SVM Aided LEDs Selection for Generalized Spatial Modulation of Indoor VLC Systems

Fangxin Zhang, Fasong Wang, Jiankang Zhang and Ting Zuo

Abstract

In order to reduce the complexity of the light-emitting diodes (LEDs) selection procedure in generalized spatial modulation (GSM) assisted indoor visible light communication (VLC) system, a support vector machine (SVM) aided low complexity and high efficiency machine learning LEDs selection algorithm is proposed for the considered GSM-VLC system. By modeling the LEDs selection problem in indoor GSM-VLC system as a multi-classification task, an optimization problem is constructed by utilizing kernel SVM. After the optimal parameters are obtained from the training stage, an LEDs selection procedure can be accomplished efficiently by SVM aided learning system for any given user's channel state information. Simulation results and complexity analysis show that, compared with traditional LEDs selection algorithms, the proposed SVM aided LED selection algorithm can achieve an ideal bit error ratio (BER) performance while having considerable lower complexity for the considered GSM-VLC system.

Index Terms

LEDs Selection, Visible Light Communication (VLC), Generalized Spatial Modulation (GSM), Support Vector Machine (SVM).

I. Introduction

With the accelerated deployment of 5th generation mobile networks (5G) system and the formal launch of the 6th generation mobile networks (6G) technology, the spectrum resources of traditional wireless communication have gradually been unable to meet human needs, so it is urgent to develop new spectrum to fundamentally solve the contradiction between the demand for super-capacity communication and the spectrum crisis. Visible light communications (VLC) has become an effective supplement to the radio frequency (RF) based wireless communication system by virtue of its rich free spectrum resources, excellent anti-electromagnetic interference capability and relatively high network security [1]–[4]. In

F. Zhang, F. Wang and T. Zuo are with the School of Information Engineering, Zhengzhou University, Zhengzhou, 450001, Henan, China. F. Wang is the *Corresponding author*. (E-mail: iefswang@zzu.edu.cn)

J. Zhang is with Department of Computing & Informatics, Bournemouth University, Poole, BH12 5BB, U.K. (E-mail: jzhang3@bournemouth.ac.uk).

particular, as a promising small cell technology, indoor VLC technology can be connected to existing wireless networks and will probably become an important part of future wireless communication [5], [6].

In terms of physical layer transmission, multi-input multi-output (MIMO) transmission is able to deeply mine wireless resources in spatial dimensions, thus significantly improving the system spectral efficiency, which has become one of the current research hot spots in academia and industry fields. Meanwhile, for indoor VLC system, the MIMO configuration provides physical alignment of light-emitting diodes (LEDs) to mobile user and the lighting function is implemented simultaneously. Therefore, the potential research of theory and experiments of VLC have recently turned to MIMO-VLC system [3].

However, in practical applications of MIMO-VLC, as the number of LEDs equipped at the transmitter ends increasing, the cost and complexity of the system also increase continuously. Moreover, the system will be seriously affected by inter-channel interference (ICI), inter-antenna synchronization (IAS) and other problems, and its overall performance will degrade. As one of successful multiple antenna schemes for wireless communication systems, generalized spatial modulation (GSM) has been extensively studied in the context of VLC, which can effectively solve the above mentioned limitations of MIMO-VLC system due to its less cost, fewer links and lower complexity [7], [8]. In practice, due to the spatial size of a typical application room and the limited luminous flux of a single LED, multiple LEDs are usually adopted for obtaining adequate illumination. If one to several LEDs are activated to transmit information signals among these LEDs, these spatially distributed LEDs can be viewed as the spatial constellation points naturally and conveying information. Therefore, the GSM is very suitable for VLC systems, and in this case, the LEDs are utilized not only for lighting, but also for data transmission. When GSM technology is utilized in the considered indoor VLC system, in order to provide ideal transmission performance, we should try to find the optimal LEDs combination to activate and transmit information during modulation process firstly. To provide ideal transmission performance of the considered indoor GSM-VLC system, for quasi-static scenarios, such as indoor industrial internet of things, meeting room and so on, the second main issue that should be addressed is how to select the optimal LEDs combination quickly. Under this condition, a valuable low complexity LEDs selection algorithm might necessitate to enhance the performance of the considered indoor GSM-VLC system.

Since common GSM-VLC system needs to select the LEDs combination to activate and transmit information during modulation process [9], how to quickly select the optimal LEDs combination online is one of the key problems to be solved. Antenna selection technology has been widely studied and applied in traditional RF aided communication system. Specifically, antenna selection problem in spatial modulation (SM) system was considered in [10] and the Euclidean distance antenna selection (EDAS) algorithm was proposed, while the high complexity of EDAS algorithm limits its application. Subsequently, in order to solve the problem of high complexity of EDAS algorithm, researchers proposed a variety of schemes to reduce the complexity of EDAS algorithm [11], [12]. However, the improved EDAS schemes in [11] and [12] are still unable to get rid of the traversal feature in the algorithms radically, and as a result the actual effect is not greatly improved. To tackle this, some scholars built a smart antenna selection scheme by using machine learning approach, which can better overcome the redundant calculation problems existed in traditional EDAS algorithms. Among them, a machine learning aided antenna selection technology was proposed in [13] for the first time, by mapping antenna selection task as a multi-classification problem, the corresponding system was constructed based on K-nearest neighbor (KNN) and support vector machine (SVM) algorithms, respectively. Following this, an antenna selection technology based on machine learning was proposed in [14] to improve the physical layer security of the considered system. Additionally, by optimizing the antenna selection method of sample features, an improved approach was proposed in [15] by utilizing machine learning technology.

However, methods proposed in [13]–[15] only considered traditional RF based communication system. In study on LEDs selection in indoor VLC system, an LEDs selection algorithm based on maximum minimum singular value was proposed in [16] by utilizing channel state information (CSI) of the transmitter to select the optimal LEDs combination at each brightness level. Furthermore, an LEDs selection algorithm was presented in [17] to minimize system mean square error (MMSE), which effectively improved the communication performance of visible light multi-user multi-input single output system under different brightness. In order to effectively solve the problem of limited number of transmitters in the generalized LED index modulation system, Tran et al. proposed an LEDs selection algorithm, which can effectively improve the performance of the original considered system [18]. However, in all LEDs selection algorithms proposed in [16]–[18] of indoor VLC system, problems of large computational load and high complexity still exist. Up to now, to the best of our knowledge, there are no machine learning aided LEDs selection algorithm in the literature for indoor GSM-VLC systems, which inspired this treatise.

Motivated by the aforementioned issues, in this paper, aided by SVM, we propose a low complexity LEDs selection algorithm of indoor GSM-VLC system. The contributions of this paper can be summarized as follows:

- According to the line of sight (LOS) channel characteristics of indoor GSM-VLC system, the training set is constructed by channel matrices generated by randomly generating user locations. Following this, the maximum minimum Euclidean distance is selected as the key performance index (KPI) of the learning system to construct the label vector of the training samples.
- For the considered indoor GSM-VLC system, the LEDs selection problem is equivalent

to the multi-classification machine learning task, and the optimization problem of LEDs selection is then constructed by using dual theory. Then, given users' given CSI, the online LEDs selection is realized by the trained system.

- Computer simulations and complexity analysis show that, compared with traditional LEDs selection algorithms, the algorithm proposed in this paper can realize online LEDs selection while keeping the performance with low bit error ratio (BER), which shows the effectiveness of the proposed algorithm.

The rest of this paper are organized as follows: The detailed mathematical model of signal and VLC channel model in GSM-VLC system are described in Section II. In Section III, the SVM aided LEDs selection scheme and algorithm are presented in detail; The performance results and related discussions are provided in section IV. Finally, we provide our concluding remarks in Section V.

II. GSM-VLC System and Signals Modelling

In this paper, we consider a GSM aided VLC system utilizing intensity modulation and direct detection (IM/DD), where the information is transmitted from a transmitter assisted by LEDs to a randomly distributed user. In the communication system, the transmitter is equipped with N_t down-facing LEDs installed on the ceiling in the service room, which are used to communicate with receiver, who has N_r photo-detectors (PDs) fixed in up-facing. For simplicity, all the LEDs and PDs are assumed to have the same parameters in this paper, although this is not necessary. Moreover, it is further assumed that the transmitter can get the location information of the receiver [19], which determines the CSI of receiver together with some other parameters that can be determined easily.

Therefore, the VLC system represents a typical MIMO-VLC Gaussian channel model, and the LEDs and the PDs are equipped in a room of size $\tilde{L} \times \tilde{W} \times \tilde{H}$ (m^3) [9], [20], as depicted in Fig. 1. The LEDs are placed at a height of 0.4 m below the ceiling and the PDs are located on a table of height 0.85 m. The observed signal by the PDs of randomly distributed user is expressed as:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w} \tag{1}$$

where $\mathbf{H} \in \mathbb{R}_+^{N_r \times N_t}$ represents the MIMO channel gains between transmitter and receiver link. The transmitted information-bearing signal vector sent by transmitter $\mathbf{x} = [x_1, x_2, \cdots, x_{N_t}]^{\mathrm{T}} \in \mathbb{R}^{N_t}$, as in (1), is assumed to be a signal vector superimposed on an identical direct current (DC) bias $I_{\mathrm{DC}} \in \mathbb{R}_+$ for the purpose to adjust the illumination level of LEDs [21]. In order to avoiding clipping distortion, preserving battery, and also for the sake of safety, the total current $I_{\mathrm{DC},i} + x_i$ for the ith LED is restricted within the range of $[(1-\tilde{\alpha})I_{\mathrm{DC}}, (1+\tilde{\alpha})I_{\mathrm{DC}}]$, where $\tilde{\alpha} \in [0, 1]$ is termed as the modulation index [21], [22]. As a result, the information-bearing signal x_i has to satisfy the peak amplitude constraint of $|x_i| \leq A, \forall i$ with $A =$

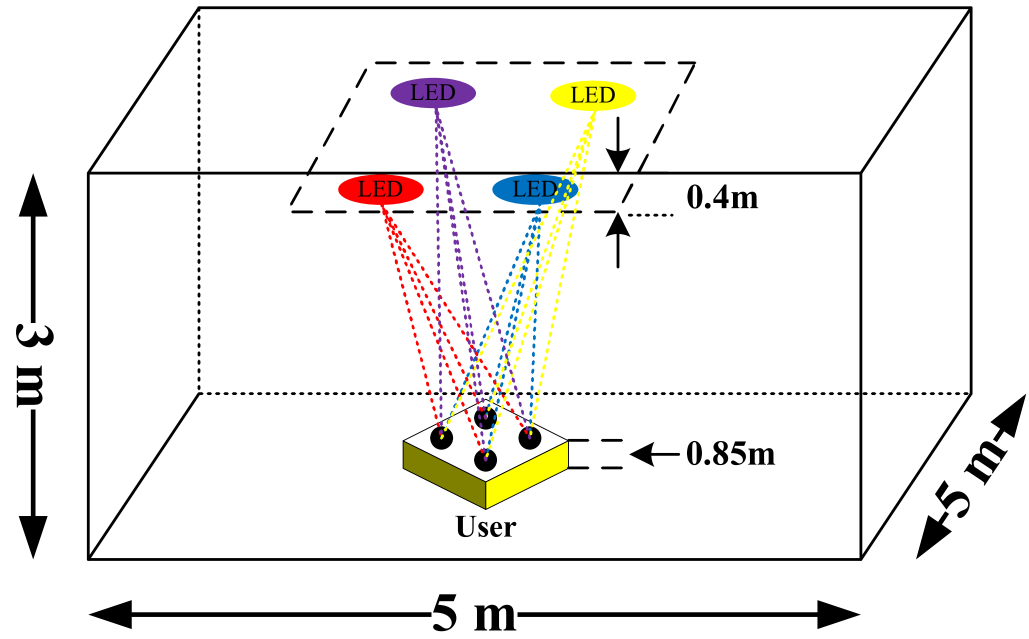

Fig. 1: An indoor GSM-VLC system contains a transmitter with 4 LEDs, one user and the receiver is equipped with 4 PDs.

$\tilde{\alpha} I_{\text{DC}} \in \mathbb{R}_{+}$. Additionally, in (1), $\mathbf{w} \sim \mathcal{N}(\mathbf{0}_{N_r}, \sigma^2 \mathbf{I}_{N_r})$ is zero-mean additive white Gaussian noise (AWGN), received at PDs of the receiver.

A. The VLC channel Model

In indoor VLC, with a generalized Lambertian emission pattern, the path gain G_{ij} between the ith LED and the jth PD of receiver can be expressed as [23]:

$$G_{ij} = \begin{cases} \frac{1}{2\pi d_{ij}^2 \sin^2(\Psi_{\text{FoV}})}(L+1) A_{\text{PD}} \beta^2 \cos^L(\phi) \cos\psi_{ij}, & \text{if} \quad 0 \le |\psi_{ij}| \le \Psi_{\text{FoV}}, \\ 0, & \text{if} \quad |\psi_{ij}| > \Psi_{\text{FoV}} \end{cases} \tag{2}$$

where $L = -\ln(2)/\ln\left(\cos(\Phi_{1/2})\right)$ is the index of Lambertian emission with half irradiance at $\Phi_{1/2}$, which is measured from the optical axis of the LED, d_{ij} is the LoS distance between the ith LED and the jth PD, ϕ is the angle of irradiance of LED, ψ_{ij} is the angle of incidence of the ith LED and jth PD optical link, which is measured from the axis normal to the receiver surface. Also, for receiver, β is the refractive index of the optical concentrator and A_{PD} is the PD area. Finally, Ψ_{FoV} is the receiver's field-of-view (FoV) semi-angle. The geometric model of LoS transmission is demonstrated in Fig. 2.

As a whole, the VLC channel gain between the ith LED and the jth PD of receiver can be expressed as:

$$h_{ij} = TR\eta G_{ij}, \quad i = 1, 2, \cdots, N_t, \; j = 1, 2, \cdots, N_r \tag{3}$$

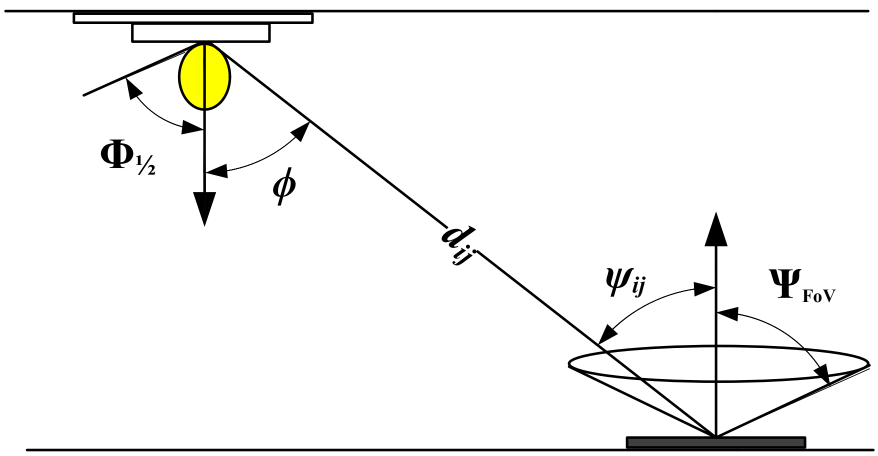

Fig. 2: Illustration of Geometric description parameters for VLC LoS path gain in the Lambertian model.

where T is the gain of a transimpedance amplifier, R is the responsivity of the PD and η is the current-to-light conversion efficiency of the LEDs, respectively.

Note that, from Eq. (2), we can readily conclude that the VLC channel gain h_{ij} depend on the specific position of the transmitter LEDs and the receiver PDs. If an LED is not in a receiver's FoV, the channel gain of the link will be zero [21].

Additionally, for indoor VLC, the received light signals by the PDs of receiver are a summation of the LoS component and multiple non-LoS (NLoS) ones due to walls reflection of the service room. However, the total received optical power conveyed by the LoS link is more than 95% at the receiver [24]. Moreover, even the strongest NLoS component is still at least 7 dB lower than the strongest LoS one [24]. Consequently, when considering that the transmit LEDs are installed on the ceiling of the service area and face down-forwards, the channel model in Eq. (2) can neglect the NLoS components, but consider only the LoS component for carrying out tractable analysis.

B. Signal Model

Let us assume that the transmitter is formed by $\tilde{N}$ LEDs in the considered service area. For proposed GSM-VLC system, we assume that N_t ($N_t \leq \tilde{N}$) LEDs are utilized for implementation of the GSM modulation among the $\tilde{N}$ LEDs. During one specific symbol duration, among the selected N_t transmit LEDs, n_t LEDs are activated to transmit a specific information symbol, while the rest ($\tilde{N} - n_t$) LEDs are only used for illumination. Hence, there are in total $M' = \binom{N_t}{n_t}$ possible combinations, among which 2^{m_l} are used transmitting m_l bits per symbol, and $m_l = \lfloor \log_2 \binom{N_t}{n_t} \rfloor$, where $\lfloor \cdot \rfloor$ denotes the floor operation.

Furthermore, the length of the remaining bits per symbol is $m_s = \log_2(|\mathbb{M}|)$, where $|\mathbb{M}|$ is the order of the IM pulse amplitude modulation (PAM) constellation and is assumed to be the power of 2, and we further assume that $\mathcal{S} = \{s_m\}_{m=1}^{|\mathbb{M}|}$, where $s_m = \frac{2Im}{|\mathbb{M}|+1}$ for $m = 1, 2, \cdots, |\mathbb{M}|$, I is the mean optical intensity emitted. Therefore, the number of binary bits per GSM symbol is $n_{\text{GSM}} = m_l + n_t m_s = \lfloor \log_2 M' \rfloor + n_t \log_2(|\mathbb{M}|)$. Specifically, we assume that an independent and identically distributed random bit sequence $\{\cdots, b_1, b_2, \cdots, b_r, \cdots\}$ puts into the GSM mapper, where the bit sequence is divided into blocks of n_{GSM} bits that are mapped into the GSM symbols $\mathbf{x}, \mathbf{x} \in \mathcal{X}$, and $\mathcal{X}$ is the $\binom{N_t}{n_t}|\mathbb{M}| = M'|\mathbb{M}|$ GSM symbols set. Aided by $\mathbf{x}$, special LEDs is selected to transmit a symbol with particular optical intensity chosen from the set of $\mathcal{S}$. Consequently, the selected LEDs will transmit the signal with a special intensity $s_m, m = 1, 2, \cdots, |\mathbb{M}|$ equiprobable at this particular time instant and all other LEDs remain silent just for lighting. The modulation process of GSM-VLC is shown in Fig.3.

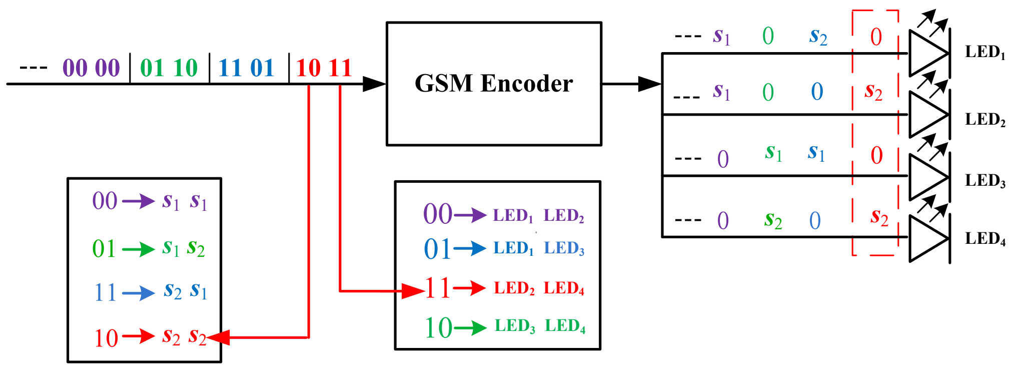

Fig. 3: Illustration of GSM-VLC system with configuration $N_t = 4, n_t = 2, N_r = 4, M = 2$

Hence, the transmitted GSM-VLC signal vector $\mathbf{x}$ can be expressed as follows:

$$\mathbf{x} = [0 \ \cdots \ 0 \ s_{i_1} \ 0 \ \cdots \ 0 \ s_{i_2} \ 0 \ \cdots \ 0 \ s_{i_{n_t}} \ 0 \ \cdots \ 0]^{\mathrm{T}}, \tag{4}$$

where $n_i \in \{1, 2, \cdots, N_t\}, i = 1, 2, \cdots, n_t$, represents the index of the n_tth activated LED, $s_{i_1}, s_{i_2}, \cdots, s_{i_{n_t}} \in \mathcal{S}$. Accordingly, the received signals $\mathbf{y} \in \mathbb{R}^{N_r}$ could be simplified from (1) as:

$$\mathbf{y} = [\mathbf{h}_{i_1} \cdots \mathbf{h}_{i_{n_t}}][s_{i_1} \cdots s_{i_{n_t}}]^{\mathrm{T}} + \mathbf{w} = \sum_{\imath=i_1}^{i_{n_t}} \mathbf{h}_\imath s_\imath + \mathbf{w} = \mathbf{H}_\imath \mathbf{s} + \mathbf{w} \tag{5}$$

where $\mathbf{H}_\imath = [\mathbf{h}_{i_1} \ \mathbf{h}_{i_2} \ \cdots \ \mathbf{h}_{i_{n_t}}]$ represents the submatrix of $\mathbf{H}$ with n_t columns, which are determined by the index of the n_tth activated LEDs, and $\mathbf{s} = [s_{i_1} \ s_{i_2} \ \cdots \ s_{i_{n_t}}]^{\mathrm{T}}$ denotes the transmit symbol vector corresponding to the LEDs $\imath = \{i_1, \ i_2, \ \cdots, i_{n_t}\}$.

Note that, the above-described GSM-VLC system is reduced to the SM-VLC system, when $n_t = 1$. Therefore, the SM-VLC scheme is a special case of our GSM-VLC, and hence, all the following analysis and results can be straightforwardly applied to the SM-VLC by letting $n_t = 1$.

III. SVM Aided LEDs Selection Scheme for GSM-VLC System

In this section, we commence to detail the proposed LEDs selection scheme for the considered GSM-VLC system, by exploiting one of the popular machine learning approach, namely, SVM. For a practical GSM-VLC system, the transmitter is always equipped with multiple LEDs. Aided by this premise, the LEDs selection can be modelled as a multiple classification problem. Specifically, according to the characteristic of the LoS channel, the feature matrix can be designed by the user position obeying uniform distribution. Then, we propose KPI based on minimum Euclidean distance maximization approach. Following this, the optimization problem for LEDs selection is constructed and it is demonstrated that this optimization problem can be solved by its convex dual problem. By utilizing the optimal solutions of the proposed optimization problem, a off-line learning system can be constructed with training samples. Finally, the trained learning system can be employed to carry out online LEDs selection efficiently.

A. *Euclidean Distance Aided LEDs Selection*

At the time of writing, most of the traditional antenna/LEDs selection treatises considered EDAS approach [10]. In a little more detail, we assume that n_t LEDs are selected from N_t transmit LEDs for delivering information, and $\mathcal{I}$ represents the set of all possible selected LEDs combinations, where $|\mathcal{I}| = M' = \binom{N_t}{n_t}$ is the number of LEDs combinations. The objective function of the EDAS can be written as:

$$\jmath_{\text{ED}} = \arg\max_{\jmath \in \mathcal{I}} \Big\{ \min_{\tilde{\mathbf{x}}_1 \neq \tilde{\mathbf{x}}_2 \in \tilde{\mathcal{X}}} \|\mathbf{H}_\jmath(\tilde{\mathbf{x}}_1 - \tilde{\mathbf{x}}_2)\|_2^2 \Big\} \tag{6}$$

where $\mathbf{H}_\jmath \in \mathbb{R}_+^{N_r \times n_t}$ is the submatrix with n_t columns given by $\jmath$ and selected from the channel matrix $\mathbf{H}$, n_t is the number of selected active antennas from N_t transmitting LEDs. $\tilde{\mathcal{X}}$ represents the set of all possible transmit information-bearing signal vectors, and the element of it has the form $\tilde{\mathbf{x}} = [\tilde{x}_{i_1}\ \tilde{x}_{i_2}\ \cdots\ \tilde{x}_{i_{n_t}}]^{\mathrm{T}}$. Generally, EDAS scheme can obtain ideal BER performance for antenna/LEDs selection in MIMO systems. However, the fundamental exhaustive search nature and high dimensional signal space render its very high computational complexity. Though some related algorithms have been optimized to reduce the complexity of EDAS [11], [12], there still exist some redundant computations for the highly correlated channel matrices of some time slots, which may lead to repetitive computations. In order to solve the redundant calculation problem of LEDs selection in GSM-VLC systems, an efficient LEDs selection shceme based on SVM will be constructed to solve this high computational complexity problem.

B. Principle of Kernel SVM

As a popular machine learning approach, the fundamental purpose of SVM is to make classification from a labeled training sample data set. Assume that we have a sample data set denoted as $\mathcal{D} = \{(\mathbf{u}_p, v_p), p = 1, \cdots, P\}$ and $P = P^+ + P^-$, where $\mathbf{u}_p \in \mathbb{R}^n$ are feature vectors, $v_p \in \{+1, -1\}$ are the labels, P^+ and P^- represent the numbers belonging to labels $+1$ and -1, respectively. If the labeled training sample data forms two disjoint convex hull in $\mathbb{R}^n$, by the strong separation theorem, there exist a hyperplane to segment these samples [26], the hyperplane can be represented as:

$$H(\mathbf{u}) = \mathbf{w}^{\mathrm{T}}\mathbf{u} + \varrho = 0, \tag{7}$$

where $\mathbf{w}$ is the normal vector which determines the direction of the hyperplane and ϱ controls the distance of the hyperplane to the origin. By this classification function, any sample $\mathbf{u}$ with $H(\mathbf{u}) > 0$ will belong to Class 1 and those with $H(\mathbf{u}) < 0$ will be recognized as Class 2.

In order to maximize the distance between the hyperplane and the training samples that are closest to the considered hyperplane, the parameters $\mathbf{w}$ and ϱ can be readjusted to make some points lie on either hyperplane $\{\mathbf{u} \mid \mathbf{w}^{\mathrm{T}}\mathbf{u} + \varrho = -1\}$ or $\{\mathbf{u} \mid \mathbf{w}^{\mathrm{T}}\mathbf{u} + \varrho = +1\}$ and these sample points are termed as the support vectors. Consequently, the distance between these two parallel hyperplane is termed as margin and equals $\frac{2}{\|\mathbf{w}\|_2^2}$, where $\|\cdot\|$ is used to denote the Euclidean norm. Then, the classification problem is turned to an optimization problem. As a result, we would like to find the separating hyperplane with the maximum margin. Accordingly, by introducing the sign variables v_p and simple operation, the optimization problem can be expressed as:

$$\begin{aligned} \min_{\mathbf{w},\varrho} \quad & \frac{1}{2}\|\mathbf{w}\|_2^2 \\ \text{s.t.} \quad & v_p(\mathbf{w}^{\mathrm{T}}\mathbf{u}_p + \varrho) \geq 1, \;\; p = 1, \cdots, P \end{aligned} \tag{8}$$

Optimization problem in Eq. (8) is a convex quadratic programming problem, it could be resolved by existing software package. Actually, problem in Eq. (8) can also be solved by its dual problem more efficiently by lagrangian multiplier method, the corresponding generalized Lagrangian function of optimization problem Eq. (8) can be written as:

$$L(\mathbf{w}, \varrho, \boldsymbol{\alpha}) = \frac{1}{2}\|\mathbf{w}\|_2^2 - \sum_{p=1}^{P} \alpha_p[(\mathbf{w}^{\mathrm{T}}\mathbf{u}_p + \varrho) - 1] \tag{9}$$

where $\boldsymbol{\alpha} = [\alpha_1, \cdots, \alpha_P]^{\mathrm{T}} \in \mathbb{R}_+^P$ is the associated Lagrange multiplier vector or termed as dual variable vector. Due to the Karush-Kuhn-Tucker conditions for differentiable convex problem, we can arrive at the dual problem of the optimization problem Eq. (8) as:

$$
\begin{aligned}
\max_{\boldsymbol{\alpha}} \quad & \sum_{p=1}^{P} \alpha_p - \frac{1}{2}\sum_{p=1}^{P}\sum_{q=1}^{P} \upsilon_p \upsilon_q \alpha_p \alpha_q \mathbf{u}_p^{\mathrm{T}} \mathbf{u}_q \\
\text{s.t.} \quad & \sum_{p=1}^{P} \upsilon_p \alpha_p = 0, \\
& \alpha_p \geq 0, \;\; p = 1, \cdots, P
\end{aligned} \tag{10}
$$

Suppose the optimal solution of the primal and dual optimization problem in Eq. (8) and Eq. (10) are denoted as $\mathbf{w}^\star, \varrho^\star, \boldsymbol{\alpha}^\star$, respectively. Then, the parameter $\mathbf{w}^\star$ and $\varrho^\star$ can be represented by $\boldsymbol{\alpha}^\star$ respectively as:

$$
\mathbf{w}^\star = \sum_{p=1}^{P} \alpha_p^\star \upsilon_p \mathbf{u}_p \tag{11}
$$

$$
\varrho^\star = \frac{1}{|\mathcal{V}|} \sum_{n \in \mathcal{V}} \left[\upsilon_n - \mathbf{w}^{\star\mathrm{T}} \mathbf{u}_n \right] \tag{12}
$$

where $\mathcal{V}$ is the index set of all support vectors and $|\mathcal{V}|$ denotes the cardinality of set $\mathcal{V}$. Hence, for any new coming data $\mathbf{u}$, the decision of classification can be obtained by:

$$
\text{sign}\left[\mathbf{w}^{\star\mathrm{T}} \mathbf{u} + \varrho^\star\right] = \text{sign}\left(\sum_{p=1}^{P} \alpha_p^\star \upsilon_p \mathbf{u}_p^{\mathrm{T}} \mathbf{u} + \varrho^\star \right) \tag{13}
$$

Nevertheless, for general classification problem, the training data samples cannot be separate by a hyperplane. In this case, the kernel SVM is proposed by using a nonlinear classification function $G(\mathbf{u}) = \mathbf{w}^{\mathrm{T}} \zeta(\mathbf{u}) + \varrho$, where $\zeta(\mathbf{u}) : \Omega \mapsto \mathbb{H}$ is a nonlinear feature mapping function from sample vector space Ω to a Hilbert space $\mathbb{H}$ termed as the feature space [29],similarly with Eq. (8), we have the following convex quadratic programming problem:

$$
\begin{aligned}
\min_{\mathbf{w}, \varrho} \quad & \frac{1}{2} \|\mathbf{w}\|_2^2 \\
\text{s.t.} \quad & \upsilon_p(\mathbf{w}^{\mathrm{T}} \zeta(\mathbf{u}_p) + \varrho) \geq 1, \;\; p = 1, \cdots, P
\end{aligned} \tag{14}
$$

Then the dual problem of Eq. (14) can be expressed as:

$$
\begin{aligned}
\max_{\boldsymbol{\alpha}} \quad & \sum_{p=1}^{P} \alpha_p - \frac{1}{2}\sum_{p=1}^{P}\sum_{q=1}^{P} \upsilon_p \upsilon_q \alpha_p \alpha_q (\zeta(\mathbf{u}_p))^{\mathrm{T}} \zeta(\mathbf{u}_q) \\
\text{s.t.} \quad & \sum_{p=1}^{P} \upsilon_p \alpha_p = 0, \\
& \alpha_p \geq 0, \;\; p = 1, \cdots, P
\end{aligned} \tag{15}
$$

By defining the kernel function $\kappa(\mathbf{a},\mathbf{b}) = (\zeta(\mathbf{a}))^{\mathrm{T}}\zeta(\mathbf{b})$, Eq. (15) can be expressed as:

$$\begin{aligned} \max_{\boldsymbol{\alpha}} \quad & \sum_{p=1}^{P}\alpha_p - \frac{1}{2}\sum_{p=1}^{P}\sum_{q=1}^{P} v_p v_q \alpha_p \alpha_q \kappa(\mathbf{u}_p, \mathbf{u}_q) \\ \text{s.t.} \quad & \sum_{p=1}^{P} v_p \alpha_p = 0, \\ & \alpha_p \geq 0, \;\; p = 1, \cdots, P \end{aligned} \tag{16}$$

For real Euclidean space, the kernel function $\kappa(\cdot,\cdot)$ can be chosen arbitrarily by the guarantee of Mercer's condition [27]. In this way, we can solve this linear classifying problem in higher-dimensions, which is equal to handling the nonlinear problem in the original dimension space. In this paper, we use the radial Gaussian kernel which is commonly adopted in SVM classification, which is defined as:

$$\kappa(\mathbf{a}_1, \mathbf{a}_2) = \exp\Big(- \frac{\|\mathbf{a}_1 - \mathbf{a}_2\|_2^2}{2\sigma^2}\Big) \tag{17}$$

Suppose the optimal solution of the primal and dual optimization problem in Eq. (14) and Eq. (15) are denoted as $\mathbf{w}^\star, \varrho^\star, \boldsymbol{\alpha}^\star$, respectively. Then, the parameter $\mathbf{w}^\star$ and $\varrho^\star$ can be represented by $\boldsymbol{\alpha}^\star$ respectively as:

$$\mathbf{w}^\star = \sum_{p=1}^{P} \alpha_p^\star v_p \zeta(\mathbf{u}_p) \tag{18}$$

$$\varrho^\star = \frac{1}{|\mathcal{V}|}\sum_{n\in\mathcal{V}} \Big[v_n - \mathbf{w}^{\star\mathrm{T}}\zeta(\mathbf{u}_n)\Big] \tag{19}$$

where $\mathcal{V}$ and $|\mathcal{V}|$ are defined as in (12). Hence, for any new coming data $\tilde{\mathbf{u}}$, the decision of classification can be obtained by Eq. (13).

C. SVM Aided LEDs Selection Scheme for GSM-VLC Systems

In order to select the optimal LEDs combination efficiently in the considered GSM-VLC system, a SVM procedure is considered in this paper, which is capable of acquiring the optimal LEDs subset from the trained classification model by SVM algorithm online. With enough LoS channel matrices of the considered system, a five stage LEDs selection strategy is proposed, which are termed as: 1) Generating training data set; 2) Designing KPI; 3) Labeling samples; 4) Constructing learning system; 5) Selecting the optimal LEDs combination. The detailed procedures are shown as follows.

1) Generating training data set: Generally, the training data set is the input of the considered SVM aided learning system. In our GSM-VLC system, the $N_r \times N_t$ dimensional LoS channel matrices $\mathbf{H}$ are utilized as training samples. For the nature of VLC system, its entries h_{ij} are real-valued, where h_{ij} is the (i,j)-th element of channel matrix $\mathbf{H}$. Hence, the channel matrix $\mathbf{H}$ need to be manipulated in order to obtain a real-valued feature vector

$\tilde{\mathbf{d}}$. Following this, the obtained feature vector will be normalized to avoid significant bias in the training process [15]. Assume that we have K LoS channel matrices $\mathbf{H}$ by randomly generating users' position, then performing the following steps to obtain desired feature matrix $\mathbf{D}$.

step 1 Generating the real-valued feature vector $\tilde{\mathbf{d}}_k \in \mathbb{R}^{1\times N}$, which is the vectorization form of channel matrix $\mathbf{H}_k$, which is defined as: $\tilde{\mathbf{d}}_\mathbf{k} = \mathbf{vec}(\mathbf{H}_k) = (h_{k,11}, h_{k,21}, \cdots, h_{k,N_r1}, h_{k,21}, h$ $\mathbb{R}^{1\times(N_r\times N_t)}$ from the channel matrix $\mathbf{H}_k, k = 1, \cdots, K$.

step 2 Repeating *step* 1 until all feature vectors $\tilde{\mathbf{d}}_k$ are generated from channel matrix $\mathbf{H}_k, k = 1, \cdots, K$.

step 3 Forming a training data set $\{\tilde{\mathbf{d}}_k, k = 1, \cdots, K\}$, then constructing training data matrix $\tilde{\mathbf{D}} = [\tilde{\mathbf{d}}_1, \cdots, \tilde{\mathbf{d}}_K]^{\mathrm{T}} \in \mathbb{R}^{K\times N}$.

step 4 Normalizing the matrix $\tilde{\mathbf{D}}$ and generating a normalized feature matrix $\mathbf{D}$, where the (i, j)-th element of $\mathbf{D}$ is expressed as:

$$d_{ij} = \frac{\tilde{d}_{ij} - \mathbb{E}_i\{\tilde{d}_{ij}\}}{\max_i\{\tilde{d}_{ij}\} - \min_i\{\tilde{d}_{ij}\}} \tag{20}$$

2) Designing KPI: The KPI is usually designed to label training samples, it can be defined by various metrics according to the considered problems in communication system, such as the norm of an effective LoS channel, the effective received signal-to-noise ratio (SNR), received signal power, BER, and so on. In this paper, as done in many antenna selection approaches [10]–[12], we utilize Eq. (6) as the KPI.

3) Labeling samples: Suppose that the set of labels and antenna combination are denoted by $\mathcal{L}$ and $\mathcal{I}$, respectively. According to the previous analysis on multiple classification and LEDs selection, a one-to-one mapping exists between the label set and the antenna combination set. It may be readily seen from Eq. (6) that the LEDs combination with less channel correlation will be more likely to be chosen for better performance of the considered system. Hence, we commence to design LEDs combinations set $\mathcal{I}$ with less channel correlation. However, if all the channel gains are uncorrelated or the correlations among channels are unknown, the LEDs combinations set $\mathcal{I}$ needs to be constructed with all possible LEDs combinations. A mapping table between ℓ and i_k is depicted in Tab. I, where $\ell \in \mathcal{L}$, $i_k \in \mathcal{I}$ and $(N_t, n_t) = (5, 2)$. The above labeling process can be summarized as the following steps:

step 5 For the kth channel matrix sample $\mathbf{H}_k$, calculating the KPI corresponding to each combination i_k and expressing it with a particular label $\ell \in \mathcal{L}$.

step 6 Find the LEDs combination i_k with best KPI and its corresponding label ℓ^*, then establish the label vector $\mathbf{c} \in \mathbb{R}^{K\times 1}$ with its kth element c_k as ℓ^*.

step 7 Repeat the above steps until the corresponding labels of all samples $\mathbf{H}_k$, $k = 1, 2, \cdots, K$, are obtained.

TABLE I: An example mapping between labels and LEDs combinations with configuration $(N_t, n_t) = (5, 2)$

lable	**LEDs combinations**
$\ell = 1$	i_1=$[1, 2]$
$\ell = 2$	i_2=$[1, 3]$
$\ell = 3$	i_3=$[1, 4]$
$\ell = 4$	i_4=$[1, 5]$
$\ell = 5$	i_5=$[2, 3]$
$\ell = 6$	i_6=$[2, 4]$
$\ell = 7$	i_7=$[2, 5]$
$\ell = 8$	i_8=$[3, 4]$
$\ell = 9$	i_9=$[3, 5]$
$\ell = 10$	i_{10}=$[4, 5]$

4) Constructing learning system: With the obtained feature matrix $\mathbf{D}$ and its corresponding label vector $\mathbf{c}$ by above steps, a learning system for multiple classifications can be constructed to select LEDs combination of the considered GSM-VLC system. For convenience, let $\mathbf{d}_{k:}$ denotes the kth row of the feature matrix $\mathbf{D}$. The detailed procedure is as follows:

step 8 Suppose that $\mathbf{D}_\ell$ is a sub-training feature matrix, the rows of $\mathbf{D}_\ell$ is composed by $\mathbf{d}_{k:}$ with $c_k = \ell$ for all $k \in \{1, \cdots, K\}$. Following this operation, another sub-training feature matrix $\mathbf{D}_{\bar{\ell}}$ can be obtained, which is a complementary matrix of $\mathbf{D}_\ell$ by eliminating the row vectors of $\mathbf{D}_\ell$ from $\mathbf{D}$. Thus, an SVM can be performed to classify these two sub-training feature matrices $\mathbf{D}_\ell$ and $\mathbf{D}_{\bar{\ell}}$.

step 9 Generating a binary label vector $\mathbf{b}_\ell = [b_{\ell 1}, \cdots, b_{\ell K}]^{\mathrm{T}}$, with its entries $b_{\ell,k}$ are defined as

$$b_{\ell,k} = \begin{cases} 1, & c_k = \ell \\ 0, & \text{else} \end{cases} \tag{21}$$

step 10 One-vs-rest (ℓ-vs-$\bar{\ell}$) binary label classification method: As stated before, for general classification problem, the training feature samples are generally unseparate by a single hyperplane linearly. In order to commence this problem, the kernel SVM is proposed by using kernel function $\kappa(\cdot, \cdot)$, following with Eq. (22), aided by slack variables and KKT conditions, the following convex quadratic programming problem are derived

$$\begin{aligned}\max_{\boldsymbol{\alpha}} \quad & \sum_{k=1}^{K}\alpha_k - \frac{1}{2}\sum_{k=1}^{K}\sum_{\tilde{k}=1}^{K} b_{\ell k}b_{\ell\tilde{k}}\alpha_k\alpha_{\tilde{k}}\kappa(\mathbf{d}_k,\mathbf{d}_{\tilde{k}})\\ \text{s.t.} \quad & \sum_{k=1}^{K} b_{\ell k}\alpha_k = 0,\\ & \alpha_k \geq 0,\;\; k=1,\cdots,K\end{aligned}\tag{22}$$

where $\boldsymbol{\alpha} \in \mathbb{R}_{+}^{P}$ is the dual variable vector. $\kappa(\mathbf{d}_k,\mathbf{d}_{\tilde{k}})$ is a kernel function of feature vectors $\mathbf{d}_k$ and $\mathbf{d}_{\tilde{k}}$, which is utilized to map unseparate linearly feature samples from lower dimensional to higher dimensional. There are several popular kernel function in practice as disscussed in [28]. The Gaussian radial basis kernel function is employed in this paper, which is defined as:

$$\kappa(\mathbf{d}_k,\mathbf{d}_{\tilde{k}}) = \exp\Big(-\frac{\|\mathbf{d}_k-\mathbf{d}_{\tilde{k}}\|_2^2}{2\sigma^2}\Big)\tag{23}$$

Aided by some famous convex optimization toolbox, convex quadratic programming problem Eq. (22) can be resolved efficiently, and the optimal solution is denoted as $\boldsymbol{\alpha}_{\ell}^{\star} = [\alpha_{\ell 1}^{\star},\cdots,\alpha_{\ell M}^{\star}]^{\mathrm{T}}$. Upon involving this optimal solution $\boldsymbol{\alpha}_{\ell}^{\star}$, the parameter $\mathbf{w}_{\ell}^{\star}$ and $\varrho_{\ell}^{\star}$ can be represented by $\boldsymbol{\alpha}_{\ell}^{\star}$ respectively as:

$$\mathbf{w}_{\ell}^{\star} = \sum_{k=1}^{K}\alpha_{\ell k}^{\star}b_{\ell k}\zeta(\mathbf{d}_k)\tag{24}$$

$$\begin{aligned}\varrho_{\ell}^{\star} &= \frac{1}{|\mathcal{V}|}\sum_{n\in\mathcal{V}}\Big[b_{\ell n} - \mathbf{w}_{\ell}^{\star\mathrm{T}}\zeta(\mathbf{d}_n)\Big]\\ &= \frac{1}{|\mathcal{V}|}\sum_{n\in\mathcal{V}}\Big[b_{\ell n} - \sum_{k=1}^{K}\alpha_{\ell k}^{\star}b_{\ell k}\zeta(\mathbf{d}_k)^{\mathrm{T}}\zeta(\mathbf{d}_n)\Big]\\ &= \frac{1}{|\mathcal{V}|}\sum_{n\in\mathcal{V}}\Big[b_{\ell n} - \sum_{k=1}^{K}\alpha_{\ell k}^{\star}b_{\ell k}\kappa(\mathbf{d}_k,\mathbf{d}_n)\Big]\end{aligned}\tag{25}$$

where $\mathcal{V}$ and $|\mathcal{V}|$ are defined as in (12).

step 11 Repeat *step* 10 for all $\ell \in \{1,\cdots,|\mathcal{L}|\}$

5) Selecting the optimal LEDs combination: Once we get all $\boldsymbol{\alpha}_{\ell}^{\star}, \ell \in \{1,\cdots,|\mathcal{L}|\}$, a LEDs selection learning system may be built according to previous detailed steps. Hence, for any new coming channel matrix, it is manipulated into a real-valued feature vector $\mathbf{d}$ and provided to the learning system. Then, the decision of SVM classification can be obtained as:

$$\text{sign}\Big[\mathbf{w}_{\ell}^{\star\mathrm{T}}\mathbf{d} + \varrho_{\ell}^{\star}\Big] = \text{sign}\left(\sum_{k=1}^{K}\alpha_{\ell k}^{\star}b_{\ell k}\kappa(\mathbf{d}_k,\mathbf{d}) + \varrho_{\ell}^{\star}\right)\tag{26}$$

And consequently, the final result is the label of the prediction class, which corresponds the selected LEDs combination.

IV. SIMULATION RESULTS

In this section, to demonstrate the efficiency of the proposed SVM-aided LEDs selection strategy of the considered GSM-VLC system, we provide numerical results for an indoor environment having the dimensions of $[5 \times 5 \times 3]$ m^3, represented by a three-dimensional (3D) Cartesian coordinate system $[O_X, O_Y, O_Z]$ with the origin being in one corner of the room. The following six system configuration schemes are considered respectively: (1) $N_t = 4, N_r = 4, n_t = 1, M = 4$; (2) $N_t = 5, N_r = 4, n_t = 1, M = 4$; (3) $N_t = 8, N_r = 4, n_t = 1, M = 4$; (4) $N_t = 4, N_r = 4, n_t = 2, M = 4$; (5) $N_t = 5, N_r = 4, n_t = 2, M = 4$; (6) $N_t = 8, N_r = 4, n_t = 2, M = 4$. Again, the transmit LEDs are assumed to be perpendicular to the ceiling and down-facing to the floor. Similarly, the receivers are located on the desks at the height of 0.85 m from the floor, which are assumed to be perpendicular to the desk and facing the ceiling. Unless otherwise specified, we assume that the LED layout is shown in Fig. 4 and the coordinate values are listed in Tab II. The detection method adopted in this paper is maximum likelihood detection [9].

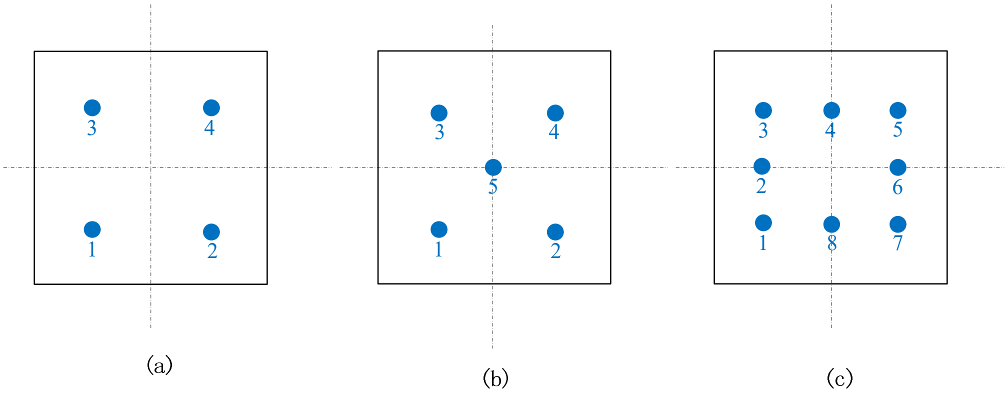

Fig. 4: Indoor LED plane layout, in which the blue solid circle is the location of the LEDs, (a) Position distribution of 4 LEDs at the transmitter; (b) Position distribution of 5 LEDs at the transmitter; (c) Position distribution of 8 LEDs at the transmitter.

The half-illuminance semi-angle $\Phi_{1/2}$ of LED is set to be $60°$, which is a typical value for commercially-available high-brightness LEDs. Receiver has a $75°$ FoV (semi-angle), the area of each PD is $A_{\text{PD}} = 1.0$ cm^2 and the responsivity is $R = 100$ μA/mW/cm^2. For convenience, all the parameters involved in our simulations are summarized in Table III. Furthermore, in order to ensure the prediction accuracy, 4×10^4 CSIs are generated from uniform distributed user's locations, these CISs are utilized as the training set to train the SVM classifier, and the resulting classification accuracy is more than 95%. After the training process is completed, a prediction model can be constructed, and the required activated LEDs are selected online for any new input feature vector by the prediction model.

TABLE II: The coordinates distribution of LEDs with different number as 4, 5 and 8.

4 LEDs		8 LEDs	
1	$(1.25, 1.25, 3.0)$ m	LED	(O_X, O_Y, O_Z)
2	$(3.75, 1.25, 3.0)$ m	1	$(1.25, 1.25, 3.0)$ m
3	$(1.25, 3.75, 3.0)$ m	2	$(1.25, 2.50, 3.0)$ m
4	$(3.75, 3.75, 3.0)$ m	3	$(1.25, 3.75, 3.0)$ m
5 LEDs		4	$(2.50, 3.75, 3.0)$ m
1	$(1.25, 1.25, 3.0)$ m	5	$(3.75, 3.75, 3.0)$ m
2	$(3.75, 1.25, 3.0)$ m	6	$(3.75, 2.50, 3.0)$ m
3	$(1.25, 3.75, 3.0)$ m	7	$(3.75, 1.25, 3.0)$ m
4	$(3.75, 3.75, 3.0)$ m	8	$(2.50, 1.25, 3.0)$ m
5	$(2.50, 2.50, 3.0)$ m		

TABLE III: System parameters in the considered indoor GSM-VLC system.

Simulation setup parameters	
Room size ($\tilde{L} \times \tilde{W} \times \tilde{H}$)	$5 \times 5 \times 3$ m^3
Number of LEDs	$4, 5, 8$
LEDs height	3 m
Receivers height	0.85 m
Transmitter parameters	
Semi-angle at half power ($\Phi_{1/2}$)	60°
Gain of an optical filter (T)	1.0
current-to-light conversion efficiency (η)	813.6 μW/mA
Modulation index ($\tilde{\alpha}$)	0.1
Receivers parameters	
Refractive index (β)	1.5
Physical area of a PD (A_{PD})	1.0 cm^2
Receiver FoV semi-angle (Ψ_{FoV})	75°
PD responsivity (R)	100 μA/mW

A. Algorithm Performance

In this subsection, in order to prove the efficiency of the proposed LEDs selection algorithm in considered GSM-VLC system with other state-of-the-art LEDs selection algorithms, which include EDAS algorithm, capacity optimized antenna selection (COAS) algorithm proposed in [11], [12] and LEDs random selection algorithm, two simulation cases are presented, the first one is to consider SM-VLC system, i.e. $n_t = 1$. Another case is the GSM-VLC system with $n_t \geq 2$. Specifically, for SM-VLC system with $n_t = 1$, Fig.5, Fig.6, Fig.7 show the comparison of the proposed SVM-based LEDs selection algorithm and the other mentioned three LEDs selection algorithms, wherein three system configurations are considered: $N_t = 4, N_r = 4, n_t = 1, M = 4$; $N_t = 5, N_r = 4, n_t = 1, M = 4$ and $N_t = 8, N_r = 4, n_t = 1, M = 4$. The LEDs location distribution is shown in Fig.4.

As can been seen from Fig.5 with $N_t = 4, N_r = 4, n_t = 1, M = 4$, when SNR is less than 30 dB, the BER performance curve of SVM aided LEDs selection algorithm nearly coincides with that of EDAS selection algorithm, and it is superior to the LEDs random selection algorithm and COAS algorithm. When SNR is more than 30 dB, compared with EDAS algorithm, the SVM aided LEDs selection algorithm has a performance loss less than 1 dB when BER is 10^{-5}. This is because by using the SVM aided learning system for LEDs selection, the prediction has an error less than 2%. Furthermore, from Fig. 6, Fig.7 we can observe that with the increase of N_t, compared with the LEDs random selection and COAS algorithms, the performance advantage of SVM and EDAS algorithm is obvious. In addition, the performance difference between SVM and EDAS algorithm increases to nearly 2 dB when BER is 10^{-5} because of the prediction accuracy declining slightly and within 5%. However, by utilizing the SVM aided LEDs selection algorithm, there is no need for a large number of redundant calculations in the actual LED selection process, which makes the computational complexity much lower than that of EDAS algorithm. Although COAS algorithm can also reduce the complexity of EDAS algorithm, its BER performance is not as good as the SVM aided LEDs selection algorithm. The LEDs random selection algorithm has the lowest computational complexity, but its BER performance is poor, which makes it cannot be adopted in practical applications. Therefore, considering the performance of all presented algorithms comprehensively, the SVM aided LEDs selection algorithm is superior to EDAS, COAS and LEDs random selection algorithms. It should be noted that, from Fig. 5, Fig. 6 and Fig.7, we can conclude that the spatial position arrangement of LEDs can affect the performance of the considered GSM-VLC system, but it is out of the topic of this paper and we will not discuss it extensively.

Then, in order to further verify the performance of the proposed SVM aided LEDs selection algorithm for indoor GSM-VLC system, three system configurations are established as $N_t = 4, N_r = 4, n_t = 2, M = 4$; $N_t = 5, N_r = 4, n_t = 2, M = 4$ and $N_t = 8, N_r = 4, n_t = 2, M = 4$. Based on these system configurations, Fig. 8, Fig. 9 and Fig.10 demonstrate the comparisons of the proposed SVM aided LEDs selection algorithm and three other LEDs selection algorithms: EDAS, COAS and LEDs random selection algorithm. The LEDs location distribution is shown in Fig. 4. As can be seen from Fig. 8, when SNR is less than 44 dB, the BER performance curve of SVM aided LEDs selection algorithm coincides with that of EDAS selection algorithm, and they are superior to the LEDs random selection algorithm and COAS algorithm. When SNR is more than 48 dB, compared with EDAS algorithm, SVM aided LEDs selection algorithm has a performance loss of less than 1 dB when BER is

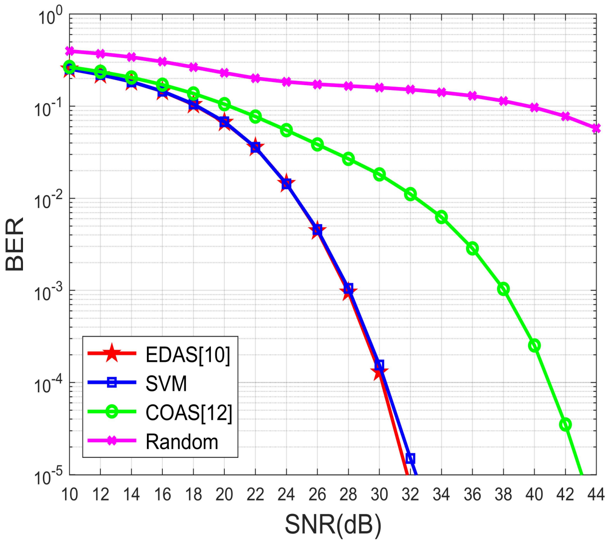

Fig. 5: Comparisons of BER performance for the proposed SVM aided LEDs selection algorithm and other three LEDs selection algorithms in SM-VLC system with system configuration as $N_t = 4, N_r = 4, n_t = 1, |\mathbb{M}| = 4$.

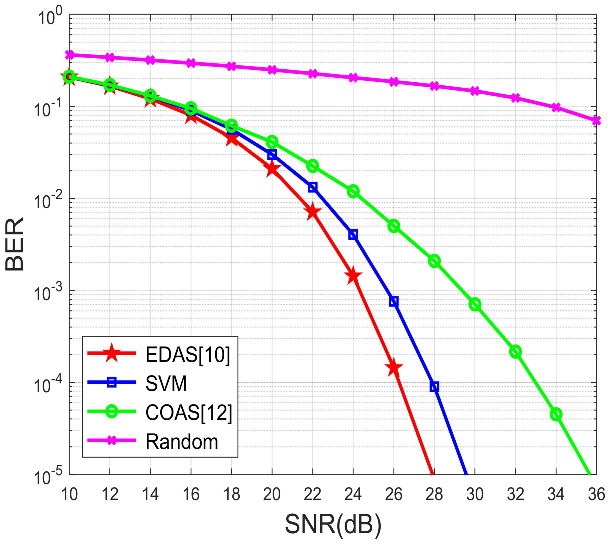

Fig. 6: Comparisons of BER performance for the proposed SVM aided LEDs selection algorithm and other three LEDs selection algorithms in SM-VLC system with system configuration as $N_t = 5, N_r = 4, n_t = 1, |\mathbb{M}| = 4$.

10^{-5}. Additionally, as depicted in Fig. 9 and Fig.10 for other two system configurations with more LEDs, when BER is 10^{-5}, the performance difference can also be limited within 1.5 dB. Hence, for the GSM-VLC systems considering the BER performance and computational complexity, based on the above results, we can conclude that the SVM aided LEDs selection

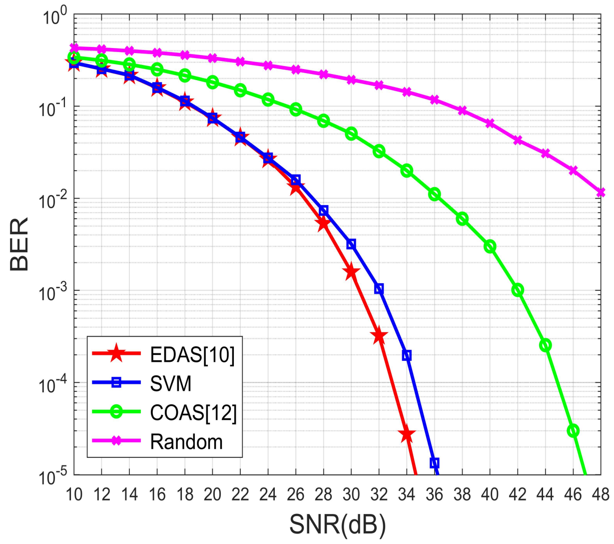

Fig. 7: Comparisons of BER performance for the proposed SVM aided LEDs selection algorithm and other three LEDs selection algorithms in SM-VLC system with system configuration as $N_t = 8, N_r = 4, n_t = 1, |\mathbb{M}| = 4$.

algorithm is superior to conventional EDAS, COAS and LEDs random selection algorithms.

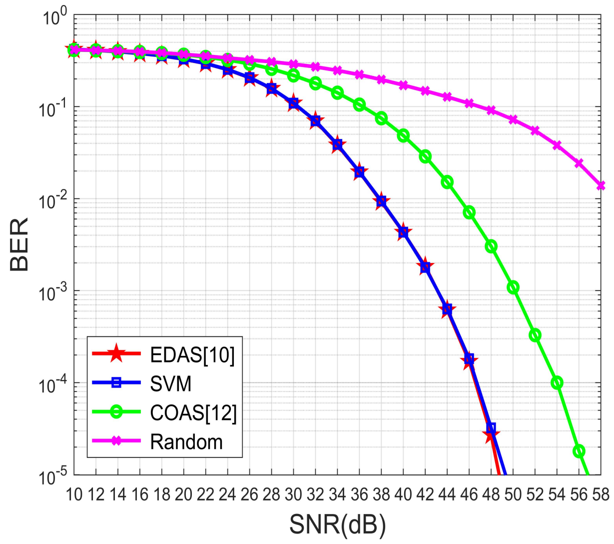

Fig. 8: Comparisons of BER performance for the proposed SVM aided LEDs selection algorithm and other three LEDs selection algorithms in GSM-VLC system with system configuration as $N_t = 4, N_r = 4, n_t = 2, |\mathbb{M}| = 4$.

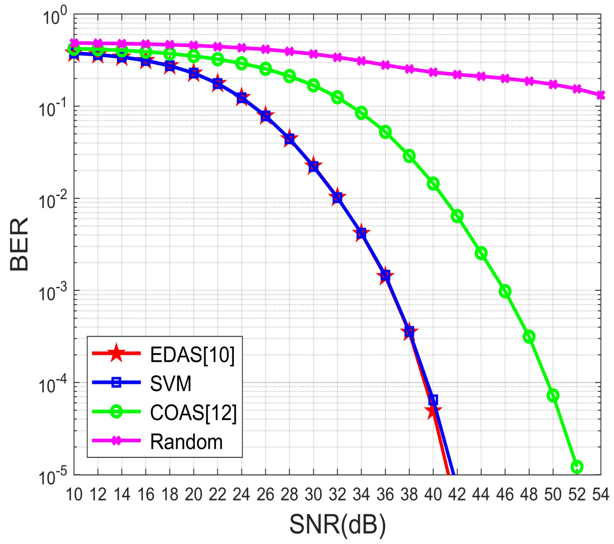

Fig. 9: Comparisons of BER performance for the proposed SVM aided LEDs selection algorithm and other three LEDs selection algorithms in GSM-VLC system with system configuration as $N_t = 5, N_r = 4, n_t = 2, |\mathbb{M}| = 4$.

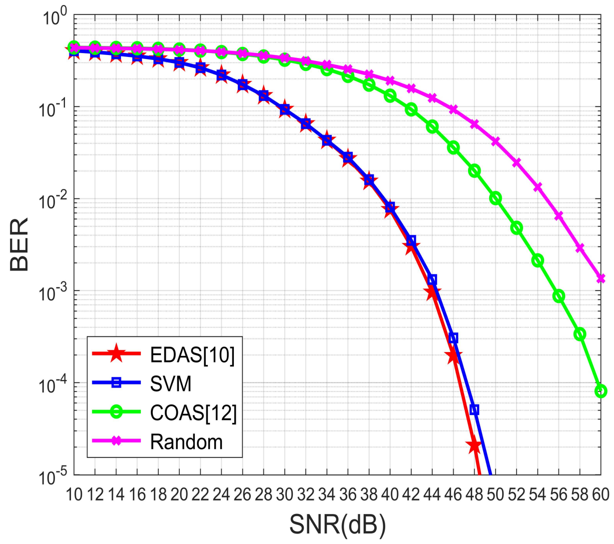

Fig. 10: Comparisons of BER performance for the proposed SVM aided LEDs selection algorithm and other three LEDs selection algorithms in GSM-VLC system with system configuration as $N_t = 8, N_r = 4, n_t = 2, |\mathbb{M}| = 4$.

B. Evaluation and comparison of complexity

In this subsection, we will compare the computational complexity of the proposed SVM aided LEDs selection with conventional LED selection algorithms EDAS, COAS and LEDs random selection algorithm. The complexity of the EDAS algorithm is obtained as follows. At first, the operation of taking 2 from $n_t|\mathbb{M}|$ requires a complexity of $\mathcal{O}(n_t^2|\mathbb{M}|^2)$. After the sub-channel matrix $\mathbf{H}_{\mathcal{I}}$ is multiplied, a vector is obtained with dimensional $N_r \times 1$, the Euclidean distance is calculated and the complexity is N_r. Then, performing heapsort on $|\mathcal{I}|$, the complexity is $|\mathcal{I}|\log|\mathcal{I}|$. Hence, the complexity of EDAS approach is $\mathcal{O}(n_t^2 N_r|\mathcal{I}||\mathbb{M}|^2\log|\mathcal{I}|)$. For COAS algorithm, the complexity of calculating the Euclidean distance of each column in the channel matrix is $N_t \times N_r$, and then perform heapsort on the N_t column vectors. Hence, the complexity is $\mathcal{O}(N_t^2 N_r \log N_t)$. Finally, for the proposed SVM algorithm, the only operation is to normalize the channel matrix, hence, the complexity is $\mathcal{O}(N_t N_r)$.

The complexities of different LEDs algorithms are presented and compared in Tab. IV. It can be observed in Tab. IV that the complexity of the proposed SVM aided LEDs selection algorithm is polynomial on N_t, which is lower than that of the conventional EDAS and COAS algorithms, especially when the number of LEDs N_t and its combinations $\mathcal{I}$ are very large. It should be noted that the computational complexity of the proposed SVM aided LEDs selection algorithm is defined as the online prediction computational complexity, the time consuming of the offline training process is not considered in complexity analysis of our proposed LEDs selection algorithm. This consideration lies that the proposed LEDs selection approach can be normally utilized before system parameters are modified once the model are trained by the offline training procedure. Hence, as other SVM aided antenna selection algorithms in [13], [15], for online prediction applications, only online prediction computational complexity is considered in our analysis.

TABLE IV: Computation complexity comparison of different LEDs selection algorithms.

Algorithms	**Complexity**
EDAS algorithm [10]	$\mathcal{O}(n_t^2 N_r|\mathcal{I}||\mathbb{M}|^2\log|\mathcal{I}|)$
COAS algorithm [12]	$\mathcal{O}(N_t^2 N_r \log N_t)$
Proposed SVM aided algorithm	$\mathcal{O}(N_t N_r)$
LEDs random selection algorithm	$\mathcal{O}(1)$

In Fig. 11, Fig. 12 and Fig. 13, the number of executions versus N_t and N_r are depicted of EDAS [10], COAS [12] and proposed SVD aided LEDs selection algorithms with system configuration $n_t = 2, |\mathbb{M}| = 4$, respectively. It can be observed that the number of executions of the proposed SVM aided LEDs selection algorithm is much lower than EDAS and COAS algorithms. Meanwhile, with the increasing of the number of N_t and N_r, the computational complexity increments of conventional EDAS and COAS algorithms are more remarkable than SVM aided LEDs selection algorithm. Additionally, even though random LEDs selection algorithm has the lowest complexity as presented in Tab. IV, this is achieved at the cost of a serious degraded BER performance as demonstrated in figures of the last subsection, which makes it unapplicable in practice.

Based on the above simulation results in this subsection, despite the proposed SVM aided LEDs selection algorithm and conventional optimal EDAS algorithm have very similar BER performance, the proposed SVM aided algorithm can efficiently eliminate redundant computation during the LEDs selection procedure, thus can greatly improve the system efficiency and achieve online LEDs selection task.

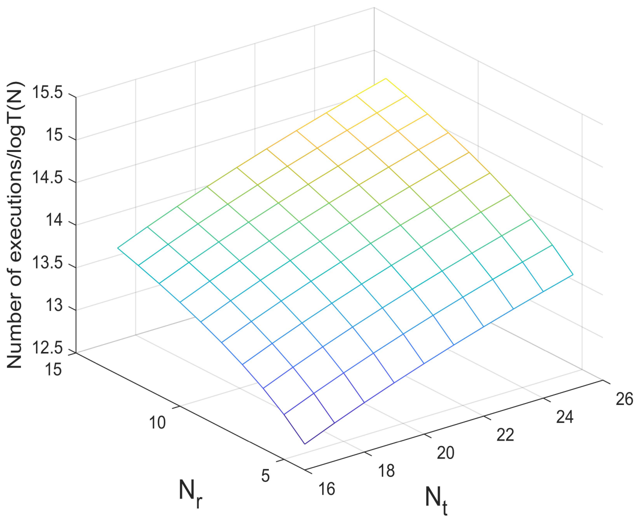

Fig. 11: Number of executions of EDAS versus N_t and N_r with configuration $n_t = 2$,$|\mathbb{M}| = 4$, T(N)=$\mathcal{O}(n_t^2 N_r |\mathcal{I}||\mathbb{M}|^2 \log |\mathcal{I}|)$)

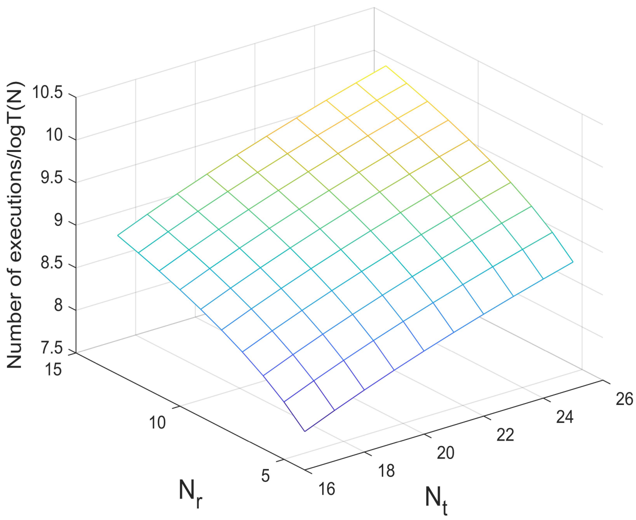

Fig. 12: Number of executions of COAS versus N_t and N_r with configuration $n_t = 2$,$|\mathbb{M}| = 4$, T(N)=$\mathcal{O}(N_t^2 N_r \log N_t)$

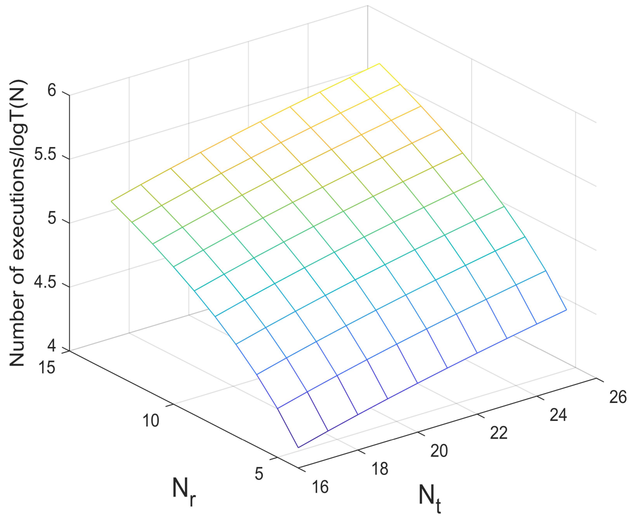

Fig. 13: Number of executions of SVM versus N_t and N_r with configuration $n_t = 2$,$|\mathbb{M}| = 4$, T(N)=$\mathcal{O}(N_t N_r)$

V. Conclusion

Aiming at the characteristics of LoS channel of indoor GSM-VLC system, this paper proposes a low complexity and high efficiency LEDs selection algorithm based on SVM of indoor GSM-VLC system, which is obtained by equivalent modeling the LEDs selection problem as a multiple classification machine learning task. Firstly, the training sample set is constructed and the feature matrix is obtained by randomly generating independent identically Uniform distributed user positions, and the minimum Euclidean distance is taken as the KPI of the training system to derive the label vector of the training samples. Then the kernel SVM is utilized to establish the optimization problem of LEDs selection, and the dual quadratic convex programming problem of the original one is obtained through dual theory, so as to obtain the optimal classification parameters of SVM efficiently. Finally, the online antenna selection for any given user channel information is realized by the learned optimal classification parameters. Through computer simulation and complexity analysis, compared with traditional LEDs selection algorithms, the algorithm proposed in this paper can realize online LEDs selection while maintain ideal BER performance, which shows the effectiveness of the proposed SVM aided algorithm.

Acknowledgment

This research was supported in part by the National Natural Science Foundation of China under Grant 61571401, U1736107 and 61901366, in part by the National Natural Science Foundation of Henan Province under Grant 192102210088 and 192102210091, in part by the Innovative Talent of Colleges and University of Henan Province under Grant 18HASTIT021, and in part by the National Key Research and Development Program under Grant 2019QY0302.

References

[1] M. Chowdhury, M. Hasan, M. Shahjalal, M. Hossan and Y. Jang, "Optical Wireless Hybrid Networks: Trends, Opportunities, Challenges, and Research Directions," IEEE Communications Surveys and Tutorials, vol. 22, no. 2, pp. 930-966, Secondquarter 2020.

[2] R. Zhang, H. Claussen, H. Haas and L. Hanzo, "Energy Efficient Visible Light Communications Relying on Amorphous Cells," IEEE Journal on Selected Areas in Communications, vol. 34, no. 4, pp. 894-906, April 2016.

[3] T. Koonen, "Indoor Optical Wireless Systems: Technology, Trends, and Applications," Journal of Lightwave Technology, vol. 36, no. 8, pp. 1459-1467, April 2018.

[4] F. Wang, C. Liu, Q. Wang, J. Zhang, R. Zhang, L. Yang and L. Hanzo, "Secrecy Analysis of Generalized Space-Shift Keying Aided Visible Light Communication," IEEE Access, vol. 6, pp. 18310-18324, January 2018.

[5] S. Wu, H. Wang and C. Youn, "Visible light communications for 5G wireless networking systems: from fixed to mobile communications," IEEE Network, vol. 28, no. 6, pp. 41-45, November 2014.

[6] Y. Zou, J. Zhu, X. Wang and L. Hanzo, "A Survey on Wireless Security: Technical Challenges, Recent Advances, and Future Trends," Proceedings of the IEEE, vol. 104, no. 9, pp. 1727-1765, September 2016.

[7] J. Wang, S. Jia and J. Song, "Generalised Spatial Modulation System with Multiple Active Transmit Antennas and Low Complexity Detection Scheme," IEEE Transactions on Wireless Communications, vol. 11, no. 4, pp. 1605-1615, April 2012.

[8] M. Renzo, H. Haas, A. Ghrayeb, S. Sugiura and L. Hanzo, "Spatial Modulation for Generalized MIMO: Challenges, Opportunities, and Implementation," Proceedings of the IEEE, vol. 102, no. 1, pp. 56-103, January 2014.

[9] S. Alaka, T. Narasimhan and A. Chockalingam, "Generalized Spatial Modulation in Indoor Wireless Visible Light Communication," 2015 IEEE Global Communications Conference (GLOBECOM), San Diego, 2015, pp.1-7.

[10] R. Rajashekar, K. Hari and L. Hanzo, "Antenna Selection in Spatial Modulation Systems," IEEE Communications Letters, vol. 17, no. 3, pp. 521-524, March 2013.

[11] Z. Zhou, N. Ge and X. Lin, "Reduced-Complexity Antenna Selection Schemes in Spatial Modulation," IEEE Communications Letters, vol. 18, no. 1, pp. 14-17, January 2014.

[12] N. Pillay and H. Xu, "Low Complexity Transmit Antenna Selection for Spatial Modulation," IET Communications, vol. 9, no. 2, pp. 239-248, June 2015.

[13] J. Joung, "Machine Learning-Based Antenna Selection in Wireless Communications," IEEE Communications Letters, vol. 20, no. 11, pp. 2241-2244, November 2016.

[14] Y. Hu, L. Li, J. Yin, H. Zhang, W. Liang, A. Gao and Z. Han, "Optimal Transmit Antenna Selection Strategy for MIMO Wiretap Channel Based on Deep Reinforcement Learning," 2018 IEEE/CIC International Conference on Communications in China (ICCC), Beijing, China, 2018, pp. 803-807.

[15] P. Yang, J. Zhu, Y. Xiao and Z. Chen, "Antenna selection for MIMO system based on pattern recognition," Digital Communications and Networks, vol. 5, pp. 34-39, October 2018.

[16] Z. Wang, C. Guo, Y. Yang and Q. Li, "Antenna selection based dimming scheme for indoor MIMO visible light communication systems utilizing multiple lamps," 2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Valencia, 2016, pp. 1-7.

[17] Z. Feng, C. Guo, Y. Yang, Z. Ghassemlooy and C. Feng, "MMSE Based Spatial Dimming Scheme for Multiuser MISO VLC Systems," 2017 IEEE Globecom Workshops (GC Wkshps), Singapore, 2017, pp. 1-6.

[18] M. Tran, S. Kim, T. Ketseoglou and E. Ayanoglu, "LED Selection and MAP Detection for Generalized LED Index Modulation," IEEE Photonics Technology Letters, vol. 30, no. 19, pp. 1695-1698, October 2018.

[19] M. Keskin, A. Sezer and S. Gezici, "Localization via Visible Light Systems," Proceedings of the IEEE, vol. 106, no. 6, pp. 1063-1088, June 2018.

[20] H. Wang, F. Wang, R. Li, "Enhancing power allocation efficiency of NOMA aided-MIMO downlink VLC networks," Optics Communications, vol. 454, pp. 124497, September 2019.

[21] F. Wang, C. Liu, Q. Wang, J. Zhang, R. Zhang, L.L. Yang and L. Hanzo, "Optical jamming enhances the secrecy performance of the generalized space shift keying aided visible light downlink," IEEE Transactions on Communications, vol. 66, no. 9, pp. 4087 – 4102, September 2018.

[22] A. Mostafa and L. Lampe, "Optimal and Robust Beamforming for Secure Transmission in MISO Visible-Light Communication Links," IEEE Transactions on Signal Processing, vol. 64, no. 24, pp. 6501-6516, December 2016.

[23] M. Jha, N. Kumar and Y. Lakshmi, "Generalized Spatial Modulation for Multi-User in Visible Light Communication," 2019 28th Wireless and Optical Communications Conference (WOCC), Beijing, China, 2019, pp. 1-5.

[24] L. Zeng, et al., "High data rate multiple input multiple output (MIMO) optical wireless communications using white led lighting," IEEE Journal on Selected Areas in Communications, vol. 27, no. 9, pp. 1654-1662, December 2009.

[25] V. Chan, "Free-Space Optical Communications," Journal of Lightwave Technology, vol. 24, no. 12, pp. 4750-4762, December 2006.

[26] P. Luo, et al, "SVM-based detection in visible light communications," Optik, vol. 151, pp. 55-64, August 2017.

[27] D. Sebald and J. Bucklew, "Support Vector Machine Techniques for Non Linear Equalization," IEEE Transactions on Signal processing, vol.48, no. 11, pp.3217-3226, November 2000.

[28] S. Alam, M. Kang, J. Pyun and G. Kwon, "Performance of classification based on PCA, linear SVM, and Multi-kernel SVM," 2016 Eighth International Conference on Ubiquitous and Future Networks (ICUFN), Vienna, 2016, pp. 987-989.

[29] S. Theodoridis, *Machine Learning: A Bayesian and Optimization Perspective*. Cambridge, MA, USA: Academic Press, 2015.