

# Image Sensors Meet LEDs

Koji KAMAKURA<sup>†a)</sup>, *Member*

**SUMMARY** A new class of visible light communication (VLC) systems, namely image sensor (IS) based VLC systems, has emerged. An IS consists of a two-dimensional (2D) array of photodetectors (PDs), and then VLC systems with an IS receiver are capable of exploiting the spatial dimensions invoked for transmitting information. This paper aims for providing a brief survey of topics related to the IS-based VLC, and then provides a matrix representation of how to map a series of one dimensional (1D) symbols onto a set of 2D symbols for efficiently exploit the associate grade of freedom offered by 2D VLC systems. As an example, the matrix representation is applied to the symbol mapping of *layered space-time coding (L-STC)*, which is presented to enlarge the coverage of IS-based VLC that is limited by pixel resolution of ISs.

**key words:** *image sensor, LED, visible light communication (VLC), optical wireless communication (OWC), two-dimensional communication, space-time coding (STC)*

## 1. Introduction

The rapid increase in the usage of LEDs has lead to the advent in recent years of visible light communication (VLC). VLC is emerging as a solution to overcome the crowded radio spectrum for wireless communication systems. In VLC, information is transmitted by modulating the intensity of an optical light source operating in visible range of the spectrum at a rate much faster than the response time of the human eye. LEDs are the most suitable light source for VLC due to their capability of high switching rate.

Initial studies on VLC using LED lighting were done in [1], where data can be provided from a local aggregation point to the luminaries via existing infrastructure like traffic lights and a single photodiode (PD) was used as a receiver. Nowadays, thanks to low-cost complementary metal oxide semiconductors (CMOSs) that consume less power while delivering high performance and advanced features, image sensors (ISs), which are an two-dimensional (2D) array of PDs, have been embedded in today's low-end cell phones and tablet devices. Therefore, in order to receive the modulated LED lights, there are two types of receiver: PDs [2]–[5] and ISs [6]–[9].

Single-element PD receivers have an advantage of converting optical signals at very high rates, but they suffer from large interference and background light noise. This results in very low SNRs and thus short communication ranges such

as offices and rooms. To increase the data rate of the single-transmitter single-receiver system using the same amount of signal power, multiple-input multiple-output (MIMO) are investigated, where the parallel data streams can be transmitted over either different spatial locations [4] or different colors of light [5]. In [5], 8 Gbps is reported to be achieved over a 1-meter indoor link with an red-blue-green-yellow LED and R/G/B/Y filters. The main drawbacks of this approach are potentially large size and high cost of the multiple receiving elements.

On the other hand, IS receivers consist of a 2D array of PDs, each of which converts incident light to an electrical signal and has it as a digitized value individually. Therefore, their outputs are essentially discrete. Furthermore, they are compact, which constitute a camera structure. The camera lens provides a different narrow field of view (FOV) for each pixel. This creates a large number of highly directional receiving pixels, which allows a degree of flexibility of selectively combining pixels that receive a strong signal from the light emitting elements that can not be attained by the type of single-element PD receivers. The tradeoff in the IS receiver, however, is a limited sampling rate of the camera receiver. The frame rate ranges from tens to hundreds frames per second for low-end cameras and a million frames per second for high-end models. Yet, it is reported in [10] that with a customized CMOS IS, a 54 Mbps is achieved, which strikes a good tradeoff between high-speed optical-signal reception and LED-transmitter detection.

This paper focuses on IS-based VLC. IS-based VLC was initially demonstrated in [11] with high-speed camera for intelligent transport systems (ITS), where LEDs are used as road side unit to transmit location or safe driving information from traffic lights to vehicles, which is now extended to communications of vehicle to infrastructure (V2I) [10] and vehicle to vehicle (V2V) [12]. It is also found in [13] that V2V communication with IS receiver is investigated for infrared (IR), instead of visible light.

Meanwhile, low-end camera-equipped cell phones and tablet devices can shoot videos at a frame rate around 20 fps, thanks to the rolling shutter. This shutter operation is investigated to receive data at a rate that far exceeds the frame rate of the camera [14]–[17].

VLC research was started off in 2003 by the Visible Light Communications Consortium (VLCC) [18], which is the successor of Visible Light Communications Association (VLCA), receiving increasing interest worldwide. They term it as image sensor communication (ISC). Con-

Manuscript received September 2, 2016.

Manuscript revised November 14, 2016.

Manuscript publicized December 14, 2016.

<sup>†</sup>The author is with Department of Computer Science, Chiba Institute of Technology, Narashino-shi, 275-8588 Japan.

a) E-mail: kama@cs.it-chiba.ac.jp

DOI: 10.1587/transcom.2016LCI0001

sequently, the visible light spectrum is now in a state of action after the IEEE standardization of VLC technology in 2011 [19]. In the Task Group (TG7r1), ISC, which is also referred to as optical camera communications (OCC), is discussed as a part of optical wireless communication (OWC) that uses the camera sensor as a receiver.

The aims of this paper are two folds. First, it aims at providing a brief survey of IS-based VLC systems, while avoiding duplication of efforts done in [12], [20], [21], which give excellent introductions to IS-based VLC. Section 2 gives a brief explanation of IS used for the receiver, and modulation with lighting constraints is mentioned in Sect.3. Since VLC and OWC have in common such as conventional digital modulation techniques, the interested reader will also be provided with references [3], [22], [23] for a more in-depth study of the techniques. Secondly, this paper aims for providing a matrix representation of how to map one-dimension (1D) symbols onto a 2D array of transmitting LEDs for IS-based VLC, which includes the area termed ISC, CamCom, and OCC. Section 4.1 mentions intrinsic limits in IS-based VLC systems, and then Sect.4.2 describes the bit-to-symbol mapping of 1D signals to 2D ones for the 2D system, with a matrix representation. The matrix representation can be easily adapted to make efficient use of spatial dimensions of rows and columns of the transmitting device array as additional dimensions invoked for transmitting information. This matrix expression is applied to a modulation for IS-based VLC, which is presented as *layered space-time coding (L-STC)* in [24]. More specifically, with the matrix expression, the symbol mapping of three-layered STC is rewritten so that readers can easily understand how it is done for expanding the coverage of IS-based VLC through L-STC in Sect. 4.3. Finally, conclusions are presented in Sect. 5.

## 2. IS-based VLC Components

### 2.1 CMOS Image Sensor

A CMOS IS has an imaging area, which consists of an array of pixels, vertical and horizontal access circuitry, and readout circuitry, as shown in Fig. 1. The imaging area is a 2D array of pixels, each of which has a PD and some transistors inside. CMOS chips have transistors at each pixel to move the charge through lines. This offers flexibility because each pixel is treated individually, unlike CCD [25], [26]. Access circuitry is used to access a pixel and read the signal value in the pixel. For this purpose, a shift resistor is used, and a decoder is used to access pixels randomly. A readout circuit is a 1D array of switches and a sample-and-hold circuit.

#### 2.1.1 Sensitivity

In terms of sensitivity of IS inside a camera, two parameters are very important: quantum efficiency and read noise.

Quantum efficiency is a measure of how efficiently the IS converts light (photons) to charge (electrons). The

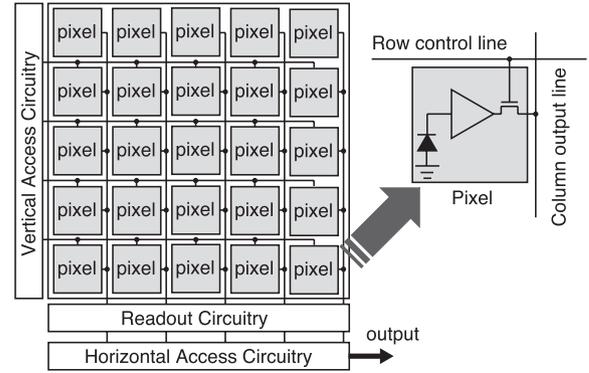


Fig. 1 Architecture of a CMOS IS.

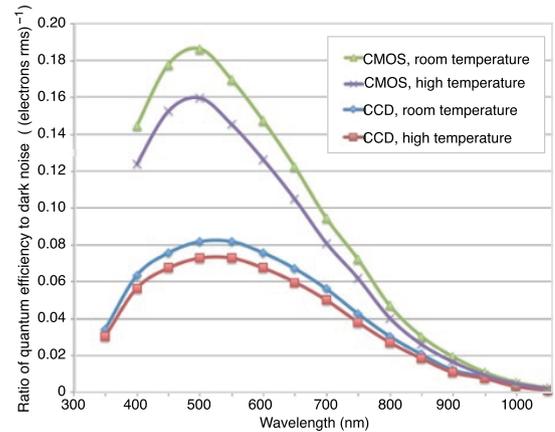


Fig. 2  $\eta_Q/N_D$  versus wavelength at the frame rate of 30 fps (by Adimec).

more electrons into a pixel during the integration period, the higher the output level of the IS, so the more sensitive the IS is for that specific wavelength of the light.

Read noise is the random noise generated within the charge-to-voltage conversion amplifier when reading out the charge. It has the noise level equivalent to the output of the IS when the IS is in the dark place during zero integration time. The lower the read noise level is, the lower the minimum number of signal electrons is that can be detected. Therefore, lower read noise results in a more sensitive IS. Recent CMOS technology offers an extremely low read noise that are close to high performance CCD ISs [27].

More precisely, the noise level at the output of the IS is not only determined by the read noise, but the dark current of a pixel also contributes to the dark noise. The dark current is dependent on the quality of the IS, and also very dependent on the temperature and the integration time. At higher temperatures the dark current increases rapidly and then the dark noise will be larger than the read noise. For temperature, the dark noise should be considered rather than the read noise [27].

With the ratio of quantum efficiency  $\eta_Q$  to dark noise  $N_D$ , as the ratio increases, the IS is more sensitive at a particular wavelength. Fig. 2 shows the ratio for recent CMOS

and CCD ISs at two temperatures: room temperature and a higher temperature. The curves show that the CMOS IS is more sensitive than the CCD one at all wavelengths. We can also see that the ISs perform worse at higher temperatures than at room temperature. The sensitivity of the camera in lower light conditions is necessary for VLC scenarios. The ratio gives the overall sensitivity of the IS, or the minimum amount of light detectable. Nowadays, CMOS ISs have a higher efficiency than CCD ones at all wavelengths.

### 2.1.2 Shutter Operations

In digital cameras, the luminance value generated by the IS depends on the voltage signal, which is induced by electron charge. The electron charge is related to the amount of light that falls on the photosensitive area, in terms of duration. They therefore require some form of shutter to control exposure time. This is generally achieved by incorporating either an external mechanical shutter in front of the IS or an on-chip electronic shutter.

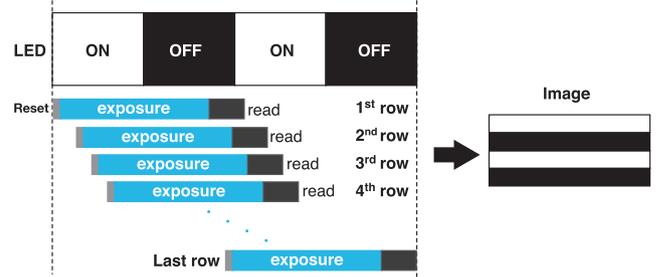
In the point of view for VLC applications, we focus on two types of electronic shutter: global shutter and rolling shutter.

**Global shutter** controls incoming light to all pixels simultaneously. At any given point in time, all pixels of a frame (image) are therefore either equally closed or equally open. All pixels are reset simultaneously, and integration of all pixels begin at the same time. Therefore, data integrated over the same time duration is output from all pixels, which is desirable for VLC receivers.

**Rolling shutter** is a readout architecture, in which, not all portions of the image start and stop receiving light at the same time. In order to maintain high fill factor and read out speed, most CMOS ISs are equipped with column-parallel readout circuits, which simultaneously read all pixels in a row into a line memory. The readout proceeds row-by-row, sequentially from top to bottom, as shown in Fig. 3. This is called “rolling shutter” operation.

High resolution image requires the increase in the number of PDs, but this significantly reduces the frame rate, which is the number of frames per second (fps) that can be captured by the IS. Thanks to the rolling shutter, recent mobile devices have ISs to capture videos at a certain frame rate. Even the least expensive smartphones nowadays are equipped with high-resolution CMOS IS inside cameras, they can shoot full HD videos at 30 or even 60 fps. However, this means that direct use of the IS for VLC receivers can provide very low data rate.

Even though the advance in CMOS technology has improved the resolutions of ISs to very high levels, it is still difficult to have high-frame-rate camera with such high-resolutions at low cost. Therefore, to use the most of the rolling shutter mechanism of CMOS ISs for receive multiple bits modulated at a high frequency within one frame, an approach is presented [9], which achieves a few kbps of throughput with the rolling shutter process of camera sensor operating at 20 fps. When capturing an image, most CMOS



**Fig. 3** Rolling shutter effect observed when capturing the LED blinking at a rate faster than the frame rate.

ISs expose one or more columns of pixels, but read out only one column at a time, sweeping across the image at a fixed scan rate to create a rolling shutter, as shown in Fig. 3. First, an LED is in ON which results in higher intensity output for pixels of the first row. At the next time instance, when the LED changes its state to OFF, the IS reads it out as low intensity output for pixels of the second row. After all the rows are scanned, the resultant image can be converted to binary data. This effect is neither visible to the naked eye, nor in a photograph that uses an auto-exposure setting. This idea is applied to for vehicular communication [28], positioning [29], and encryption [30].

## 3. Modulation under Lighting Constraints

### 3.1 Dimming

Different levels of illuminance are required when performing different types of activities [31]. For example, an illuminance in the range of 30–100 lux is often said to be enough for simple visual tasks performed in most public places, while office or residential applications require higher level of illuminance in the range of 100–1000 lux. The luminance is measured in lux ( $\text{lux} = \text{lm}/\text{m}^2$ ), which is the luminous flux incident per unit area. The lm (lumen) is the unit of luminous flux, which represents the light power of sources as perceived by human eye. It is defined as a monochromatic light source emitting an optical power of 1/683 watt at 555 nm has a luminous flux of one lumen. The light of a full moon provides an illumination of about one lux, and on a sunny day outdoors, the illumination would be about 100,000 lux.

With the advancements in LED driver circuits, it has become possible to dim an LED to an arbitrary level depending on the application requirement to save energy. If an LED can be dimmed to an arbitrary level, it is also necessary to understand its impact on the human perceived light. It was first shown in [32] that the relation between the measured light and the perceived light is nonlinear. The perceived light can be calculated from the measured light as [32], [33]

$$\text{Perceived light (\%)} = 100 \times \sqrt{\frac{\text{Measured light (\%)}}{100}}. \quad (1)$$

For example, at full brightness, the measured light in a space

is 60 foot-candles (fcd). Eq. (1) indicates that a lamp that is dimmed 1% (0.6 fcd) of its measured light is perceived to be 10% dimmed by the human eye. This is important in terms of VLC because a user may choose an arbitrary level of dimming depending on the application or desired energy savings, but the communication should not be affected by the dimming; i.e., the data should be modulated in such a way that any desired level of dimming is supported.

### 3.2 Flicker Constraint

An additional constraint on any VLC modulation is that it should not result in human perceivable fluctuations in the brightness of the light. Since the flicker has little impact on human eye perception when light pulses blink at 200 Hz or higher frequency [19], the average intensity is constrained. This constraint relates the intensity and the signal level itself. This new constraint, which is referred to as the dimming target, introduces a new domain of system design which has rarely been considered in existing communication media.

### 3.3 Modulation with Lighting Constrains

A typical non-return-to-zero (NRZ) on-off keying (OOK) has 50% average intensity for equiprobable binary symbols. For the lighting constraint of 75% dimming, it simply shifts the OFF symbol level from 0% intensity level to 50%, which is referred to as analog dimming. However, a non-linear characteristic of LEDs poses some difficulties and further the reduced level margin faces on difficulties, and the reduced margin between the two intensity levels degrades the signal detection performance.

One solution to mitigate long runs of 1s and 0s as on and off states is to compensate the intensity difference in time. For a general data transmission rate with equiprobable binary symbols, the same duration of dummy ON transmission is appended to meet the target 75% dimming. If the target is below 50%, a dummy OFF symbol interval is appended to resolve the difference. The dummy transmission is presented in the IEEE standard [32]. On the other hand, pulse width modulation (PWM) is used for providing a marginal rate enhancement [34]. PWM can be also superposed with OOK and pulse position modulation (PPM) for dimming support [35].

Another solution is to use *run length limited* (RLL) codes. RLL codes ensure that the output symbols have balanced repetition of 0s and 1s. Examples of commonly used RLL codes include Manchester, 4B6B, and 8B10B coding. In Manchester coding, a 0 is replaced with a “down” transition (one-to-zero) and a 1 is replaced with an “up” transition (zero-to-one). 4B6B coding maps a 4-bit symbol to a 6-bit symbol that has balanced repetition. Similarly, 8B10B maps an 8-bit symbol to a 10-bit symbol. Since the Manchester coding has the highest in the number of additional bits added, it can be a suitable choice for low data rate services that require better DC balancing. On the other hand, 8B10B

reduces the number of additional bits added, but it performs poorly in terms of the DC balancing.

Furthermore, a dimming technique in VLC is inverse source coding [36], which uses the ratio of 1s and 0s for satisfying the required dimming target. For example, if the dimming ratio is 70%, then 70% of the output bits must be 1s. The encoding scheme increases the number of transmitted data bits to achieve the dimming ratio. Huffman coding is generally used as the encoding scheme, where only an approximation to the required dimming ratio is achieved.

## 4. Two-Dimensional Systems

### 4.1 Four Distortions

Most of the luminaries typically contain multiple LEDs to provide sufficient illumination. When these multiple LEDs can be treated as an array of multiple transmitters that works as a 2D transmitter. VLC systems between such a transmitter array and the corresponding IS receiver constitutes a 2D system. The 2D system has attractive mapping options due to additional dimensions invoked for transmitting data.

The IS inside cameras can provide a wide receiving FOV that allows for node mobility without the need to realign the receiver. Yet, because of the camera design, each pixel has a very narrow FOV, allowing high gain communication. The camera lens creates the effect of each pixel being angled to a slightly different part of the scene, so that the combination of all LEDs of the array generates an image with a wide FOV.

The 2D system provides a new domain to transmitting data, but exhibit four main types of distortions below.

**1) Perspective distortion** is a common phenomenon in our daily life. When we take a look at a rectangular array from a certain angle, the image on the plane is like a trapezoid, as shown in Fig. 4(a). In the typical case of a projector sitting on a table and looking upwards to the wall screen, the image is larger at the top than on the bottom. Consequently, some LEDs on an array expand at the IS plane inside camera, while others shrink [37].

**2) Blurring** occurs when the camera moves while cap-

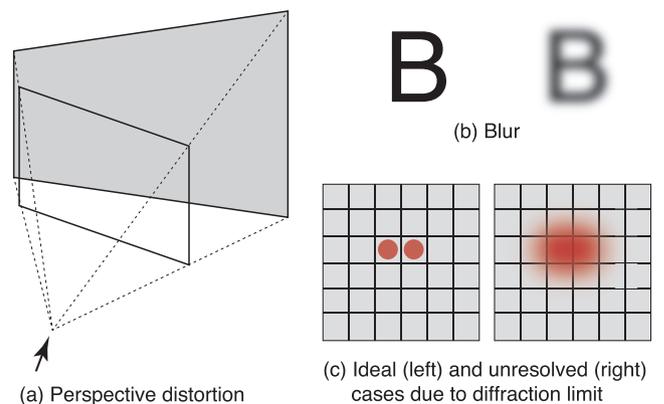


Fig. 4 Examples of distortions considered in the imaging system.

turing an image or a lack of focus can introduce blur in the image, which causes the pixels to blend together, as in Fig. 4(b). An array-camera communication system must be able to deal with such blending and still successfully recover the transmitted bits. In the frequency domain, blur can be considered a low-pass filter where high frequency attenuates much more than the low frequency.

**3) Ambient Light** is a source of noise for array-camera links because it changes the luminance of the received pixels. This can cause errors in the information bits decoded from the pixels. In frequency domain, it can be considered the DC component.

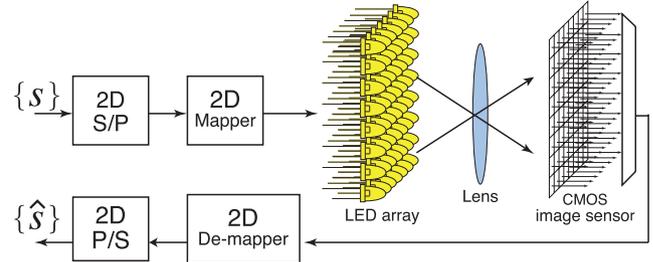
**4) Diffraction limit** determines the angular resolution, thus limiting the pixel resolution in any imaging system. Light is collected by a lens which limits the spatial extent over which it can be received. If the wavelength of light is  $\lambda$  and the diameter of the lens' aperture is  $D$ , the angular resolution set by the diffraction limit is  $\theta = 1.22\lambda/D$  radians [24]. For example, in an imaging system operating at  $\lambda = 630$  nm with  $D = 2$  mm, the angular resolution is about  $3.84 \times 10^{-4}$  radians (79.3 arc-seconds), corresponding to a spatial resolution of 76.9 mm at a distance of 200 m. This means that IS receivers cannot discern two points at the distance of 200 m when they are separated from each other by 76.9 mm or less. In such a diffraction-limited region, intensity-modulated signals transmitted from two 20 mm-away LEDs disperse and become unresolved LED pair, as shown in Fig. 4(c), thus impossible to demodulate from them.

In real world scenarios, we further face i) camera motion, ii) illumination variation, and iii) background distracters such as other vehicles on the road [38]. Camera motion is inherently present in the visual MIMO communications system because the camera at the receiver and LED transmitters are on different mobile nodes. Consequently, the geometry of the image formation process varies, i.e., the position and orientation of the camera center with respect to the scene varies. As the camera moves further, the object of interest appears to become smaller. Because of this perspective projection, the LED array undergoes arbitrary scaling, and the standard communication approach of template matching with matched filters or with correlation-based detectors might be insufficient. The computer vision literature has numerous methods for achieving scale invariance in object recognition, which often require heavy processing.

## 4.2 Matrix Expression of Transmitter Design of 2D Systems

In order to efficiently exploit the spatial degree of freedom of the 2D system, this section gives an matrix expression.

Consider a 2D communication system, composed of an LED array transmitter, an optical channel, and an IS receiver, as shown in Fig. 5. We assume an LED array as the 2D transmitting device here, but it can be a pixelated array such as computer display, screen, or digital signage [7], [37], [41]. In general, an input to the 2D system can be



**Fig. 5** An overview of a 2D optical wireless communication system between an LED array and an IS.

a one-dimension stream of complex symbols. The output of the 2D mapper inside the 2D system will be a set of unipolar signals (i.e., real but non-negative), and then are transmitted as intensity signals from the LED array. Such signals pass through the optical channel into the IS receiver inside cameras, followed by the 2D demapper with the receiver's output data stream.

Let  $\{s\}$  be a stream of complex symbols into the 2D system. The function of the 2D mapper is to convert  $\{s\}$  into a series of 2D symbols. When  $N_1 N_2$  symbols from  $\{s\}$  are fed to the mapper, it outputs the complex symbol matrix  $\mathbf{S} = \mathbf{S}_I + j \cdot \mathbf{S}_Q$  of size  $N_1 \times N_2$ .  $\mathbf{S}$  should be converted to a unipolar symbol matrix  $\mathbf{X}$  of size  $(M_1 N_1) \times (M_2 N_2)$ , which is an expansion by  $M_1 \times M_2$  inside the 2D mapper. The complex input  $\mathbf{S}$  and the unipolar output  $\mathbf{X}$  are considered in the spatial domain.

The process of the 2D mapper is detailed. Let  $\mathbf{A}_I$  and  $\mathbf{A}_Q$  be two matrices, each of which has size  $M_1 \times M_2$ .  $\mathbf{A}_I$  and  $\mathbf{A}_Q$  have orthonormality constraints, given by

$$\mathbf{A}_I \bullet \mathbf{A}_I = \mathbf{A}_Q \bullet \mathbf{A}_Q = 1, \quad \mathbf{A}_I \bullet \mathbf{A}_Q = 0, \quad (2)$$

where  $\bullet$  denotes the Frobenius inner product. The Frobenius inner product of two real matrices is a scalar, defined as

$$\mathbf{A} \bullet \mathbf{B} = \text{trace}(\mathbf{A}\mathbf{B}^T) = \text{trace}(\mathbf{B}\mathbf{A}^T). \quad (3)$$

The conditions of Eq. (2) do not uniquely determine  $\mathbf{A}_I$  and  $\mathbf{A}_Q$ , leaving enough degrees of freedom while specifying them. We can use this freedom to satisfy a diverse range of practical requirements and to design a meritorious transmission scheme.

To take the benefits of spatial domain consisting of  $(N_1 M_1) \times (N_2 M_2)$  transmitting devices, the mapper generates symbols to modulate the individual transmitting devices separately. Through the expansion process, the two real bipolar matrices  $\mathbf{S}_I$  and  $\mathbf{S}_Q$  are expanded into  $\mathbf{G}_I$  and  $\mathbf{G}_Q$  of size  $(M_1 N_1) \times (M_2 N_2)$  by

$$\mathbf{G}_I = \mathbf{S}_I \otimes \mathbf{E}_{1,1}, \quad \mathbf{G}_Q = \mathbf{S}_Q \otimes \mathbf{E}_{1,1}, \quad (4)$$

where  $\mathbf{E}_{1,1}$  is an  $M_1 \times M_2$  matrix with all zero elements except a single one at the index  $(1, 1)$ , and  $\otimes$  is the tensor multiplication operation of the two matrices. If  $\mathbf{P}$  is an  $m \times n$  matrix and  $\mathbf{Q}$  is an  $x \times y$  matrix, then the tensor multiplication  $\mathbf{P} \otimes \mathbf{Q}$  is the  $mx \times ny$  matrix:

$$\mathbf{P} \otimes \mathbf{Q} = \begin{bmatrix} p_{11}\mathbf{Q} & \cdots & p_{1n}\mathbf{Q} \\ \vdots & \ddots & \vdots \\ p_{m1}\mathbf{Q} & \cdots & p_{mn}\mathbf{Q} \end{bmatrix}, \quad (5)$$

each element of which is expressed as

$$p_{ab}\mathbf{Q} = \begin{bmatrix} p_{ab}q_{11} & \cdots & p_{ab}q_{1y} \\ \vdots & \ddots & \vdots \\ p_{ab}q_{x1} & \cdots & p_{ab}q_{xy} \end{bmatrix}, \quad (6)$$

$a \in \{1, 2, \dots, m\}, b \in \{1, 2, \dots, n\}.$

With this zero padding,  $\mathbf{G}_I$  and  $\mathbf{G}_Q$  have zeros between each of the elements of  $\mathbf{S}_I$  and  $\mathbf{S}_Q$ , respectively. Further, convolving  $\mathbf{G}_I$  and  $\mathbf{G}_Q$  with  $\mathbf{A}_I$  and  $\mathbf{A}_Q$  results in real bipolar 2D signals  $\mathbf{X}_I$  and  $\mathbf{X}_Q$ , respectively, of size  $(M_1N_1) \times (M_2N_2)$ : i.e.,

$$\mathbf{X}_I = \mathbf{G}_I * \mathbf{A}_I, \quad \mathbf{X}_Q = \mathbf{G}_Q * \mathbf{A}_Q, \quad (7)$$

where  $*$  denotes the convolution operation.

The orthogonality of Eq. (2) allows to combine the matrices  $\mathbf{X}_I$  and  $\mathbf{X}_Q$  into a unipolar matrix  $\mathbf{X}$

$$\mathbf{X} = \frac{\mathbf{X}_I + \mathbf{X}_Q}{2c_2} + \frac{c_1}{2}, \quad (8)$$

where  $c_1$  is the peak amplitude of the signal.  $s_I$  and  $s_Q$  are normalized so that each element satisfies  $|s_I|, |s_Q| \leq c_1$ . Meanwhile,  $c_2$  in Eq. (8) is chosen to set the dynamic range of the output signal, according to the VLC channel limitations, thus ensuring unipolar output:  $\mathbf{X} \geq 0$  for any  $\mathbf{S}$ .

### 4.3 Layered Space-Time Coding (L-STC)

With the mathematical expression above, this subsection gives an example, which is a 2D system called *layered space-time coding (L-STC)* introduced in [24], [39]. L-STC is considered binary. Therefore, this example allows us to treat  $\mathbf{S}_Q = 0$ : i.e., we just consider  $\mathbf{S}_I$  as  $\mathbf{S}$  with element  $s_{i,j} \in \{1, 0\}$ ; while it can be complex in 2D systems with orthogonal frequency division multiplexing (OFDM) in [7], [37], [41] to shape the spectrum of the optical output signal.

More specifically, three-layered STC is considered, where the input to the 2D system is three different bit streams called  $\{b\}$ ,  $\{c\}$ , and  $\{d\}$ , each of which is separately space-time coded, before inputting to the 2D system. Two bits  $s_1$  and  $s_2$  in  $\{s\} \in \{b, c, d\}$  are fed to the individual STC encoder, it generates  $s_1 = (s_1, s_2)$  and  $s_2 = (s_2, s_1^*)$ . Note that although  $s_1^*$  indicates the complex conjugate of  $s_1$  in the complex system, in the unipolar system we treat it as the opposite binary state [42]: i.e.,  $s_1^* = \bar{s}_1$ . The purpose of STC encoding is to expand the communication range limited by the pixel resolution of imaging system. In the region where two LEDs are decoupled in the image plane (i.e., detected by one pixel), the combination of two values of the pixel obtained in two frames, the bits are decoded through STC decoding. Note also that although the multiplication of  $j$

will be used in the encoding process explained below, it will be equivalent to  $j \cdot s = \bar{s}$  in the unipolar system.

As shown at the top of Fig. 6, when two bits  $s_1$  and  $s_2$ ,  $\{s\} \in \{b, c, d\}$ , from each of the bit streams are fed to the individual STC encoder, it outputs two matrices  $\mathbf{S}_1$  and  $\mathbf{S}_2$ ,  $\mathbf{S} \in \{\mathbf{B}, \mathbf{C}, \mathbf{D}\}$ , which are of size  $2 \times 2$  with the two bits located in the main diagonal, as

$$\mathbf{B}_1 = \begin{bmatrix} b_1 & 0 \\ 0 & b_2 \end{bmatrix}, \quad \mathbf{B}_2 = \begin{bmatrix} b_2 & 0 \\ 0 & \bar{b}_1 \end{bmatrix}, \quad (9)$$

$$\mathbf{C}_1 = \begin{bmatrix} c_1 & 0 \\ 0 & c_2 \end{bmatrix}, \quad \mathbf{C}_2 = \begin{bmatrix} c_2 & 0 \\ 0 & \bar{c}_1 \end{bmatrix}, \quad (10)$$

$$\mathbf{D}_1 = \begin{bmatrix} d_1 & 0 \\ 0 & d_2 \end{bmatrix}, \quad \mathbf{D}_2 = \begin{bmatrix} d_2 & 0 \\ 0 & \bar{d}_1 \end{bmatrix}. \quad (11)$$

For  $\mathbf{B}_i$ ,  $\mathbf{C}_i$ , and  $\mathbf{D}_i$ ,  $i \in \{1, 2\}$ , we define expansion matrices with different sizes  $4 \times 4$ ,  $2 \times 2$ , and  $1 \times 1$  as  $\mathbf{E}_{1,1}^{(4)}$ ,  $\mathbf{E}_{1,1}^{(2)}$ , and  $\mathbf{E}_{1,1}^{(1)}$ , respectively. Using these expansion matrices according to Eq. (4), we have

$$\mathbf{B}_1^{(8)} = \mathbf{B}_1 \otimes \mathbf{E}_{1,1}^{(4)}, \quad \mathbf{B}_2^{(8)} = \mathbf{B}_2 \otimes \mathbf{E}_{1,1}^{(4)}, \quad (12)$$

$$\mathbf{C}_1^{(4)} = \mathbf{C}_1 \otimes \mathbf{E}_{1,1}^{(2)}, \quad \mathbf{C}_2^{(4)} = \mathbf{C}_2 \otimes \mathbf{E}_{1,1}^{(2)}, \quad (13)$$

$$\mathbf{D}_1^{(2)} = \mathbf{D}_1 \otimes \mathbf{E}_{1,1}^{(1)}, \quad \mathbf{D}_2^{(2)} = \mathbf{D}_2 \otimes \mathbf{E}_{1,1}^{(1)}, \quad (14)$$

where the elements of each matrix are seen in Fig. 6.

Then, in three-layered STC, we define the coding matrices operating on the first and second input matrices  $\mathbf{B}_i^{(8)}$ ,  $\mathbf{C}_i^{(4)}$ , and  $\mathbf{D}_i^{(2)}$  as

$$\mathcal{E}_1^B = [1], \quad \mathcal{E}_2^B = [j], \quad (15)$$

$$\mathcal{E}_1^C = \begin{bmatrix} 0 & j \\ 1 & 0 \end{bmatrix}, \quad \mathcal{E}_2^C = \begin{bmatrix} 0 & 1 \\ j & 0 \end{bmatrix} \quad (16)$$

$$\mathcal{E}_1^D = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & j & 0 \\ 0 & j & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}, \quad \mathcal{E}_2^D = \begin{bmatrix} 0 & 0 & 0 & j \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ j & 0 & 0 & 0 \end{bmatrix}. \quad (17)$$

The encoding process in the spatial domain is done as

$$\mathbf{G}_1^B = \mathcal{E}_1^B \otimes \mathbf{B}_1^{(8)}, \quad \mathbf{G}_2^B = \mathcal{E}_2^B \otimes \mathbf{B}_2^{(8)}, \quad (18)$$

$$\mathbf{G}_1^C = \mathcal{E}_1^C \otimes \mathbf{C}_1^{(4)}, \quad \mathbf{G}_2^C = \mathcal{E}_2^C \otimes \mathbf{C}_2^{(4)}, \quad (19)$$

$$\mathbf{G}_1^D = \mathcal{E}_1^D \otimes \mathbf{D}_1^{(2)}, \quad \mathbf{G}_2^D = \mathcal{E}_2^D \otimes \mathbf{D}_2^{(2)}, \quad (20)$$

where an  $8 \times 8$  matrix  $\mathbf{G}_i^S$  is formed by multiplying  $\mathbf{S}_i$  by each element of  $\mathcal{E}_i^S$ . Note again that the  $j$ -operator indicates the binary opposite.

Then, convolving  $\mathbf{G}_i^S$  with convolution matrices defined as

$$\mathbf{A}^B = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}, \quad \mathbf{A}^C = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad \mathbf{A}^D = [1], \quad (21)$$

results in

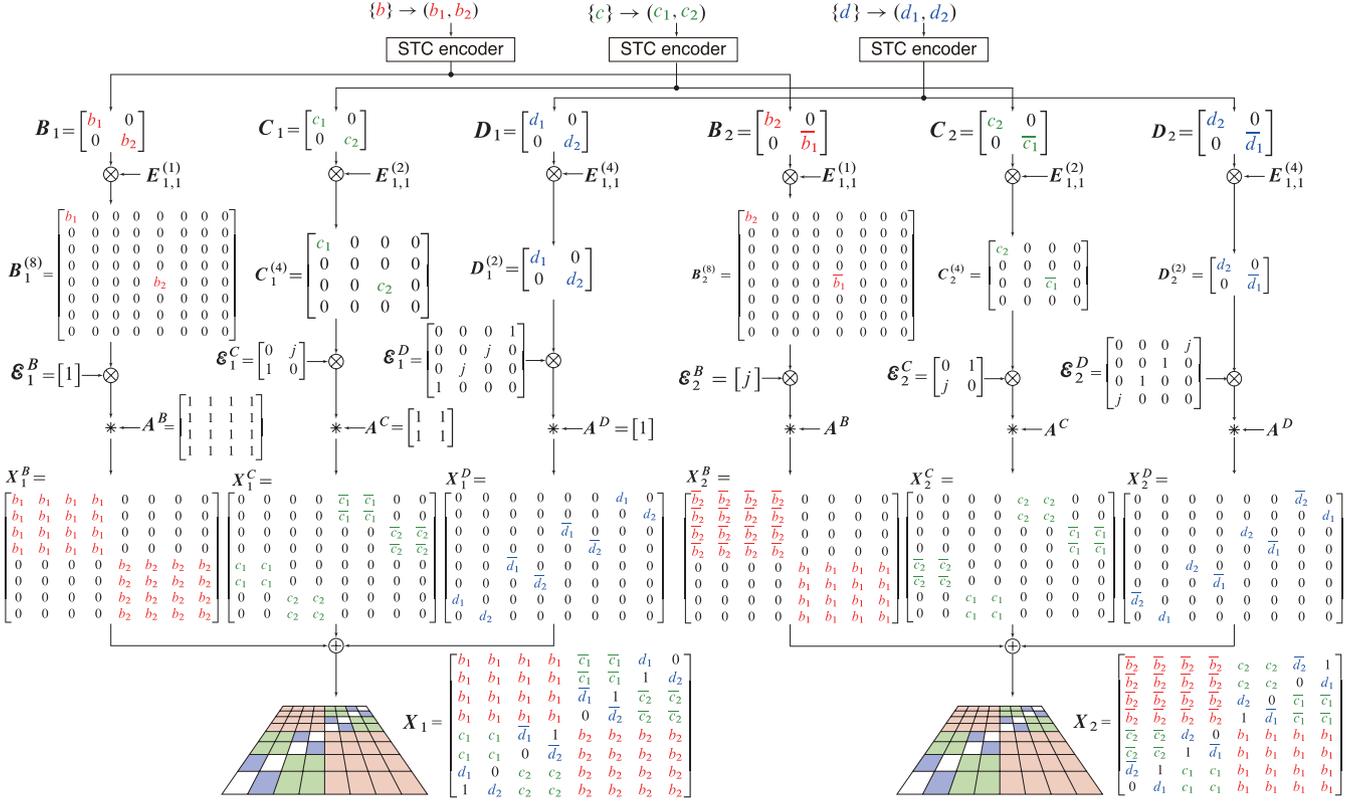


Fig. 6 Matrix notations in the three-layered STC.

$$X_1^B = G_1^B * A^B, \quad X_2^B = G_2^B * A^B, \quad (22)$$

$$X_1^C = G_1^C * A^C, \quad X_2^C = G_2^C * A^C, \quad (23)$$

$$X_1^D = G_1^D * A^D, \quad X_2^D = G_2^D * A^D, \quad (24)$$

where convolution we consider here is different from the normal one. Although in 2D convolution there are wide and narrow convolutions, the one considered here is a hybrid. Zero padding is done for elements at top or left edges of the input matrix (i.e.,  $S$ ) that does not have any neighboring elements to the top and left, while no zero-padding is done for elements at bottom or right edges of the input matrix. To put it more specifically, in convolving  $G^B$  of size  $4 \times 4$  with  $A^B$  of size  $4 \times 4$ , narrow convolution is done by appending three rows and three columns with all zero elements to  $B$ .

Finally, the output matrices of the first and second durations are expressed as, respectively,

$$X_1 = X_1^B + X_1^C + X_1^D, \quad (25)$$

$$X_2 = X_2^B + X_2^C + X_2^D, \quad (26)$$

where the bits are spread over the  $8 \times 8$  matrices, shown at the bottom of Fig. 6.

Two raw images of size  $6 \times 6$  in pixel are shown in Fig. 7, which is captured by high-speed camera operating at 1000 fps at 55 m away from an  $8 \times 8$  LED array. They are resized to  $8 \times 8$  luminance matrices, shown in Fig. 7, through bilinear interpolation. From the two luminance matrices, six bits are successfully extracted. The decision statistics are obtained as combinations of addition and subtraction of

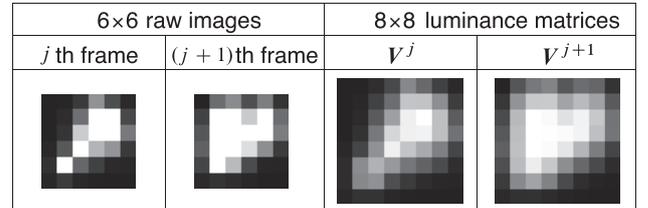
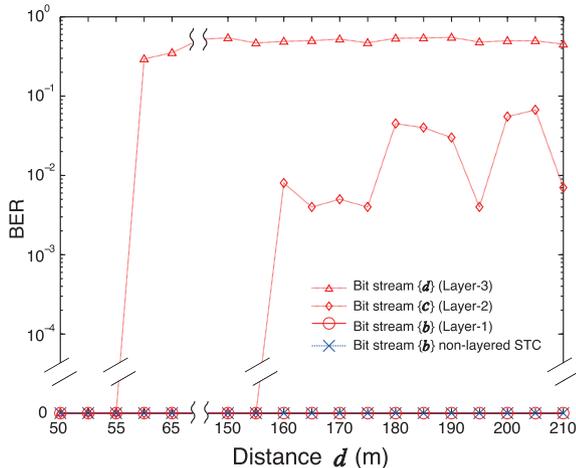


Fig. 7 Two raw images with size  $6 \times 6$  trimmed from the whole one and its  $8 \times 8$  luminance matrices  $V^j$  and  $V^{j+1}$  after resizing image processing are shown, when an  $8 \times 8$  LED array was captured by the high-speed camera at the distance of 55 m. Reproduced from [24].

luminance values, though the detail is not mentioned here (see [24]). It was confirmed from Fig. 8 that the received bit stream of  $\{b\}$  is error-free all over the measured range, and the received bit streams of  $\{c\}$  and  $\{d\}$  are also error-free in the range  $D \leq 155$  m and 55 m, respectively. This is a bit surprise that the sizes of the raw image of the LED array at 210 m, 155 m, and 55 m are  $2 \times 2$ ,  $2 \times 2$ , and  $5 \times 5$  pixels, respectively. This can be said a good example of exploiting spatial dimensions for extending the coverage of IS-based VLC.

It is shown in [43], [44] that applying spatial modulation (SM) into the L-STC allows to further increase in the reception rate without deteriorating the bit reception quality of the original L-STC.



**Fig. 8** Measured BER of the three-layered and non-layered STC systems is shown as a function of the distance  $d$  of the receiver from the LED array, where the lens with the focal length 12.5 mm are used. Reproduced from [24].

## 5. Conclusion

This paper presented a survey on the novel type of VLC using IS receivers. Current VLC receivers either use a PD or an IS for receiving the VLC signals. The use of PD is suitable for stationary clients where its FOV can be aligned to the LED fixture for high received optical power. On the mobile devices or moving situations, the IS can be used since they have comparatively larger FOV (due to wider concentration lens), making the moving device more robust to movements and FOV misalignment. Due to a large number of PDs operating inside IS might be slow and energy expensive. This is natural, given that IS was primarily designed for image and video capture, and not for receivers. Thus, it is challenging to design an array transmitter and IS receiver pair to satisfy a diverse range of practical requirements, with robustness to device movements and FOV misalignment, making the most use of the 2D systems, for meritorious transmission techniques.

## Acknowledgment

The author would like to thank Prof. Takaya Yamazato of Nagoya University for his insightful comments on an earlier version of this paper.

## References

- [1] M. Akanegawa, Y. Tanaka, and M. Nakagawa, "Basic study on traffic information system using LED traffic lights," *IEEE Trans. Intell. Transp. Syst.*, vol.2, no.4, pp.197–203, Dec. 2001.
- [2] H. Haas, L. Yin, Y. Wang, and C. Chen, "What is LiFi?," *J. Lightw. Technol.*, vol.34, no.6, pp.1533–1544, Mar. 2016.
- [3] P.H. Pathak, X. Feng, P. Hu, and P. Mohapatra, "Visible light communication, networking, and sensing: A survey, potential and challenges," *IEEE Commun. Surveys Tuts.*, vol.17, no.4, pp.2047–2077, 4th Quater 2015.

- [4] L. Zeng, D.C. O'Brien, H.L. Minh, G.E. Faulkner, K. Lee, D. Jung, Y. Oh, and E. T. Won, "High data rate multiple input multiple output (MIMO) optical wireless communications using white LED lighting," *IEEE J. Sel. Areas Commun.*, vol.27, no.9, pp.1654–1662, Dec. 2009.
- [5] Y. Wang, L. Tao, X. Huang, J. Shi, and N. Chi, "8-Gb/s RGBY LED-based WDM VLC system employing high-order CAP modulation and hybrid post equalizer," *IEEE Photon. J.*, vol.7, no.6, 7904507, Dec. 2015.
- [6] H.B.C. Wook, S. Haruyama, and M. Nakagawa, "Visible light communication with LED traffic lights using 2-dimensional image sensor," *IEICE Trans. Fundamentals*, vol.E89-A, no.3, pp.654–659, March 2006.
- [7] M. Rubaiyat, H. Mondal, and J. Armstrong, "Analysis of the effect of vignetting on MIMO optical wireless systems using spatial OFDM," *J. Lightw. Technol.*, vol.32, no.5, pp.922–929, May 2014.
- [8] A. Dabbo and S. Hranilovic, "Receiver design for wireless optical MIMO channels with magnification," *Proc. ConTEL 2009*, pp.51–58, Zagreb, Croatia, June 2009.
- [9] C. Danakis, M. Afgani, G. Povey, I. Underwood, and H. Haas, "Using a CMOS camera sensor for visible light communication," *Proc. IEEE Globecom Workshops 2012*, pp.1244–1248, Anaheim, CA, USA, Dec. 2012.
- [10] Y. Goto, I. Takai, T. Yamazato, H. Okada, T. Fujii, S. Kawahito, S. Arai, T. Yendo, and K. Kamakura, "A new automotive VLC system using optical communication image sensor," *IEEE Photon. J.*, vol.8, no.3, 6802716, June 2016.
- [11] K. Masuda, T. Yamazato, H. Okada, M. Katayama, and T. Fujii, "A proposal on hierarchical coding scheme of visible light communication using led traffic lights," *Proc. ISITA 2004*, pp.187–190, Parma, Italy, Oct. 2004.
- [12] T. Yamazato, I. Takai, H. Okada, T. Fujii, T. Yendo, S. Arai, M. Andoh, T. Harada, K. Yasutomi, K. Kagawa, and S. Kawahito, "Image-sensor-based visible light communication for automotive applications," *IEEE Commun. Mag.*, vol.52, no.7, pp.88–97, July 2014.
- [13] I. Takai, S. Ito, K. Yasutomi, K. Kagawa, M. Andoh, and S. Kawahito, "LED and CMOS image sensor based optical wireless communication system for automotive applications," *IEEE Photon. J.*, vol.5, no.5, 6801418, Oct. 2013.
- [14] R.D. Roberts, "A MIMO protocol for camera communications (CamCom) using undersampled frequency shift on-off keying (UF-SOOK)," *Proc. GLOBECOM Wkshps*, pp.1052–1057, Atlanta, GA, USA, Dec. 2013.
- [15] W. Du, J.C. Liando, and M. Li, "Soft hint enabled adaptive visible light communication over screen-camera links," *IEEE Trans. Mobile Comput.*, vol.16, no.2, pp.527–537, Feb. 2017.
- [16] B. Xie, K. Chen, G. Tan, M. Lu, Y. Liu, T. He, and J. Wu, "LIPS: A light intensity based positioning system for indoor environments," *ACM Trans. Sensor Networks*, vol.12, no.4, pp.1–27, 2016.
- [17] T. Nguyen, C.H. Hong, N.T. Le, and Y.M. Jang, "High-speed asynchronous optical camera communication using LED and rolling shutter camera," *Proc. IEEE ICUFN 2015*, pp.214–219, Sapporo, Japan, July 2015.
- [18] VLCC [online] [http://www.vlcc.net/modules/xpage1/?ml\\_lang=en](http://www.vlcc.net/modules/xpage1/?ml_lang=en)
- [19] 802.15.7-2011-IEEE standard for local and metropolitan area networks-Part 15.7: Short-range wireless optical communication using visible light.
- [20] S. Haruyama and T. Yamazato, "Image sensor based visible light communication," in *Visible Light Communication*, S. Arnon Ed., ch.9, pp.181–205, Cambridge University Press, Cambridge, UK, 2015.
- [21] T. Yamazato and S. Haruyama, "Image sensor based visible light communication and its application to pose, position, and range estimations," *IEICE Trans. Commun.*, vol. E97-B, no.9, pp.1759–1765, Sept. 2014.
- [22] S. Arnon, *Advanced Optical Wireless Communication Systems*,

- Cambridge University Press, Cambridge, UK, 2012.
- [23] Z. Ghassemlooy, W. Popoola, and S. Rajbhandari, *Optical wireless communications: System and channel modelling with MATLAB*, CRC Press, Boca Raton, FL, 2012.
- [24] K. Ebihara, K. Kamakura, and T. Yamazato, "Layered transmission of space-time coded signals for image-sensor-based visible light communications," *J. Lightw. Technol.*, vol.33, no.20, pp.4193–4206, Oct. 2015.
- [25] A.E. Gamal and H. Eltouky, "CMOS image sensors," *IEEE Circuits Devices Mag.*, vol.21, no.3, pp.6–20, May/June 2005.
- [26] L.J. Kozlowski, D.L. Standley, J. Luo, A. Tomasini, A.M. Gallagher, R. Mann, B.C. Hsieh, T. Liu, and W.E. Kleinhans, "Theoretical basis and experimental confirmation: Why a CMOS imager is superior to a CCD," *Proc. SPIE 3698, Infrared Technology and Applications XXV*, pp.388–396, July 1999.
- [27] J. Janesick and J. Tower, "Particle and photon detection: Counting and energy measurement," *Sensors*, vol.16, no.5, 688, pp.1–17, May 2016.
- [28] P. Ji, H.-M. Tsai, C. Wang, and F. Liu, "Optical vehicle-to-vehicle communication system using LED transmitter and camera receiver," *Proc. IEEE VTC Spring 2014*, pp.1–6, Seoul, Korea, May 2014.
- [29] Y.-S. Kuo, P. Pannuto, K.-J. Hsiao, and P. Dutta, "Luxapose: Indoor positioning with mobile phones and visible light," *Proc. MobiCom 2014*, pp.447–458, Maui, HI, USA, Sept. 2014.
- [30] Y. Liu, K. Liang, H.-Y. Chen, L.-Y. Wei, C.-W. Hsu, C.-W. Chow, and C.-H. Yeh, "Light encryption scheme using light-emitting diode and camera image sensor," *IEEE Photon. J.*, vol.8, no.1, 7801107, Feb. 2016.
- [31] E.F. Schubert, *Light-Emitting Diodes*, 2nd Ed., ch.5, pp.93–97, Cambridge University Press, Cambridge, UK, 2006.
- [32] S. Rajagopal, R.D. Roberts, and S.-K. Lim, "IEEE 802.15.7 visible light communication: modulation schemes and dimming support," *IEEE Commun. Mag.*, vol.50, no.3, pp.72–82, March 2012.
- [33] *IESNA Lighting Handbook*, 9th Edition, ch.27, Figure 27-4, IESNA, New York, 2000.
- [34] G. Ntogari, T. Kamalakis, J.W. Walewski, and T. Sphicopoulos, "Combining illumination dimming based on pulse-width modulation with visible-light communications based on discrete multitone," *J. Opt. Commun. Netw.*, vol.3, no.1, pp.56–65, Jan. 2011.
- [35] H.-J. Jang, J.-H. Choi, Z. Ghassemlooy, and C.G. Lee, "PWM-based PPM format for dimming control in visible light communication system," *Proc. CSNDSP 2012*, pp.1–5, Poznan, Poland, July 2012.
- [36] J.K. Kwon, "Inverse source coding for dimming in visible light communications using NRZ-OOK on reliable links," *IEEE Photon. Technol. Lett.*, vol.22, no.19, pp.1455–1457, Oct. 2010.
- [37] S.D. Perli, N. Ahmed, and D. Katabi, "PixNet: Interference-free wireless links using LCD-camera pairs," *Proc. ACM MobiCom 2010*, pp.137–148, Chicago, IL, USA, Sept. 2010.
- [38] M. Kinoshita, T. Yamazato, H. Okada, S. Arai, E. Souke, T. Yendo, T. Fujii, and K. Kamakura, "Motion modeling of mobile transmitter for image sensor based I2V-VLC, V2I-VLC, and V2V-VLC," *Proc. GLOBECOM Wkshps 2014*, pp.450–455, Austin, TX, USA, Dec. 2014.
- [39] K. Ebihara, K. Kamakura, and T. Yamazato, "Layered space-time coding using LED array for image-sensor-based visible light communications," *Proc. IEEE ICC 2014*, pp.5048–5053, London, UK, June 2015.
- [40] T. Nguyen, N.T. Le, and Y.M. Jang, "Practical design of screen-to-camera based optical camera communication," *Proc. IEEE ICOIN 2015*, pp.369–374, Siem Reap, Cambodia, Jan. 2015.
- [41] S. Hranilovic and F.R. Kschischang, "A pixelated MIMO wireless optical communication system," *IEEE J. Sel. Topics Quantum Electron.*, vol.12, no.4, pp.859–874, July/Aug. 2006.
- [42] Y. Amano, K. Kamakura, and T. Yamazato, "Alamouti-type coding for visible light communication based on direct detection using image sensor," *Proc. IEEE GLOBECOM 2013*, pp.2430–2435, Atlanta, GA, USA, Dec. 2013.
- [43] K. Masuda, K. Kamakura, and T. Yamazato, "Spatial modulation in layered space-time coding for image-sensor-based visible light communication," *Proc. PIMRC 2016*, pp.1–6, Valencia, Spain, Sep. 2016.
- [44] K. Masuda and K. Kamakura, "Spatial modulation for layered space-time coding used in image-sensor-based visible light communication," *IEICE Trans. Commun., Special Issue on Visible Light Communications in Conjunction with Topics of ICEVLC 2015*, Article ID: 2016LCP0001, Dec. 2016.



**Koji Kamakura** received the B.E., M.E., and Ph.D. degrees in electrical engineering from Keio University, Yokohama, Japan, in 1997, 1999, and 2002, respectively. He is a Professor at Department of Computer Science, Chiba Institute of Technology, Chiba, Japan. From 2002 to 2006, he was an Assistant Professor at the Department of Electronics and Mechanical Engineering, Chiba University, Chiba, Japan. From 2006 to 2015, he was an Associate Professor with the Department of Computer Science, Chiba Institute of Technology, Chiba, Japan. He was a Visiting Professor at Heudiasyc, Université de Technologie de Compiègne, France, from April 2013 to March 2014. He was a Visiting Scientist at the School of Information Technology and Engineering, University of Ottawa, Ottawa, ON, Canada, in 2002 and 2003. From 2000 to 2002, he was a Special Researcher of Fellowships of the Japan Society for the Promotion for Science, for Japanese Junior Scientists. His research interests include optical communication theory and system analysis. He is a Member of the IEICE. He received the 14th Telecom System Technology Award for Students from the Telecommunications Advancement Foundation in 1999 and the Ericsson Young Scientist Award in 2002.