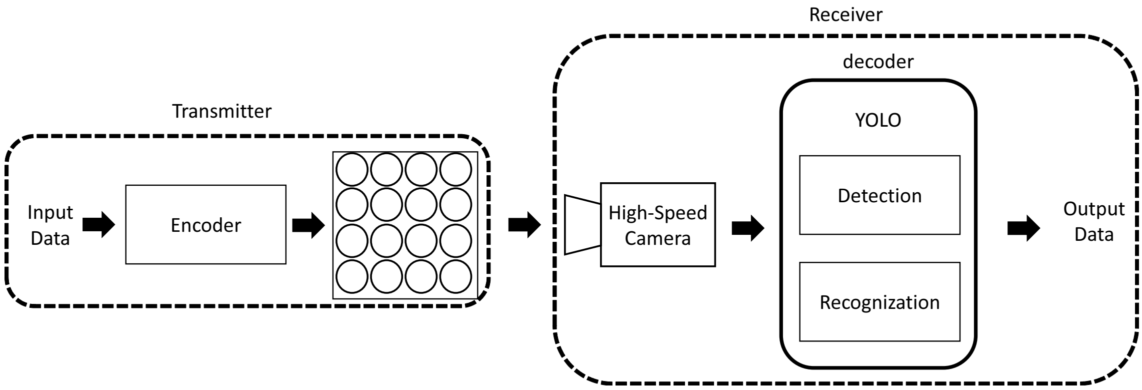
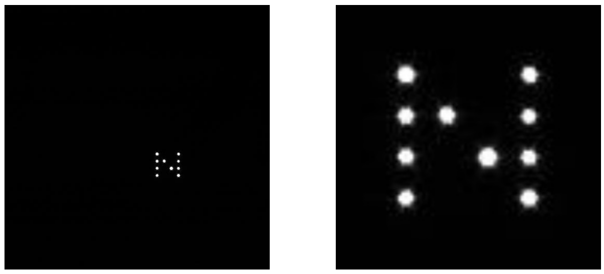


Figure 1: The system model assumed in this study

3. Proposed system model of the image sensor communication

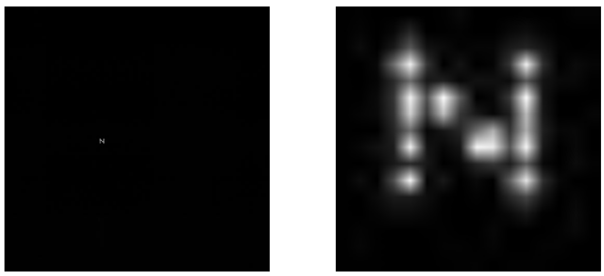
This section explains a system model of our image sensor communication shown in Fig. 1. The transmitter is composed of an encoder, a mapper, and an LED array arranged in a matrix of $M \times N$ LEDs. Parallel data transmission is performed by modulating different transmitted data via on-off-keying (OOK: On = 1, Off = 0) with each LED.

The receiver is composed of a high-speed image sensor and a decoder. The transmitted optical signals pass through the spatial optical channel before they arrive at the receiver. The receiver uses the image sensor to capture these optical signals at a high frame rate and converts them to electrical signals. This study uses the YOLO as the decoder unit.



(a) Captured image (b) Enlarged image

Figure 2: One of the non-degraded image



(a) Captured image (b) Enlarged image

Figure 3: One of the degraded image

4. Simulation

In this paper, we create images that are supposed to capture in the VCL and investigate whether LEDs are detectable using the YOLO. Table 1 summarises the capturing condition of the input images corresponding to the transmitted data. We use 4×4 LED array, that models LED traffic lights in the transmitter. Although we can obtain $(4 \times 4)^2 = 256$ lightning patterns using the LED array in theory, this study considers only 38 different lighting patterns that contains 32 patterns and based on the assumption that 5 bit data 5 patterns similar to the 32 patterns.

The images for training and testing contains non-degraded and degraded images. Figure 2 is one of the degraded images using in this study. Because the non-degraded images were captured from near the LED array, the LED lighting pattern is clearly recognized in the captured image. Figure 3 is one of the degraded images using in this study. The degraded images were captured by placing the camera at a distance from the LED array, and the image includes blur.

Table 2 shows the simulation conditions. In this study, we performed three kinds of simulations using different images with different features for the training and testing dataset. A Simulation 1 uses non-degraded images for both training and testing images. A Simulation 2 uses non-degraded images for training as Simulation 1 used, however, it used both degraded and non-degraded images for testing. A Simulation 3 uses both non-degraded and degraded images for both training and testing. We perform the simulations under the above three conditions to investigate the accuracy and confidence score. The training and testing images are randomly selected from the dataset consisting of the non-degraded and degraded images.

Table 1: Capturing conditions of the input images

	Non-degraded images	Degraded images
Patterns	38	
Shooting distance	3–5m	7–11m
Focal length	35mm	12.5mm
Focus	Infinity	
Aperture	F 16, F 4	F 16

Table 2: Simulation conditions

	Simulation 1	Simulation 2	Simulation 3
Network	YOLO (ver.5)		
GPU	NVIDIA GeForce GTX 1050 Ti		
Pattern	38		
Training image	Non-degraded		Non-degraded and degraded
Number of training images	11400		22800
Batch size	10		
Epoch	30		
Testing image	Non-degraded	Non-degraded and degraded	
Number of testing images	3800	7600	

5. Simulation results

The results of three simulations are summarized in Table 3. In this table, the accuracy are calculated by the average of 10 runs that uses different the testing images. Detection time per image written in Table 3 is the processing time required to detect LEDs in each image. The confidence scores written in Table 3 is the mean value of the confidence scores for all lighting patterns.

In the Simulation 1 and Simulation 3, we can see that the proposed method demodulated the transmitted LED patterns with 100%. Although extremely rare, as shown in Fig. 4, the proposed method wrongly detected a LED that does not exist in the detected area in the image. However, this confidence score that detected lighting pattern is very low. Therefore, it is possible to deal with this problem without outputting unless the confidence score is above a certain value. Figure 5 shows the average value of confidence score of each lighting pattern on the Simulation 3. We can see that all the confidence scores show also above 80%.

In the Simulation 2, the decoding accuracy was 51.7%. The results means although the proposed method can identify the non-degraded images with high accuracy, it was hard to exactly recognize the degraded images. Then, it exactly decorded only 208 images out of 3800 degraded images. Furthermore, the confidence score of degraded images was 30% that is very low.

From these results, we conclude that the proposed method can decode the lighting pattern with high accuracy and high confidence score by training both degraded images and non-degraded images.

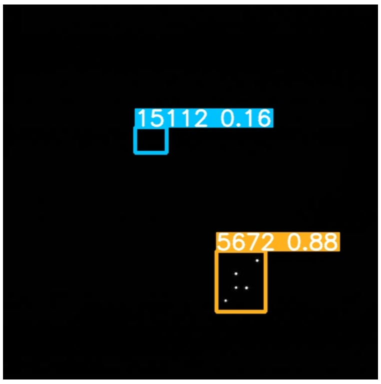


Figure 4: One of results in the Simulation 1. The proposed method wrongly detected a LED although no LED exists in the blue rectangle area in the image.

6. Conclusions

In this paper, we have proposed the YOLO-based decoding method for VLC. The simulation results have shown that the proposed method can decode the lighting pattern with high accuracy and high confidence score by training both degraded images and non-degraded images.

In this study, we considered the transmitted image including a single LED array. In our future work, we should consider the decoding method from the transmitted images including multiple LED arrays.

References

[1] T. Yamazato, "Visible Light Communications Using LED Array and High-speed Camera for ITS Applica-

Table 3: Simulation results

	Simulation 1	Simulation 2	Simulation 3
Training time	10.751 hours		21.407 hours
Testing image	Non-degraded	Non-degraded and degraded	
Number of testing images	3800	7600	
Number of detectable images	3800	3932	7600
Accuracy	99.9%	51.7%	100%
Detection time per image	0.02121 s	0.02132 s	0.02139 s
Confidence score	87%	45%	84%

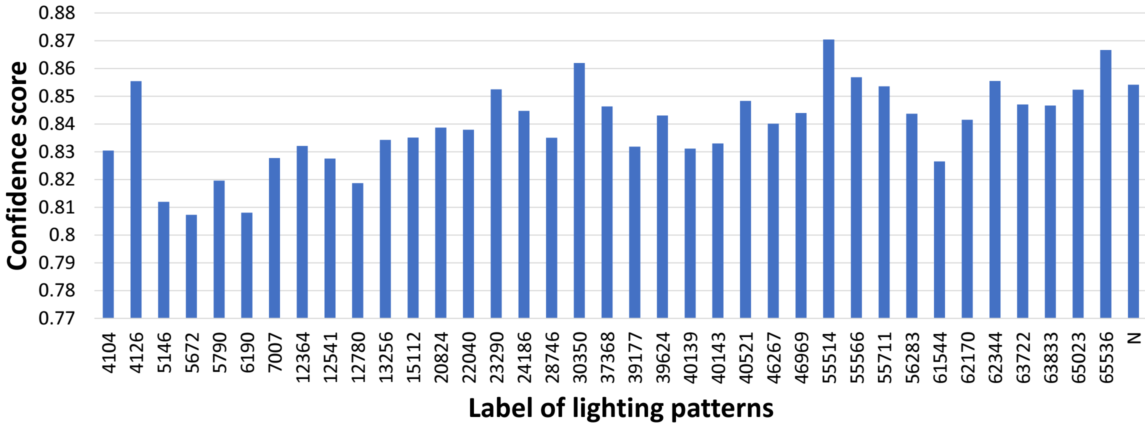


Figure 5: The confidence scores in the Simulation 3

tions", *IEICE ESS, Fundamentals Review*, Vol. 3, NO. 2, pp. 45–53, 2009 (in Japanese).

[2] Y. Ohira, S. Arai, T. Yendo, T. Yamazato, H. Okada, T. Fujii and K. Kamakura, "Signal Demodulation Method from Degraded Image for Visible Light Communication Using LED and High-Speed Camera", *IEICE Trans. B*,Vol. J98-B, No. 7, pp. 696–706, 2015

[3] H. Matsushita and S. Arai, "A Study on Image Recognition Using Convolutional Neural Network for Image Sensor Communication", *2019 IEICE NOLTA Society Conference*, No. NLS-16, p. 16, 2019 (in Japanese).

[4] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection", *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779–788, 2016.