Non-Line-of-Sight MIMO Space-Time Division Multiplexing Visible Light Optical Camera Communications

Navid Bani Hassan, Zabih Ghassemlooy, *Senior Member, IEEE,* Stanislav Zvanovec, *Senior Member, IEEE,* Mauro Biagi, *Senior Member, IEEE,* Anna Maria Vegni, *Senior Member, IEEE,* Min Zhang, and Pengfei Luo

***Abstract*—In this paper, a non-line-of-sight multiple-input multiple-output space and time division multiple access optical camera communications system is proposed for an indoor environment. Mask matching and equal-gain combining (EGC) schemes as well as differential modulation and frame subtraction are used. We propose a unique packet structure to label the transmitters and a new detection method for data extraction from the captured video streams. We outline a comprehensive theoretical model and have developed an experimental testbed to evaluate the performance of the proposed system. The results highlight that zooming and defocusing of the camera does not have a significant impact on the system performance, therefore the aperture can be set to its maximum value. The system performs well over a link span of 10 m with a low transmit power of 12 mW and in the presence of ambient light due to the non-linear conversion of RAW to JPEG. Using mask matching and EGC improves the tolerance of the system to the noise.**

***Index Terms*—non-line-of-sight; camera communications; MIMO, space division multiplexing; equal-gain combining.**

I. Introduction

APPLICATIONS of solid-state light emitting diodes (LEDs) for illuminations, indoor data communications and localization, as well as sensing have been growing over in the last few years [1], [2]. Visible light communications (VLC) systems, which uses the LED-based lights installed in indoor and outdoor environments to provide wireless services, offers a potential licence-free bandwidth B orders of magnitude higher than the radio frequency technologies [2]. In VLC there are two types of detectors are used (i) photodiodes (PD) with a wide B range (a few MHz to beyond GHz depending on the size) [1]; and (ii) image sensors (ISs) (a multi-array PD) with much lower B [1]–[5]. Short range PD-based VLC systems offer higher data rates $R_{\rm b}$ but at the cost of reduced mobility and increased level of shadowing and blocking in indoor environments [2]. However, both mobility and shadowing have been addressed using multi-array transmitter (Tx) and receivers (Rxs) and/or non-line-of-sight (NLOS) VLC links, but at much reduced $R_{\rm b}$ due to multipath (see [2], [6] and the references within). Note that, in an outdoor environment a single PD-based VLC system will suffer from a high level ambient light (i.e., sunlight) induced noise, thus resulting in significant deterioration of the link performance. Unlike PDs, the complementary metal-oxide semiconductor (CMOS) based cameras offer both vision and spatial separation of multiple Txs (i.e., sunlight, street light, vehicle lights etc.), which is highly useful in VLC [2]–[5]. Such systems also offer imaging multiple-input multiple-output (MIMO) capability [3], [7] for use in indoor [8], [9] and outdoor [10] applications including intelligent transportation systems (ITS), indoor localizations, and device-to-device communications [11], [12]. In ITS, the MIMO feature of cameras can be exploited for collision warning, adaptive cruise control, enhancing driving safety, people and object detection, range estimations and data communications in vehicle-to-vehicle and vehicle-to-road side infrastructure communications [10], [13]–[15]. There are a number of challenges in IS-based VLC, also known as optical camera communications (OCC), as outlined below:

1) Transmission links: are usually considered as line-of-sight (LOS). However, in some scenarios the Txs may not be within the field of view of the LOS path such as (*i*) two vehicles approaching a road junction from different directions; (*ii*) vehicles travelling in the same direction; and (*iii*) device-to-device communications, where smart devices are not pointing directly towards the Txs or LOS paths are partially or fully blocked, therefore communications via NLOS paths [5], [16]. In NLOS links, the reflected optical beams normally have a large illumination footprint, thus offering an increased level of mobility and link tolerance to the cameras movements.

2) Detection and tracking of Txs: given that Txs are part of the image a number of solutions have been proposed such as frame subtraction in LOS [11], [17] and NLOS [18], block matching [12], and vision-based image processing [19], [20]. However in image processing schemes the processing time is considerably high, and the offered $R_{\rm b}$ is very low (i.e., kbps) due to the low camera frame rate $R_{\rm f}$. One solution is to use the rolling shutter (RS) effect, which enables $R_{\rm b} > R_{\rm f}$ [21]. In [5] and [22] RS cameras were employed in a NLOS OCC

N. Bani Hassan and Z. Ghassemlooy are with the Dep. of Physics, Mathematics, and Electrical Engineering, Northumbria University, Newcastle upon Tyne, e-mail: {navid.hassan, z.ghassemlooy}@northumbria.ac.uk.

S. Zvanovec is with the Dep. of Electromagnetic Field, Czech Technical University in Prague, Technicka 2, 16627 Prague, Czech Republic, e-mail: xzvanove@fel.cvut.cz.

M. Biagi is with the Dep. of DIET engineering, Sapienza University of Rome, Via Eudossiana 18, Rome, Italy. e-mail: mauro.biagi@uniroma1.it.

A. M Vegni is with the Department of Engineering, Roma Tre University, Rome, Italy. e-mail: annamaria.vegni@uniroma3.it.

M. Zhang is with the State Key Laboratory of Information Photonics and Optical Communications, Beijing Univ. of Posts & Telecom, Beijing, China, e-mail: mzhang@bupt.edu.cn.

P. Luo is with Research Department of HiSilicon, Huawei Technologies Co., Ltd, Beijing 100085, P. R. China, e-mail: oliver.luo@hisilicon.com.

Manuscript received July 05, 2018; revised Month DD, YYYY.

to detect the fast changing intensity of reflected lights in an indoor environment. However, these systems have a number of problems including (*i*) non-stationary camera; (*ii*) a complex detection mechanism and the need for synchronization between the camera and the Tx; and (*iii*) limited $L_{\rm s}$ [22]. Note, with the CMOS-based IS technology, the quality of captured videos using digital cameras has improved significantly. Cameras in new smartphones can record video streams up $R_{\rm f}$ of 960 frames per second (fps) at a resolution of 720p [23]. In addition, there are commercially available high speed cameras with $R_{\rm f}$ of 25.7 kfps and 1 Mfps at resolutions of 1280×800 and 128×32, respectively [24]. In [10] a CMOS IS with a data capture rate of 20-Mb/s/p was reported. Note that, VLC links with low $R_{\rm b}$ can be effectively used in a number of applications such as ITS, indoor localization, sensing, etc., where high $R_{\rm b}$ is not a requirement [2].

3) Flickering: is due to low-frequency changes in the intensity of the Tx. To address this problem high-speed cameras or the RS effect can be used to increase $R_{\rm b}$. In [25] and [26] an undersample frequency and phase shift on-off keying modulation technique were investigated for OCC with low-speed cameras in order to increase the chip rate.

4) Multiple access and multiplexing: where signals from independent sources are combined at the Rx. In [20] a space division multiplexed system was demonstrated for an OCC LOS link. In [27] color-shift keying and code-division multiple-access schemes were proposed for OCC, which employed a complex Tx and Rx modules.

In this paper, an indoor NLOS MIMO space-time division multiplexed (STDM) OCC system is proposed. By employing a combination of successive frame subtraction, mask matching, equal-gain combining (EGC), and differential modulation schemes for detection and transmission, respectively, the proposed system can recover from the captured images the transmitted data via both LOS and NLOS paths. The differential modulation and the frame subtraction are used at the Tx and the Rx, respectively to remove the unwanted background objects within frames, whereas the mask matching allows to determine the optical footprint of the desired Tx in a given frame. A MIMO-based space division multiplexing along with time division multiplexing (TDM) are employed to spatially separate the Txs and recover the data successfully even with the overlap of optical footprints. To recover the data from the diversity Rxs as well as to improve the signal-to-noise ratio (SNR) we use EGC. We give a detailed mathematical analysis of the received optical power considering the crosstalk. The proposed OCC system is experimentally implemented and evaluated under different channel conditions. We show that for $L_{\rm s}$ of 10 m very low transmit power $P_{\rm t}$ levels of 12 mW and 2 mW with and without the ambient light, respectively, are required to achieve a bit error rate (BER) of 10^{-3}, which is just below the forward error correction limit of 3.8×10^{-3}. We also show that, (*i*) there is no need for the Txs to be placed in focus, hence larger aperture sizes can be used; and (*ii*) using JPEG images instead of RAW images, the system performance improves under the low-to-medium ambient light. Moreover, we show that in the MIMO system beyond a certain value of the transmit power the BER is flat, since the interference is the dominant source.

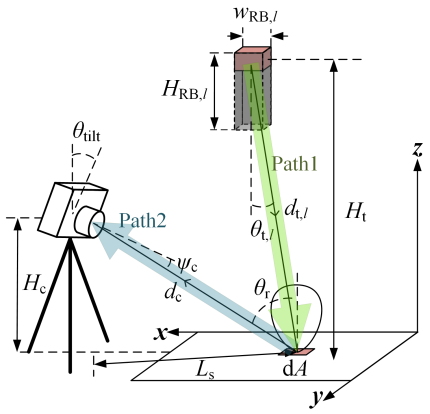

Fig. 1. Configuration of the proposed system.

The rest of the paper is organized as follows: In Section II, the system model is presented, and in Section III the proposed detection algorithm is given. Section IV outlines experimental test setup and results. Section IV concludes the paper.

II. Theory

Assuming a linear shift-invariant channel and considering a single reflection, the received optical signal is given by:

$$\boldsymbol{y}_l(t) = P_{\mathrm{t},l}\eta g_{\mathrm{lens}} x_{\mathrm{op},l}(t) \otimes \boldsymbol{h}_l(t) + \boldsymbol{n}(t) + \boldsymbol{I}(t), \qquad (1)$$

where $P_{\mathrm{t},l}$ is the transmit power of the l-th Tx located at $(x_{\mathrm{t},l}, y_{\mathrm{t},l}, H_{\mathrm{t},l})$, η is the quantum efficiency of the IS array with a size of $U \times V$, where U and V are the number of rows and columns in the image, respectively, and g_{lens} presents the gain of the camera lenses. $x_{\mathrm{op},l}(t)$ is the transmitted optical signal of the l-th Tx, which is a scalar number, and $\boldsymbol{n}(t)$ is the noise, which is modeled as a Gaussian distributed random variable (GDRV) with the mean of μ_{N} and variance of σ_{N}^2 [28]. $\boldsymbol{I}(t)$ is the interference from other Txs. Note, $\boldsymbol{h}_l(t) = \boldsymbol{h}_{\mathrm{ch},l}(t) \otimes \boldsymbol{h}_{\mathrm{cp}}(t)$, which is a matrix of size $U \times V$, where $h_{\mathrm{ch},l}(t)$ is the impulse response of the channel, and $\boldsymbol{h}_{\mathrm{cp}}(t)$ is the impulse response of the CMOS with an active pixel matrix given by [29]:

$$\boldsymbol{h}_{\mathrm{cp}}(\omega) = \boldsymbol{G} e^{-j\omega T_{\mathrm{exp}}/2} \frac{2\sin(0.5\omega T_{\mathrm{exp}})}{\omega}, \qquad (2)$$

where $\boldsymbol{G}$ is the gain matrix of the preamplifiers and ω is the angular frequency. Here, $\otimes$ denotes the convolution operation, which acts element-wise on the matrix.

In the proposed NLOS model, see Fig. 1, the trajectory of light beams to the camera located at $(x_{\mathrm{c}}, y_{\mathrm{c}}, H_{\mathrm{c}})$ can be divided into two LOS paths i.e., Path1 and Path2 that exist between the Tx and the floor, and between the element area $\mathrm{d}A = \mathrm{d}x\mathrm{d}y$ on the floor and the camera, respectively. Note, the light intensity in $\mathrm{d}A$ is assumed to be constant. In both paths light transmission is via LOS with the attenuation coefficients for Paths1&2 are given as [30]:

$$\alpha_{\text{t-f},l}(x,y) = \frac{\mathrm{d}A}{d_{\mathrm{t},l(x,y)^2}}, \qquad (3)$$

$$\alpha_{\text{f-c}}(x,y) = \frac{A_{\text{lens,eff}}}{d_{\text{c}}^2(x,y)} \times \rho, \tag{4}$$

where $A_{\text{lens,eff}} = \gamma \frac{f}{f_{\text{stop}}}$ and f are effective area of the lens and the focal length of the lens, respectively, f_{stop} is the f-stop number for the aperture, γ is the factor that relates the aperture area to $A_{\text{lens,eff}}$, and ρ is the reflection coefficient of the surface. $d_{\text{t},l}(x,y)$ and $d_{\text{c}}(x,y)$ are the distances between the l-th Tx and the $\mathrm{d}A(x,y,0)$, and between the camera and $\mathrm{d}A$, respectively defined as [30], [31]:

$$d_{\text{t},l}(x,y) = \sqrt{(x-x_{\text{t},l})^2 + (y-y_{\text{t},l})^2 + H_{\text{t},l}^2}. \tag{5}$$

$$d_{\text{c}}(x,y) = \sqrt{(x-x_{\text{c}})^2 + (y-y_{\text{c}})^2 + H_{\text{c}}^2}. \tag{6}$$

The overall received power from the l-th Tx at the pixel (i,j) of the camera is given by:

$$\begin{aligned} h_{\text{ch},l}(u,v) = \int_{x_{u-1}}^{x_u} \int_{y_{v-1}(x)}^{y_v(x)} & R_{\text{t},l}(x,y)\alpha_{\text{t-f},l}(x,y) \\ & \times R_{\text{r}}(x,y)\alpha_{\text{f-c}}(x,y) \\ & \times \cos(\psi_{\text{c}}(x,y)). \end{aligned} \tag{7}$$

$R_{\text{t},l}(x,y)$ and $R_{\text{r}}(x,y)$ are LED beam profile and Lambertian radiation pattern of $\mathrm{d}A$, respectively, which is given by [30]:

$$R_{\text{t},l}(x,y) = \frac{m_{\text{t},l}+1}{2\pi} \cos^{m_{\text{t},l}}\left(\theta_{\text{t},l}(x,y)\right). \tag{8}$$

$$R_{\text{r}}(x,y) = \frac{m_{\text{r}}+1}{2\pi} \cos^{m_{\text{r}}}\left(\theta_{\text{r}}(x,y)\right). \tag{9}$$

where $m_{\text{t},l}$ and m_{r} are Lambertian order for the l-th LED and the reflected beam profile, respectively. $\theta_{\text{t},l}(x,y)$ and $\theta_{\text{r}}(x,y)$ are emission angles of the Tx and the reflection to the Rx, respectively given as:

$$\theta_{\text{t},l}(x,y) = \tan^{-1}\left(\frac{\sqrt{(x-x_{\text{t},l})^2 + (y-y_{\text{t},l})^2}}{H_{\text{t},l}}\right). \tag{10}$$

$$\theta_{\text{r}}(x,y) = \tan^{-1}\left(\frac{\sqrt{(x-x_{\text{c}})^2 + (y-y_{\text{c}})^2}}{H_{\text{c}}}\right). \tag{11}$$

In (9) the incident angle of the camera $\psi_{\text{c}}(x,y) = \cos^{-1}(\overrightarrow{\boldsymbol{n}_{\text{I}}}.\overrightarrow{\boldsymbol{n}_{\text{c}}})$ where $\overrightarrow{\boldsymbol{n}_{\text{I}}} = (x-x_{\text{c}}, y-y_{\text{c}}, -H_{\text{c}})$ and $\overrightarrow{\boldsymbol{n}_{\text{c}}} = (-\cos(\phi_{\text{tilt}})\cos(\theta_{\text{tilt}}), \sin(\phi_{\text{tilt}})\cos(\theta_{\text{tilt}}), -\sin(\theta_{\text{tilt}}))$ denote the unitized incident beam and the camera plain normal vector, respectively and the dot denotes the inner product. Here, ϕ_{tilt} and θ_{tilt} are azimuth and elevation tilting angles of the camera, respectively, see Fig. 1. In (9), x_u and $y_v(x)$ are the boundaries of the area that pixel (u,v) cover in x and y dimensions on the floor surface, respectively given as:

$$x_u = x_{\text{c}} - H_{\text{c}} \times \frac{f\cos(\theta_{\text{tilt}}) + (u-U/2)a\sin(\theta_{\text{tilt}})}{f\sin(\theta_{\text{tilt}}) - (u-U/2)a\cos(\theta_{\text{tilt}})}, \tag{12}$$

$$y_v(x) = \frac{v(\beta-\alpha)}{Vx_U}(x-x_0) + \frac{v\alpha}{V}, \quad -\frac{V}{2} \le v \le \frac{V}{2}, \tag{13}$$

where $\alpha = a\frac{VH_{\text{c}}}{g\sin(\theta_{\text{tilt}}+\theta_{\text{FoV}}/2)}$, $\beta = \frac{\sin(\theta_{\text{FoV}}/2-\theta_{\text{tilt}})}{\sin(\theta_{\text{FoV}}/2+\theta_{\text{tilt}})}\alpha$, $\theta_{\text{FoV}} = 2\tan^{-1}\left(\frac{Ua}{2f}\right)$, and $g = \sqrt{(Ua/2)^2 + f^2}$. Further details of the analysis are given in Appendices A and B.

In this work, the main noise sources considered are the shot noise, reset noise, readout noise, dark current shot noise, fixed pattern noise, and quantization error noise [32], [33], which are modeled as an additive GDRV [28]. Note that, in IS-based Rxs the dark current, readout, and quantization noise sources are negiligable and therefore are ignored and only the shot and reset noise sources are considered. Hence, $\sigma_N^2 = \sigma_{\text{res}}^2 + \sigma_{\text{shot}}^2$. Here, we consider the non-return to zero on and off keying signal format, where the BER is defined as:

$$\text{BER} = Q\left(\frac{\mu_1 - \mu_0}{\sigma_1 + \sigma_0}\right), \tag{14}$$

where $Q(.)$ is Gaussiang Q-function, μ_1 and μ_0 are the mean of bit "1" and "0", respectively, and $\sigma_1 = \sqrt{\sigma_{\text{res}}^2 + \sigma_{\text{shot}}^2}$ and $\sigma_0 = \sigma_{\text{res}}$ are the variance of bit "1" and "0", respectively. Given that, lights from other Txs can results in crosstalk due to overlapping optical footprints, which will effect the link performance depending on the size of the overlapping of area. The overall received signal at the IS with crosstalk is given by:

$$\begin{aligned} I(u,v) = \int_{x_{u-1}}^{x_u} \int_{y_{v-1}(x)}^{y_v(x)} & \sum_{k=1,k\neq l}^{K} P_{\text{t},k} x_{\text{op},k} R_{\text{t},l}(x,y) \\ & \times \alpha_{\text{t-f},l}(x,y) R_{\text{r}}(x,y) \alpha_{\text{f-c}}(x,y) \cos(\psi_{\text{c}}(x,y)). \end{aligned} \tag{15}$$

III. System Model

The block diagram of the proposed system is shown in Fig. 2(a). In [34], it was shown that if the overlapping area of optical footprints is $> 30\%$ of the area of each optical footprint the full recovery of the signal using a camera was not possible. Therefore, here we consider a TDM-based scheme where the Txs are divided into groups with the overlapping area $< 30\%$ of the each footprint area.

1) Packet Generator: To mark the beginning of the packet a *preamble* sequence N_{p} is added to M independent pseudo-random sequences of data (i.e., *payload*) d_l of length N_{d} where $l = 1, \ldots, M$, using the packet generator block, see Fig. 2(b). Note, N_{p} length should be selected such that both the overhead and the likelihood of appearing in the *payload* which can be obtained as:

$$P_X(P \in D) = \sum_{i=1}^{\lfloor N_{\text{d}}/N_{\text{p}} \rfloor} \sum_{j=1}^{N_{\text{d}}-iN_{\text{p}}} \frac{(N_{\text{d}} - i\,(N_{\text{p}}-1))!}{(N_{\text{d}} - iN_{\text{p}} - j)!i!j!}, \tag{16}$$

are reduced. For a payload of length 1000 bits, a 64-bit *preamble* will give $P_X(P \in D) = 5 \times 10^{-17}$. The field *group* and *mask* are used for the Txs's group ID, and their positions, respectively as well as representing the Tx's ID, see Fig. 2(b).

2) Differential Signaling and Interleaver: The output of packet generator $b_{k,l}$ is differentially encoded with $x_{k,l} = x_{k-1,l} \oplus b_{k,l}$, where $\oplus$ denotes the modulo-2 addition, $x_{k,l}$ is the output of differential signaling block, and $x_{0,l} = 0$. To reduce the flickering effect, a bit by bit-based interleaver is used, in which the i-th bit of g-th group of Txs is followed by the i-th bit of the $(g+1)$-th group of Txs, and the $(i+1)$-th bit of the g-th group of Txs follows the i-th bit of the $(g-1)$-th group of Tx, see Fig. 2(b). Accordingly, for the Txs $R_{\text{b}} = (KR_{\text{f}})/N_{\text{g}}$, where K is the number of symbols per

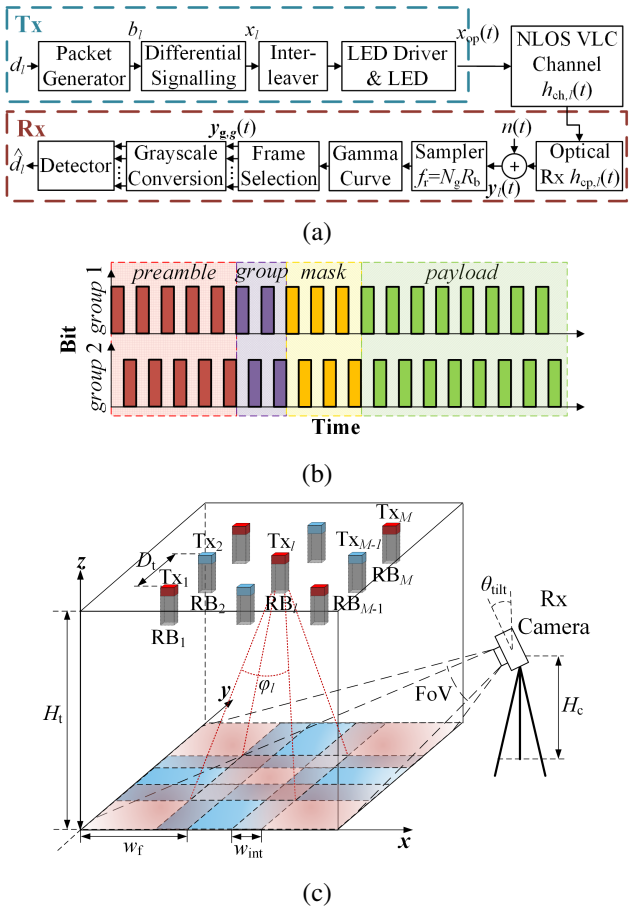

Fig. 2. Proposed OCC system: (a) schematic block diagram, (b) proposed packet structure, and (c) system orientation with Txs optical footprints with two groups of Txs. The camera is used to capture reflected lights.

frame (for OOK-NRZ $K = 1$) and N_g is the number of Txs's group. Note, we assumed that R_f is fixed, but for cameras with variable R_f the condition of $R_b < \min(R_f)$ must be met. However, the additional frames captured needs to be removed by means of thresholding.

3) Light Source: Array LEDs located within rectangular boxes (RBs), see Fig. 2(c), are used to illuminate the floor with the required rectangular optical footprints, which is as:

$$w_{f,l} = \frac{w_{RB,l}}{H_{RB,l}} H_t, \tag{17}$$

where H_t is the height of the room, $w_{RB,l}$ and $H_{RB,l}$ are the width and the height of the l-th RB, respectively. Note, for the l-th Tx $(x_{t,l}, y_{t,l}, H_t)$ the viewing angle $\phi_l = 2 \times \tan^{-1}\left(\frac{w_{RB,l}}{2H_{RB,l}}\right)$.

4) NLOS VLC Channel and Sampler: At the Rx the captured video streams of the reflected lights from the floor surface are processed frame by frame in Matlab in order to recover d_l. The camera parameters such as ISO, aperture, and exposure time T_{exp} play major roles in improving signal detection particularly in environments with background lights. Increasing ISO, changes the gain of the camera, and increases the brightness level and the perceptible noise of the captured image. The aperture, which controls the amount of light passing through the camera lens, also affects blurriness of out-

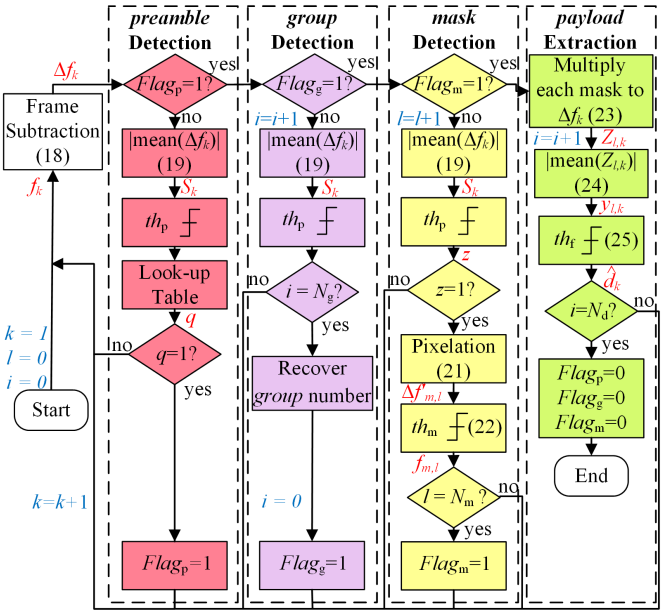

Fig. 3. Flowchart of the proposed detection scheme composed of four stages: preamble, group, mask, and payload.

of-focus objects (i.e., also known as Bokeh). T_{exp}, also known as the shutter speed, determines how long the IS is exposed to the incoming light. The slower the shutter speed, the brighter the image is, but at the cost of motion blur. Note, in order to enhance the visual aspects of the image, the camera applies the gamma correction curve to the image, i.e., the image intensity is not linearly proportional to the received light intensity.

5) Grayscale Conversion: In offline processing, following selection of frames, the received frame first is converted to the gray scale (in line with the recommendation BT.601-7 [35]) to generate a $U \times V$ matrix, which is given as:

$$\begin{aligned} f_k(u,v) = &0.2989 \times f_{k,\mathrm{RGB}}(u,v,\mathrm{R}) + \\ &0.5870 \times f_{k,\mathrm{RGB}}(u,v,\mathrm{G}) + \\ &0.1140 \times f_{k,\mathrm{RGB}}(u,v,\mathrm{B}), \end{aligned} \tag{18}$$

where $f_{k,\mathrm{RGB}}(u,v,\mathrm{R})$, $f_{k,\mathrm{RGB}}(u,v,\mathrm{G})$, and $f_{k,\mathrm{RGB}}(u,v,\mathrm{B})$ denote the red, green, and blue components of the (u,v)-th element in the image matrix, respectively.

IV. Detector

1) Frame Subtraction: Following the selection of odd and even frames, frame subtraction is first carried out for all frames per group, and then for frames of other groups for the entire video frame (i.e., $U \times V$) in order to eliminate the background objects in the image, as given by:

$$\boldsymbol{\Delta f}_{k,g} = \boldsymbol{f}_{k,g} - \boldsymbol{f}_{k-1,g} \tag{19}$$

where $\boldsymbol{f}_{k,g}$ is the matrix of the k-th frame of the g-th group.

The operation of the detection algorithm for the proposed system is best described by the flow chart shown in Fig. 3. The algorithm works in four stages i.e., (*i*) *preamble* detection; (*ii*) *group* detection; (*iii*) *mask* detection; and (*iv*) *payload* extraction, where the first three stages are controlled using

three flags of $Flag_{\mathrm{p}}$, $Flag_{\mathrm{g}}$, and $Flag_{\mathrm{m}}$, respectively. The initial value for all flags is set to zero.

2) EGC: In the first stage of detection, we take the average over the entire matrix of $\boldsymbol{\Delta f}_{k,g}$ as given by:

$$S_{k,g} = \left| \underset{i,j}{\mathrm{mean}} \left(\Delta f_{k,g}(u,v) \right) \right|, \quad u = 1, \ldots, U, \quad v = 1, \ldots, V. \tag{20}$$

$S_{k,g}$ is then compared with a threshold level th_{p} with the output given as:

$$R_{k,g} = \begin{cases} 1, & S_{k,g} \geq th_{\mathrm{p}} \\ 0, & S_{k,g} < th_{\mathrm{p}} \end{cases} \quad l = 1, \ldots, M, \quad i = 1, \ldots, U, \quad j = 1, \ldots, V. \tag{21}$$

$R_{k,g}$ is then correlated with preamble and when matched $Flag_{\mathrm{p}}$ is toggled to "1". The next step is to find the group ID, which together with the mask, represents the Tx's ID. Using (17) and (18) the detector determine N_{g} number of iterations and recovers the group number for all Txs in the same group all at the same time, and then $Flag_{\mathrm{g}}$ is toggled to "1".

S_k is again compared with th_{p} to check for a mask within a particular frame. If $S_k > th_{\mathrm{p}}$, a pixelated version of $\boldsymbol{\Delta f}_k$ is generated by averaging over blocks of $m \times n$ as given by:

$$\Delta f'_{\mathrm{m},l}(u,v) = \underset{u,v}{\mathrm{mean}} \left(\Delta f_{\mathrm{m},l}(u,v) \right), \quad u = (\xi - 1)m + 1, \ldots, \xi m, \quad v = (\zeta - 1)n + 1, \ldots, \zeta n, \quad \xi = 1, \ldots, U/m, \quad \zeta = 1, \ldots, V/n. \tag{22}$$

Lemma 1: For N number of independent normally distributed random variables of X_i, the mean and the variance of $Y = \frac{1}{N} \sum_{i=1}^{N} X_i$ are given by, $\mu_Y = \frac{1}{N} \sum_{i=1}^{N} \mu_X$ and $\sigma_Y^2 = \frac{1}{N^2} \sum_{i=1}^{N} \sigma_X^2$, respectively, where μ_X and σ_X^2 are the mean and variance of X_i, respectively.

Note that, with $n(t)$ considered as GRDV, from Lemma 1, it can be seen that by taking the average over a block of pixels σ_X^2 decreases, radically. Figures 4(a) and (b) illustrates the probability mass function (PMF) of an arbitrary distribution of light for a row of subtraction frame before and after pixilation, respectively. For higher values of m and n, σ_X^2 decreases more, which represents a highly accurate mask. However, the resolution of the mask frame reduces, thus resulting in no detection of the signal from the Txs with footprints smaller than $(m \times n)$ pixels. Following pixelation, every element of $\boldsymbol{\Delta f'}_{\mathrm{m},l}$ is binarized forming a matrix of $U \times V$ as:

$$f_{\mathrm{m},l}(u,v) = \begin{cases} 1, & \Delta f'_{\mathrm{m},l}(u,v) \geq th_{\mathrm{m}} \\ 0, & \Delta f'_{\mathrm{m},l}(u,v) < th_{\mathrm{m}} \end{cases} \quad l = 1, \ldots, M, \quad i = 1, \ldots, U, \quad j = 1, \ldots, V. \tag{23}$$

where $f_{\mathrm{m},l}(u,v)$ denotes the (u,v)-th pixel of the l-th mask frame. Fig. 4 depicts the mask frame extraction process (c) after frame subtraction, (d) after pixelation, and (e) the output mask frame. Since masks are unique for each Tx, they also represent the Tx ID and refer to the area illuminated exclusively by the l-th Tx. Therefore, when a subtracted frame is element-wise-multiplied by the mask matrix, all pixels are set to zero if not illuminated by the l-th Tx. Otherwise, pixels

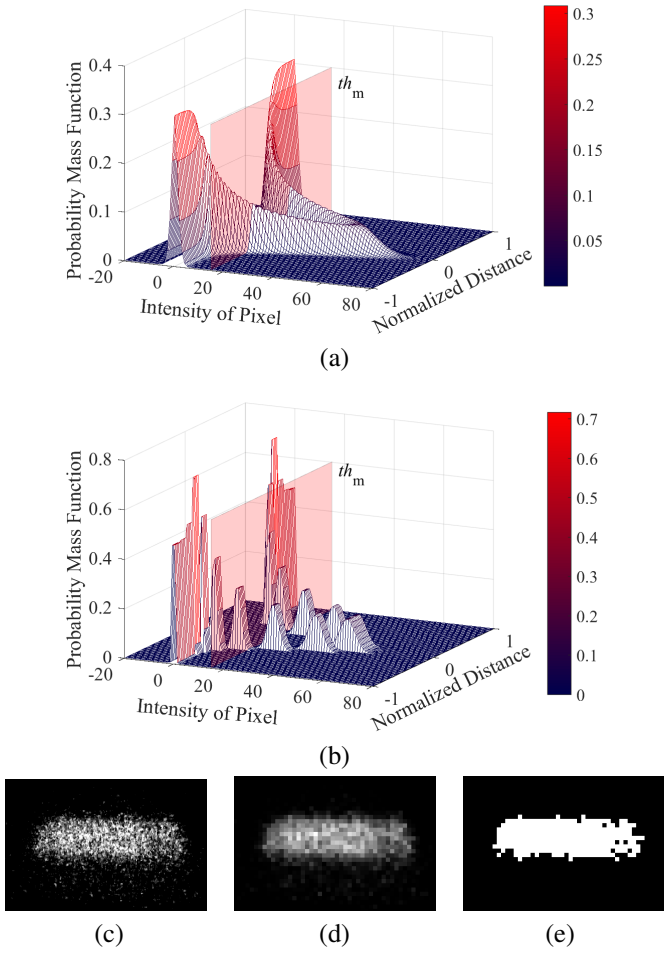

Fig. 4. PMF of intensity of pixels of a row of the subtracted image with Lambertian Tx for pixilation: (a) before, and (b) after and *mask* frame extraction process: (c) frame subtraction (d) pixelation, and (e) mask frame.

remain unchanged. This procedure is performed for M frames and then the mask stage flag is toggled to "1".

3) Mask Matching and EGC: Finally, once masks are identified, the proceeding frames are element-wise-multiplied by all masks as given by:

$$\boldsymbol{Z}_{l,k} = \boldsymbol{f}_{\mathrm{m},l} . \boldsymbol{\Delta f}_k. \tag{24}$$

With the EGC-based detector the average over $\boldsymbol{Z}_{l,k}$, which is a matrix with a size of $U \times V$, is given as:

$$y_{l,k} = \underset{u,v}{\mathrm{mean}} (Z_{l,k}(u,v)) \tag{25}$$

(25) is compared to th_{d} to extract the data of the l-th Tx as:

$$\hat{d}_{l,k} = \begin{cases} 1, & y_{l,k} \geq th_{\mathrm{d}} \\ 0, & y_{l,k} < th_{\mathrm{d}} \end{cases} \tag{26}$$

This approach increases the tolerance of the system to the noise; hence higher SNR levels.

V. Experimental Results

In this work, we investigate the impact of channel parameters on the system performance using a single Tx. Fig. 5 shows the experimental test-bed, which is used to evaluate

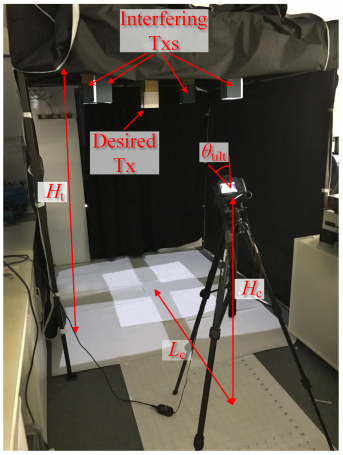

Fig. 5. Experimental setup for the proposed system.

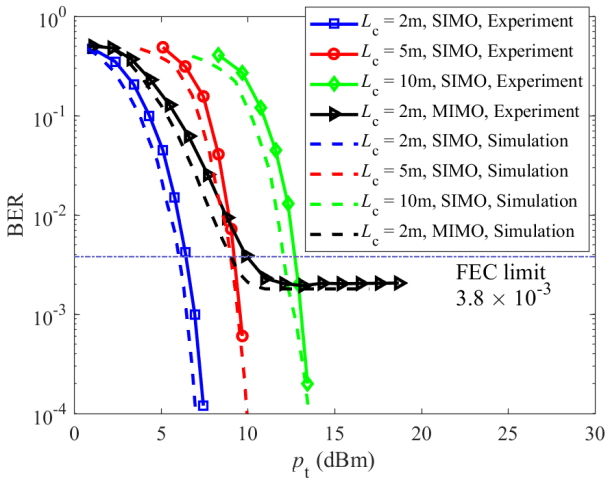

Fig. 6. Predicted and measured BER vs. the transmit power for different L_ss.

the proposed system depicted in Fig. 2(c). For each Tx, an independent 1068-bit long packet composed of 64-bit preamble, 2-bit group, 2-bit mask, and 1000-bit payload created in Matlab was generated in the OOK-NRZ format using an arbitrary waveform generator (TTi TGA12104). For each Tx, we have used a Luxeon rebel (coulour temperature 5600K) LED positioned at the height of 1.8 m from the floor. The LEDs beam have almost first order Lambertian profile. An RB of the size $13 \times 13 \times 35$ cm^3 was used to create a footprint of 0.9×0.9 m^2. The floor surface was covered by white papers with a reflection coefficient of $\rho \cong 0.67$. At the Rx, a Canon Rebel SL1 (EOS 100D) camera with a sensor size of 22.3×14.9 mm and a Canon EF-S 18-55 mm lens positioned 1.3 m above the floor level were used to record a 3 minutes long RGB video stream with a 720p resolution at 60 fps. Note, The typical value of η in case of an Advanced Photo System (APS) sensor is $\sim$0.37 [36]. The white balance was set to off-mode. The pixel area was 18.4 μm^2, θ_{tilt} was 20^o, and JPEG image resolution was 1280$\times$720. The video streams were then processed off-line in Matlab. In the detection process, the mask was created by taking average over blocks of 10×10 pixels. The pixel size of the CMOS-based IS is in $\sim\mu$m, which is far larger than the half wavelength of the LED light and thus

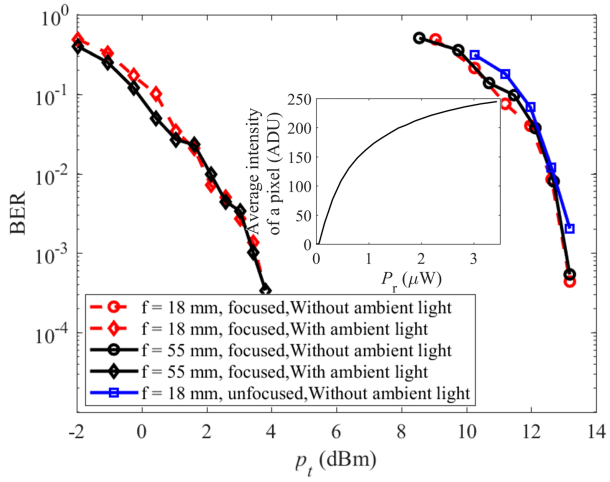

Fig. 7. The BER vs. the transmit power for different focal lengths, ambient light levels, focused and defocused modes.

the camera is able to serve as a spatial diversity Rx. Other system parameters are shown in Table I. In this paper, by applying STDM, described in Section III, we ensure that the interference does not affect the system performance. Hence, in our experimental test bed, we consider only a single LED light source, which can readily be generalized to M-LED.

To assess the system BER as a function of P_t we captured video streams for $f = 18$ mm, aperture of $f/4$, $T_{\text{exp}} = 1/100$ s, and ISO level of 6400, and for SIMO with L_s of 2 m, 5 m and 10 m, and MIMO with $L_s = 2$ m and a 60% overlap area, see Fig. 6. Also illustrated are the simulation plots where the noise variance was experimentally measured as a function of the pixels mean intensity in ADU, which show a close match. It can be seen that shorter L_s requires lower P_t for a given BER value. Note, increasing L_s results in reduced size of the optical footprint in the same rate as the drop in the signal level, hence the mean power in the illuminated area of the image is at a fixed level. Thus, based on Lemma 1 the power penalty due to L_s is related to the noise, which is averaged over a reduced number of pixels. The results also show that, a long-range and low data rate communications system can be established with a very low transmit power of $\sim$13 dBm. Note, MIMO displays higher BER compared to SIMO, which is due to increased level of interference. Moreover, for $P_t >$ $\sim$10 dBm the BER reached a constant level (BER floor) since interference became dominant. Within the experiment, we used 4 independent interfering Txs, where the overlap area can be increased by more than 30% compared to [34]. We attribute this to reduced probability of all interfering Txs being on at the same time.

Next, we located the camera at a distance of 5 m from the Tx, see Fig. 7. We measured the BER against P_t the transmit power for f of 18 and 55 mm with and without the ambient light, see Fig. 7(a). At higher f the size of the reflection footprint increases covering more pixels with the aperture size become larger. Therefore, we observed no improvement in BER when zooming. The results highlight an $\sim$10 dB improvement with the ambient light. This is due

to the non-linear gamma correction curve in the camera (see the inset in Fig. 7(a)). Therefore, the ambient light does not degrade the system performance, but acts as a bias point at the Tx by shifting the signal into the linear region of the camera. Next, for the aperture of $f/4$, T_{exp} of 1/100 s, and a ISO level of 6400 we measured the BER against P_{t} for focused and unfocused cases with no ambient light, see Fig. 7(b). In this case, we observe a slight improvement in BER for the focused case. Changing the camera focusing distance slightly changes the focal point of the lens, but aperture remains the same. Accordingly, we observed a slight increase in the image size for the defocusing case, which resulted in light distribution over a higher number of pixels; hence lower SNR/pixel. We can, therefore, conclude that, although large aperture leads to out-of-focus footprints, the system SNR performance improves. Note that, in order to reduce the frames processing time one option would be to use binning, where the image resolution is reduced by taking the average of $m \times m$ blocks [37].

VI. Conclusion

In this paper, a novel NLOS MIMO STDM system was proposed. To avoid data loss due to interchannel interference, STDM was adopted to allocate time slots to groups of Txs. We proposed a dedicated packet structure based on differential modulation in order to uniquely label the Txs and to ease the reception of the packets. A comprehensive theoretical model and a novel effective detection algorithm were proposed to accurately extract the information from video streams. In the proposed algorithm frame subtraction and a mask were used to remove unwanted background objects, and to effectively locate reflections from a specified Tx to extract the data, respectively. An EGC technique was used to extract the data from the diversity Rx, which made the system more tolerant to the noise. The proposed system was evaluated under fixed conditions and for different channel and camera parameters showing transmission over longer L_{s} under very low transmit power levels. We demonstrated that the system performance improved in the presence of ambient light because of non-linearity of the gamma correction. We also showed that in MIMO, the BER level is lower bound limited due to the interference being the dominant impairment of the system. Finally, we showed that camera defocusing do not have a significant impact on the link performance, hence the camera aperture could be set to the maximum to take advantage of higher received power levels.

Appendix A

Fig. 8 illustrated the boundaries that each pixel covers on the floor plane in the $\overrightarrow{y}$ direction, which can be obtained as:

$$x_u = H_{\text{c}} \times \tan\left(\sum_{i=0}^{u} \theta_i\right) + x_{\text{c}}, \tag{27}$$

where x_{c} and H_{c} are the camera position in the y direction and its height from floor, respectively. Here, $\theta_0 = \theta_{\text{T}} - \frac{\theta_{\text{FoV}}}{2}$ is

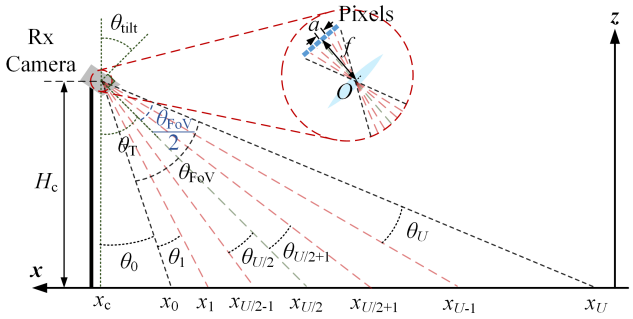

Fig. 8. Side view of areas that each photodetector covers.

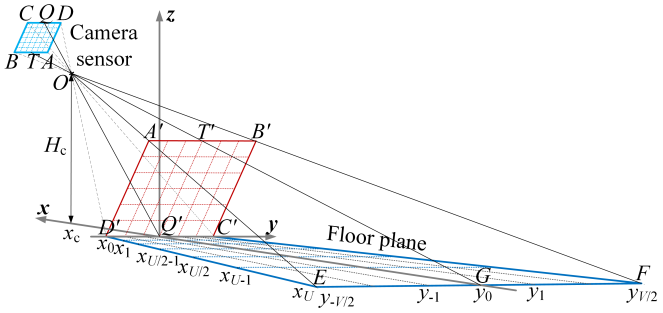

Fig. 9. 3D view of areas covered by PD.

the angle between the normal line to the floor that passes the center of the lens and lower boundary of FoV.

$$\sum_{i=0}^{u} \theta_i = \theta_{\text{T}} + \tan^{-1}\left(\frac{(u - U/2)\, a}{f}\right), \quad 1 \le u \le U \tag{28}$$

where $\theta_{\text{T}} = \pi/2 - \theta_{\text{tilt}}$, a is the width of the pixel. Ultimately, the boundaries can be calculated as:

$$x_u = x_{\text{c}} - H_{\text{c}} \times \frac{(u - U/2) a \cos(\theta_{\text{T}}) + f \sin(\theta_{\text{T}})}{f \cos(\theta_{\text{T}}) - (u - U/2) a \sin(\theta_{\text{T}})}. \tag{29}$$

Appendix B

Fig. 9 shows a 3D illustration of the areas covered by each pixel. $A'B'C'D'$ is a rectangle proportional to and in parallel with $ABCD$, with the lower side touching the floor plane and the line connecting the center of $ABCD$ and the center of $A'B'C'D'$ crosses O. Since $\overline{AB} \| \overline{CD}$, according to the proportionality theorem $\frac{\overline{A'B'}}{\overline{EF}} = \frac{\overline{OT'}}{\overline{OG}}$. Since $OQ'T'$ is an isosceles triangle, $\overline{OT'} = \overline{OQ'} = \frac{H_{\text{c}}}{\cos(\theta_{\text{T}} - \theta_{\text{FoV}}/2}$ and $\overline{OG} = \frac{H_{\text{c}}}{\cos(\theta_{\text{T}}) + \theta_{\text{FoV}}/2}$. Accordingly, if the length of $\overline{A'B'}$ is α then the length of $\overline{EF}$ is given by:

$$\beta = \frac{\cos(\theta_{\text{T}} + \theta_{\text{FoV}}/2)}{\cos(\theta_{\text{T}} - \theta_{\text{FoV}}/2)} \alpha. \tag{30}$$

Knowing that the length of $\overline{OT}$ is $g = \sqrt{(aU/2)^2 + f^2}$, and due to proportionality of OTQ and $OT'Q'$ we have:

$$\alpha = \frac{V H_{\text{c}}}{g \cos(\theta_{\text{T}} - \theta_{\text{FoV}}/2)} a, \tag{31}$$

where M is the number of photodetectors in a row of the camera sensor. The area boundary on the floor plane, which each photodetector covers in $\overrightarrow{x}$ direction, is given by:

$$y_v(x) = \frac{v(\beta-\alpha)}{Vx_U}(x-x_0) + \frac{v\alpha}{V}, \quad -\frac{V}{2} \leq v \leq \frac{V}{2}. \quad (32)$$

ACKNOWLEDGMENT

One of the authors is granted a scholarship by Northumbria University. This work was supported by the UK EPSRC grant EP/P006280/1: Multifunctional Polymer Light-Emitting Diodes with Visible Light Communications (MARVEL) and the GACR 17-17538S.

REFERENCES

[1] P. H. Pathak, X. Feng, P. Hu, and P. Mohapatra, "Visible light communication, networking, and sensing: A survey, potential and challenges," *ieee communications surveys & tutorials*, vol. 17, no. 4, pp. 2047–2077, 2015.

[2] Z. Ghassemlooy, L. N. Alves, S. Zvanovec, and M.-A. Khalighi, *Visible Light Communications: Theory and Applications*. CRC Press, 2017.

[3] S. Hranilovic and F. R. Kschischang, "Short-range wireless optical communication using pixilated transmitters and imaging receivers," in *Communications, 2004 IEEE International Conference on*, vol. 2. IEEE, 2004, pp. 891–895.

[4] P. Luo, M. Zhang, Z. Ghassemlooy, H. Le Minh, H.-M. Tsai, X. Tang, and D. Han, "Experimental demonstration of a 1024-qam optical camera communication system," *IEEE Photonics Technology Letters*, vol. 28, no. 2, pp. 139–142, 2016.

[5] C. Danakis, M. Afgani, G. Povey, I. Underwood, and H. Haas, "Using a cmos camera sensor for visible light communication," in *Globecom Workshops (GC Wkshps), 2012 IEEE*. IEEE, 2012, pp. 1244–1248.

[6] A. Burton, H. Minh, Z. Ghassemlooy, E. Bentley, and C. Botella, "Experimental demonstration of 50-mb/s visible light communications using 4× 4 mimo," *IEEE Photonics Technology Letters*, vol. 26, no. 9, pp. 945–948, 2014.

[7] J. Perez-Ramirez and D. K. Borah, "A single-input multiple-output optical system for mobile communication: Modeling and validation," *IEEE photonics technology letters*, vol. 26, no. 4, pp. 368–371, 2014.

[8] S. Itoh, I. Takai, M. S. Z. Sarker, M. Hamai, K. Yasutomi, M. Andoh, and S. Kawahito, "A cmos image sensor for 10mb/s 70m-range led-based spatial optical communication," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2010 IEEE International*. IEEE, 2010, pp. 402–403.

[9] S.-H. Yang, H.-S. Kim, Y.-H. Son, and S.-K. Han, "Reduction of optical interference by wavelength filtering in rgb-led based indoor vlc system," in *Opto-Electronics and Communications Conference (OECC), 2011 16th*. IEEE, 2011, pp. 551–552.

[10] 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 Photonics Journal*, vol. 5, no. 5, pp. 6 801 418–6 801 418, 2013.

[11] R. Boubezari, H. Le Minh, Z. Ghassemlooy, and A. Bouridane, "Smartphone camera based visible light communication," *Journal of Lightwave Technology*, vol. 34, no. 17, pp. 4121–4127, 2016.

[12] R. Boubezari, H. Le-Minh, Z. Ghassemlooy, and A. Bouridane, "Novel detection technique for smartphone to smartphone visible light communications," in *Communication Systems, Networks and Digital Signal Processing (CSNDSP), 2016 10th International Symposium on*. IEEE, 2016, pp. 1–5.

[13] T.-C. Bui and S. Kiravittaya, "Demonstration of using camera communication based infrared led for uplink in indoor visible light communication," in *Communications and Electronics (ICCE), 2016 IEEE Sixth International Conference on*. IEEE, 2016, pp. 71–76.

[14] T. Yamazato, I. Takai, H. Okada, T. Fujii, T. Yendo, S. Arai, M. Andoh, T. Harada, K. Yasutomi, K. Kagawa *et al.*, "Image-sensor-based visible light communication for automotive applications," *IEEE Communications Magazine*, vol. 52, no. 7, pp. 88–97, 2014.

[15] S. Pergoloni, M. Biagi, S. Colonnese, R. Cusani, and G. Scarano, "A space-time rls algorithm for adaptive equalization: The camera communication case," *Journal of Lightwave Technology*, vol. 35, no. 10, pp. 1811–1820, 2017.

[16] H. Chinthaka, N. Premachandra, T. Yendo, T. Yamasato, T. Fujii, M. Tanimoto, and Y. Kimura, "Detection of led traffic light by image processing for visible light communication system," in *Intelligent Vehicles Symposium, 2009 IEEE*. IEEE, 2009, pp. 179–184.

[17] 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 Photonics Journal*, vol. 8, no. 3, pp. 1–17, 2016.

[18] Y. Kawai, T. Yamazato, H. Okada, T. Fujii, T. Yendo, S. Arai, and K. Kamakura, "Tracking of led headlights considering nlos for an image sensor based v2i-vlc," in *International Conference and Exhibition on Visible Light Communications*, 2015.

[19] T. Nagura, T. Yamazato, M. Katayama, T. Yendo, T. Fujii, and H. Okada, "Tracking an led array transmitter for visible light communications in the driving situation," in *Wireless Communication Systems (ISWCS), 2010 7th International Symposium on*. IEEE, 2010, pp. 765–769.

[20] S. Arai, Y. Shiraki, T. Yamazato, H. Okada, T. Fujii, and T. Yendo, "Multiple led arrays acquisition for image-sensor-based i2v-vlc using block matching," in *Consumer Communications and Networking Conference (CCNC), 2014 IEEE 11th*. IEEE, 2014, pp. 605–610.

[21] C.-W. Chow, C.-Y. Chen, and S.-H. Chen, "Visible light communication using mobile-phone camera with data rate higher than frame rate," *Optics express*, vol. 23, no. 20, pp. 26 080–26 085, 2015.

[22] W.-C. Wang, C.-W. Chow, L.-Y. Wei, Y. Liu, and C.-H. Yeh, "Long distance non-line-of-sight (nlos) visible light signal detection based on rolling-shutter-patterning of mobile-phone camera," *Optics express*, vol. 25, no. 9, pp. 10 103–10 108, 2017.

[23] T. Haruta, T. Nakajima, J. Hashizume, T. Umebayashi, H. Takahashi, K. Taniguchi, M. Kuroda, H. Sumihiro, K. Enoki, T. Yamasaki *et al.*, "4.6 a 1/2.3 inch 20mpixel 3-layer stacked cmos image sensor with dram," in *Solid-State Circuits Conference (ISSCC), 2017 IEEE International*. IEEE, 2017, pp. 76–77.

[24] Phantom, "Phantom v2512," June 2017. [Online]. Available: http://www.phantomhighspeed.com/Products/Ultrahigh-Speed-Cameras/v2512

[25] R. D. Roberts, "Undersampled frequency shift on-off keying (ufsook) for camera communications (camcom)," in *Wireless and Optical Communication Conference (WOCC), 2013 22nd*. IEEE, 2013, pp. 645–648.

[26] P. Luo, Z. Ghassemlooy, H. Le Minh, X. Tang, and H.-M. Tsai, "Undersampled phase shift on-off keying for camera communication," in *Wireless Communications and Signal Processing (WCSP), 2014 Sixth International Conference on*. IEEE, 2014, pp. 1–6.

[27] S.-H. Chen and C.-W. Chow, "Color-shift keying and code-division multiple-access transmission for rgb-led visible light communications using mobile phone camera," *IEEE Photonics Journal*, vol. 6, no. 6, pp. 1–6, 2014.

[28] T. Seybold, C. Keimel, M. Knopp, and W. Stechele, "Towards an evaluation of denoising algorithms with respect to realistic camera noise," in *Multimedia (ISM), 2013 IEEE International Symposium on*. IEEE, 2013, pp. 203–210.

[29] J. C. Chau and T. D. Little, "Analysis of cmos active pixel sensors as linear shift-invariant receivers," in *Communication Workshop (ICCW), 2015 IEEE International Conference on*. IEEE, 2015, pp. 1398–1403.

[30] J. M. Kahn and J. R. Barry, "Wireless infrared communications," *Proceedings of the IEEE*, vol. 85, no. 2, pp. 265–298, 1997.

[31] W. Gu, M. Aminikashani, P. Deng, and M. Kavehrad, "Impact of multipath reflections on the performance of indoor visible light positioning systems," *Journal of Lightwave Technology*, vol. 34, no. 10, pp. 2578–2587, 2016.

[32] H. J. Trussell and M. J. Vrhel, *Fundamentals of digital imaging*. Cambridge University Press, 2008.

[33] H. Tian, "Noise analysis in cmos image sensors," 2000.

[34] N. B. Hassan, Z. Ghassemlooy, S. Zvanovec, P. Luo, and H. Le-Minh, "Non-line-of-sight $2 \times n$ indoor optical camera communications," *Appl. Opt.*, vol. 57, no. 7, pp. B144–B149, Mar 2018.

[35] ITU, "Parameter values for the hdtv standards for production and international programme exchange," February 2018. [Online]. Available: https://www.itu.int/rec/R-REC-BT.709/

[36] B. A. Fowler, A. El Gamal, D. X. Yang, and H. Tian, "Method for estimating quantum efficiency for cmos image sensors," in *Solid State Sensor Arrays: Development and Applications II*, vol. 3301. International Society for Optics and Photonics, 1998, pp. 178–186.

[37] Z. Zhou, B. Pain, and E. R. Fossum, "Frame-transfer cmos active pixel sensor with pixel binning," *IEEE Transactions on electron devices*, vol. 44, no. 10, pp. 1764–1768, 1997.