Enhancing Urban Traffic Management with Visible Light Communication and Reinforcement Learning

Gonçalo Galvão[1], *Manuel Augusto* Vieira[1,2], *Manuela* Vieira[1,2*], *Paula* Louro[1,2], *Mário* Véstias[1,3]
[1]Electronics Telecommunications and Computer Dept. ISEL/IPL, R. Conselheiro Emídio Navarro, 1949-014 Lisboa, Portugal.
[2] CTS-UNINOVA and LASI, Monte da Caparica, 2829-516, Caparica, Portugal
[3] INESC INOV-Lab, Lisboa, Portugal
* manuela.vieira@isel.pt

Abstract. This paper introduces Visible Light Communication (VLC) to enhance traffic signal efficiency and vehicle trajectory management at urban intersections. A multi-intersection traffic control system is proposed, integrating VLC localization services with learning-based traffic signal control. VLC facilitates communication between connected vehicles and infrastructure using headlights, streetlights, and traffic signals to transmit information. By leveraging vehicle-to-vehicle (V2V) and infrastructure-to-vehicle (I2V) interactions, joint transmission and data collection are achieved via mobile optical receivers. The system aims to reduce waiting times for pedestrians and vehicles while improving overall traffic safety. It is designed to be flexible and adaptive, accommodating diverse traffic movements during multiple signal phases. VLC cooperative mechanisms, transmission range, relative pose concepts, and queue/request/response interactions help balance traffic flow between intersections, enhancing the overall road network performance.

1 Introduction

Smart cities are evolving to be more comfortable, fast, and safe for travel. Technological advancements have equipped even non-autonomous vehicles with sophisticated sensors and computers, significantly improving road safety [1] [2]. However, accidents and traffic congestion still persist, causing long waiting and travel times. Additionally, the overload of radio frequencies poses a challenge for stable communication technology. Visible Light Communication (VLC), using LEDs and photodetectors, offers a solution by modulating light for communication, complementing RF communications in traffic scenarios.

Leveraging Intelligent Transport System (ITS) [3] technology and VLC, road safety and traffic management can be optimized through communication between vehicles (V2V), vehicles and infrastructure (V2I), infrastructure and vehicles (I2V), and pedestrians [4]. While VLC physical transmission has been well-researched, there is a gap in its application for vehicular and pedestrian traffic control, which this paper aims to address. The proposed traffic control system model with VLC targets both vehicles and pedestrians, addressing the need for adaptive traffic control that responds in real-time to traffic demands. VLC provides detailed information on vehicle position and speed, enabling better decision-making for traffic management.

The system operates on a Reinforcement Learning (RL) method, specifically using Deep Q-Learning algorithms. By collecting and analyzing reward samples, the algorithm learns to optimize traffic control decisions over time [5].

The paper is organized as follows: Section II describes the VLC system and traffic environment, presenting simulated results of VLC communication between vehicles and pedestrians. Section III details the intelligent traffic system and agent learning method using Deep Q-Learning. Section IV compares the intelligent system with a dynamic traffic control system, highlighting the improvements achieved.

2 V-VLC SYSTEM

2.1 V-VLC Communication Link

The Vehicular VLC system (V-VLC) comprises a transmitter that generates modulated light and a receiver located in infrastructures, driving to detect the received light variation, as illustrated in Fig. **1**. Both the transmitter and receiver are connected through the wireless channel. In this system, the light produced by the LED is modulated using ON-OFF-keying (OOK) amplitude modulation. The environment is defined by a cluster of square unit cells arranged in an orthogonal geometry. Different data channels are provided by tetra-

chromatic white light (WLEDs) sources positioned at the corners of the square unit cells distributed along the road and at the crossroads. The white WLEDs sources consist of Red (R: 626 nm), the Green (G: 530 nm), the Blue (B: 470 nm) or the Violet (V: 390 nm) chips and combine the lights in correct proportion to generate white light [6] [7].

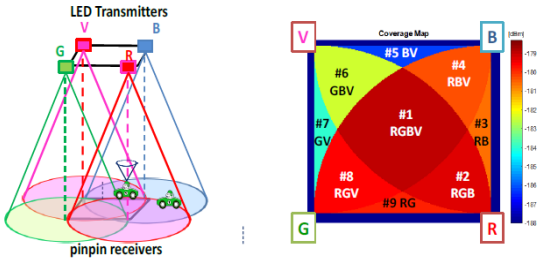

Fig. 1 - Illustration of the coverage map in the unit cell: footprint regions (#1 #9).

Each of the RGBV signals sent has a wavelength calibrated amplitude that defines it. Since each VLC infrastructure has four independent emitters, the optical signal generated in the receiver can have one, two, three, or even four optical excitations, resulting in 2^4 different optical combinations and 16 different photocurrent levels at the photodetector. Filtering is accomplished using a PIN-PIN demultiplexer [8]. The PIN-PIN demultiplexer plays a crucial role in the decoding process, ensuring accurate retrieval of the original message. It receives the combined OOK signals and armed with prior knowledge of the calibrated amplitudes, decodes the sent message.

The communication protocol defines the structure and rules governing the exchange of information. It includes specifications for the synchronization, identification, and payload portions of the transmitted frame. The communication protocol is presented in Table 1.

Table 1 - Message protocol defined for each of the V-VLC communications.

		COM	Position												
L2V	Sync	1	x	y	END	Hour	Min	Sec	Payload (32 bits)						EOF
V2V	Sync	2	x	y	Lane (0–7)	N° Veic.	END	Hour	Min	Sec	Car IDx	Car IDy	n° behind		EOF
V2I	Sync	3	x	y	TL (0–15)	N° Veic.	END	Hour	Min	Sec	Car IDx	Car IDy	n° behind		EOF
I2V	Sync	4	x	y	TL (0–15)	ID veic	END	Hour	Min	Sec	Car IDx	Car IDy	n° behind	Phase	EOF
P2I	Sync	5	x	y	TL (0–15)	Direct.	END	Hour	Min	Sec	payload				EOF
I2P	Sync	6	x	y	TL (0–15)	Phase	END	Hour	Min	Sec	Payload				EOF

Starting the analysis of the communication protocol by examining the frame structure it begins with a synchronization block (Sync) of five bits, indicated by the pattern [10101]. This is used to synchronize the receivers and identify the start of a new frame. Advancing to the identification blocks (ID), these includes the type of communication (COM), localization of transmitters (x, y coordinates), and timeline information (END, Hour, Min, Sec). The time sub-block begins with the pattern [111] to alert the decoder that the following bit sequence (6+6+6) corresponds to time identification rather than payload. Other ID blocks include the necessary number and temporary identification of vehicles following the leader. Include also information related to the occupied lane (Lane 0-7), traffic light (TL) signal requested (TL 0-15), cardinal direction, or active phase provided by the infrastructure in a "response" or "request" message at the intersection. In relation to the traffic messages, this block includes vehicle information such as x, y coordinates and order of cars behind the leader that request/receive permission to cross the intersection (CarIDx, CarIDy, nº behind). It also includes traffic information (payload) such as road conditions, average waiting time or weather conditions. The frame concludes with a 4-bit EoF block, defined by the pattern [0000], indicating the end of the frame.

2.2 Traffic Environment

The traffic scenario studied in this paper is shown in Fig. 2. It features two four-arm connected intersections with two lanes per arm, totaling 160 meters in length, designed for vehicles and pedestrians. Right lanes are designated for straight or right turns, while left lanes are for left turns only. Each lane contains footprint cells (1-9) for vehicle detection. The positions of the streetlights (R, G, B, V), which are modulated for communication, are also displayed. These streetlights serve as geo-transmitters and are strategically positioned 20 meters apart along the roadside. Each LED transmitter emits an I2V message, including synchronization, physical ID, and traffic information. When a vehicle or pedestrian enters the capture range of a streetlight, the receiver assigns a unique ID (x, y, t) and provides relevant traffic information.

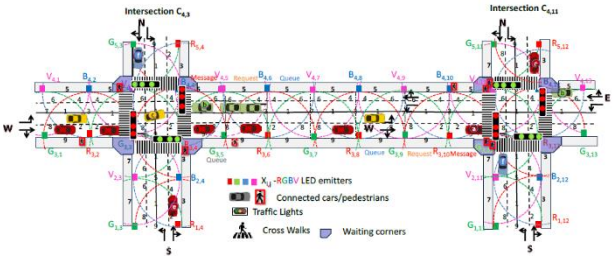

Fig. 2 - Traffic scenario considered with the respective footprints (1-9), connected vehicles and pedestrians.

2.3 V-VLC Experimental Results

Laboratory tests [9] were conducted in a controlled environment to simulate various communications between vehicles and pedestrians. The transmitted information must be decoded from the measured photocurrent signal by the receiver, requiring a predefined calibration curve. This bit sequence allows for all sixteen possible on/off combinations of the four optical channels. Each of the 16 distinct photocurrent levels on the calibration curve corresponds to one of the 24 possible RGBV channel combinations.

Fig. 3 illustrates vehicle positions, the communication protocol in Table 1, and the technique for decoding calibrated signals emitted by transmitters. In Fig. 3a, decoded optical signals and received signals (MUX) are displayed for a V2V (COM 2) and V2I

(COM 3) communication scenario involving a leader vehicle at position ($R_{3,10}$, $G_{3,11}$, $B_{4,10}$). This vehicle communicates with the IM at the second intersection (C2) on lane L0 (direction E) at 10:25:46 and is followed by three vehicles (V_1, V_2, and V_3) at positions ($R_{3,8}$, $G_{3,6}$, and $R_{3,4}$).

Fig. 3b exemplifies responses (I2V and I2P) from two traffic lights (TL10 and TL13) to a crossing request from vehicle ($R_{3,10}$, $G_{3,11}$, $B_{4,10}$) and a pedestrian located in the "waiting corner" of the first intersection ($R_{3,4}$, $G_{3,5}$). Timestamps "10:25:46" and "10:28:66" indicate when each response was generated.

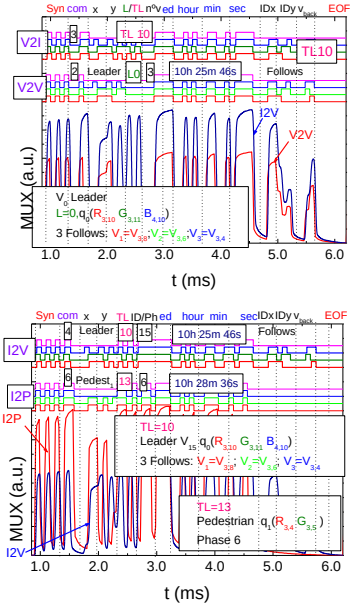

Fig. 3 - MUX signal request (a) and responses (b) assigned to different types of V-VLC communication.

Fig. 4 a illustrates the MUX signal sent to the traffic lights (TL's) by pedestrians to cross both intersections (C1 and C2) while waiting in the corners ($P_{1,2}$2I). In this figure, the top part displays the decoded messages, and on the right-hand side, the content of the message is outlined. Furthermore, Fig. 4 b demonstrates the MUX signal received by the traffic lights ($I2P_{1,2}$).

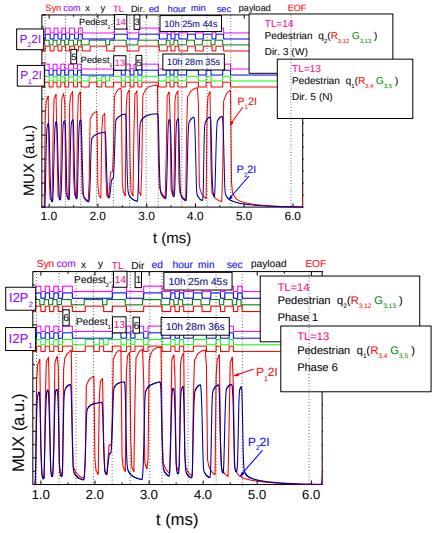

Fig. 4 - Normalized MUX signal responses and the corresponding decoded messages, sent and received by pedestrians.

The results reveal that the pedestrian begins walking on the sidewalk lane towards the west (W), intending to cross at C2, and waits at positions $R_{3,12}$-$G_{3,13}$. At 10:25:44, the pedestrian initiates communication with the traffic light ($P_2$2I), and by 10:25:45, receives a response (I2P2). The pedestrian remains in the waiting zone until the pedestrian phase becomes active. The traffic light indicates that the current active phase is N-S (Phase 1), so the pedestrian must wait for 120 seconds before crossing. The pedestrian then crosses the crosswalk, reaching the next intersection in approximately 1 minute and 50 seconds. Upon arrival, the pedestrian waits in the designated zone at $R_{3,4}$-$G_{3,5}$ until the pedestrian phase is active again. At 10:28:35, the pedestrian communicates with the traffic light at C1 ($P_1$2I), and the traffic light responds at 10:28:36 ($I2P_1$), indicating that the current active phase is the final one in the cycle (Phase 6). These interactions demonstrate the effectiveness of the communication system, allowing the pedestrian to stay informed about the active phase, waiting time, and make timely decisions accordingly.

3 Intelligent Traffic control system

With the data collected on vehicles via VLC through the cells in Fig. 1, implemented via lamps along the roads as shown in Fig. 2, an intelligent traffic system must be developed to optimize traffic flow at intersections. This system utilizes reinforcement learning (RL), a machine learning paradigm where an agent learns to make decisions by interacting with its environment. Agents in RL aim to achieve a goal in uncertain, potentially complex environments by receiving feedback in the form of rewards or punishments. The fundamental idea is for the agent to learn optimal behavior or strategies through trial and error.

At each time step t, the agent receives a state input s_t, based on the observation of the environment and then executes an action a_t, that transforms the state observed to a next state s_{t+1}. Then the reward r_t, a metric that defines how good the action was for the environment, is calculated. In this case, the reward is defined by the equation (1), using the accumulated total waiting time, atw_t, as a metric for vehicles (*veh*) and pedestrians (*ped*). $atwt_t$ and $atwt_{t-1}$ are the accumulated total waiting time of all the cars/pedestrians in the intersection captured respectively at agentstep t and agentstep $t-1$. The weights of the p_{veh} and p_{ped} are set based on the desired priority that the agent should have towards vehicles and pedestrians during network training. The agent will learn a policy that benefits one more than the other, or keeps the system balanced if the weights are equal.

If the agent's behavior leads to positive environmental reward, which indicates that the waiting time is longer in the past, *t-1*, than at the present moment, *t*, then the tendency of producing this behavior by the agent will be strengthened, and vice versa. The goal is to maximize the cumulative discounted reward, r_t , as described in Equation 1.

$$r_t = p_{veh}(atwt_{veh,t-1} - atwt_{veh,t}) + p_{ped}(atwt_{ped,t-1} - atwt_{ped,t}) \quad (1)$$

This experience $e_x = (s_t, a_t, r_t, s_{t+1})$ will be stored in the replay memory, to be used in the future to train the agent. The replay memory is a dataset of an agent's experiences $D_t = (e_1, e_2, \dots, e_t)$, which are gathered when the agent interact with the environment as time goes by (t = 1, 2…, n).

To train the agent, the deep Q-Learning technique is employed, leveraging the Q-Learning algorithm [10] [11] [12] [13]. The Q-value represents the expected cumulative reward of taking a particular action in a particular state and following the optimal policy thereafter. These Q-values are predicted by a neural network (NN) that takes the state as input and outputs Q-values for each possible action, as illustrated in Fig.5.

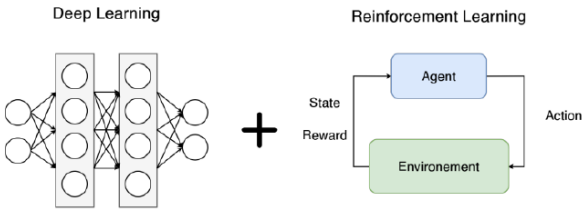

Fig. 5 - Deep Reinforcement Learning.

The Q-value represents the expected cumulative reward of taking a specific action in each state while following the optimal policy thereafter.

A neural network (NN) predicts these Q-values by taking the state as input and outputting Q-values for each possible action. The state of the environment comprises 100 cells at each intersection, indicating the presence of vehicles or pedestrians. These cells are set to '1' if occupied and '0' if not. Each lane, divided into 10 cells, indicates vehicle movement toward the intersection, with cell sizes increasing farther from the intersection. With 8 lanes per junction, there are 80 vehicle cells per intersection. For pedestrians, only the waiting zones are considered, each divided into 5 cells, totaling 20 pedestrian cells per intersection as draft in Fig. 5.

The neural network's input layer consists of 100 neurons representing the state of the environment. This is followed by five hidden layers, each with 400 neurons using rectified linear units (ReLUs). The output layer features nine neurons, each representing the Q-values for potential actions. To refine Q-value predictions, a Mean Squared Error (MSE) function quantifies the disparity between predicted and target Q-values, enhancing the learning process. N represents the number of samples stored in memory, and Qtarget and Qpred denote the target and predicted values, respectively. After each training episode, target Q-values for action-state pairs are calculated based on equation 2.

$$\mathrm{MSE_{Loss}} = \frac{1}{N}\sum_{i=1}^{N}\left(Q_{target} - Q_{pred}\right)^2 \quad (2)$$

N is the number of samples stored in memory, and the target and predicted value, Q_{target} and Q_{pred}, respectively. After each episode of training, the target Q-values for action-state pairs are calculated based on equation 3.

$$Q_{target} = r_t + \gamma . max Q_{pred}(s_{t+1}, a') \quad (3)$$

Where a' is the action and γ is a discount factor that diminishes the reward over time.

The nine Q-values at the neural network's output correspond to the nine actions shown in Fig. 6. The agent selects the action that best suits the current traffic situation, without following a predefined order. Conversely, today's dynamic traffic systems at junctions follow a fixed sequence of phases, as shown in Fig. 6. This can result in activating a phase that does not align with current traffic needs. The next section compares these two systems to highlight their differences and evaluate their effectiveness.

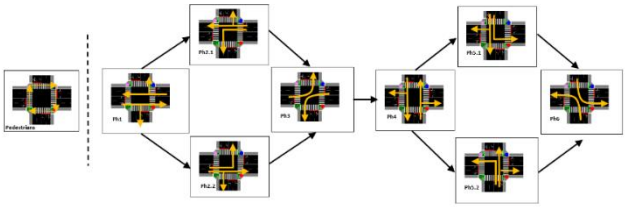

Fig. 6 - Nine possible actions that can be chosen by the agent.

4 Simulation results

4.1 Training Results

To evaluate the behaviour of the intelligent traffic control system in relation to pedestrian and vehicle scenarios, a comparison was made with the dynamic traffic control system. The neural network used was trained with a reward system that weighted the waiting times for vehicles (p_{veh}) and pedestrians (p_{ped}) equally, for 300 epochs, each lasting one hour. Both systems considered the same generation rates for pedestrians and vehicles, totalling 2300 vehicles and 11000 pedestrians in the traffic scenario.

Fig. 7 - Cumulative Negative reward for both agents in training.

Fig. 7 shows the cumulative negative reward from training the network for both agents. Both agents evolved and learned from their traffic experiences throughout the episodes. The curves converged towards less negative reward values, indicating better decision-making over time.

After training the network, tests were conducted to compare both systems under two traffic scenarios representing peak hour conditions. The first scenario involved high vehicle and pedestrian traffic (High-High)

with 2300 cars and 11000 pedestrians. The second scenario had high vehicle traffic but low pedestrian traffic (High-Low), with 2300 cars and 5600 pedestrians.

4.2 Testing results – High-High and High-Low scenarios

Figures 8a and 8b display the number of pedestrians waiting in zones at the two junctions for both traffic scenarios. In Figure 8a, which represents the high vehicle and pedestrian scenario, the intelligent system significantly outperforms the dynamic system. The dynamic system peaks at around 1500 waiting pedestrians in the first 25 minutes, while the intelligent system peaks at just 400 pedestrians.

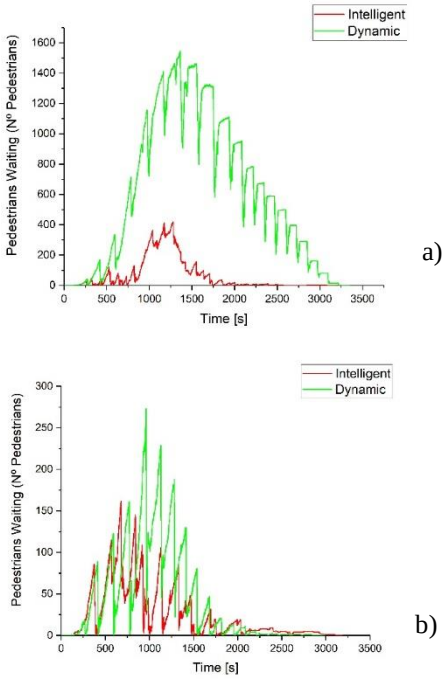

Fig. 8 - Comparison of the number of pedestrians stopped waiting in both systems for the High-High (a) and High-Low (b) scenarios.

Fig. 8b shows a smaller difference in the high vehicle and low pedestrian scenario, where the dynamic system peaks at 275 pedestrians in the first 15 minutes, compared to 150 for the intelligent system. This disparity arises because the intelligent system adapts phases dynamically to current traffic conditions, unlike the dynamic system, which follows a fixed cycle. The pedestrian phase in the dynamic system, appearing every 120 seconds, results in periodic peaks in waiting pedestrians.

Fig. 9a and 9b illustrate vehicle waiting times under both scenarios. Fig. 9a shows that the intelligent system reaches a peak of waiting vehicles between 8 and 15 minutes due to higher pedestrian traffic affecting vehicle flow. In contrast, Fig. 9b indicates a peak at around 15 minutes when pedestrian traffic is lower, allowing the system to balance vehicle and pedestrian phases better. Despite both scenarios having high vehicle traffic, the intelligent system manages fewer waiting vehicles in the low pedestrian scenario. The dynamic system shows consistent behavior with a 120-second cycle time, but the high pedestrian count negatively impacts vehicle dispatch, suggesting that a large number of waiting pedestrians might lead to poor vehicle flow.

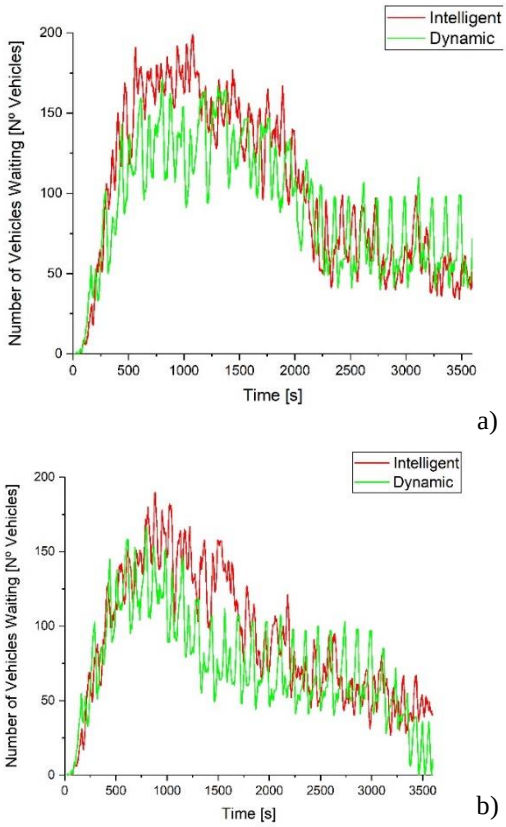

Fig. 9 - Comparison of the number of cars in the entire environment for both systems for the High-High (a) and High-Low (b) scenarios.

5 Conclusions

After analyzing both scenarios, it is evident that the intelligent system generally outperforms the dynamic system in traffic management. The dynamic system achieves better results in terms of vehicle flow, with fewer cars waiting at junctions. However, this advantage comes at the cost of longer pedestrian wait times. The dynamic system, with its fixed 120-second phase cycle, prioritizes vehicles, allowing only a brief 10% of the time for pedestrian crossings. This results in many pedestrians waiting, which can lead to congestion in pedestrian areas.

In contrast, the intelligent system adapts in real time to current traffic conditions. It initially prioritizes pedestrians when their numbers are high but ensures that vehicle flow is not compromised excessively. While the intelligent system does see a higher number of vehicles waiting at times, it does not result in the same level of pedestrian congestion observed with the dynamic system. The intelligent system manages to balance the needs of both vehicles and pedestrians more effectively by adjusting its priorities based on real-time traffic data.

Overall, the intelligent system demonstrates superior adaptability and efficiency. It manages to reduce pedestrian waiting times while still maintaining a

reasonable level of vehicle flow. In comparison, the dynamic system's fixed cycle often leads to longer pedestrian wait times, which can cause significant congestion. Therefore, the intelligent system proves to be more effective in handling the traffic scenarios studied, providing a better balance between vehicle and pedestrian needs.

References

1. E. Uhlemann, “Introducing connected vehicles [Connected vehicles],” IEEE Vehicular Technology Magazine, vol. 10, no. 1. Institute of Electrical and Electronics Engineers Inc., Mar. 01, 2015. doi: 10.1109/MVT.2015.2390920.
2. N. Lu, N. Cheng, N. Zhang, X. Shen, and J. W. Mark, “Connected vehicles: Solutions and challenges,” IEEE Internet of Things Journal, vol. 1, no. 4. Institute of Electrical and Electronics Engineers Inc., pp. 289–299, Aug. 01, 2014. doi: 10.1109/JIOT.2014.2327587.
3. M. M. Rana and K. Hossain, “Connected and Autonomous Vehicles and Infrastructures: A Literature Review,” International Journal of Pavement Research and Technology, vol. 16, no. 2. Springer, pp. 264–284, Mar. 01, 2023. doi: 10.1007/s42947-021-00130-1..
4. D. O’Brien,; H. Le Minh, L. Zeng, G. Faulkner, K. Lee, D. Jung, Y. Oh, E. T. Won, E.T. Indoor Visible Light Communications: Challenges and prospects. In Free-Space Laser Communications VIII, Proceedings of the Optical Engineering + Applications, San Diego, CA, USA, 10–14 August 2008; SPIE: Bellingham, WA, USA, 2008; Volume 7091, pp. 60–68.
5. D. Liu, L. Li, A traffic light control method based on multi-agent deep reinforcement learning algorithm. Sci. Rep. 2023, 13, 9396.
6. R. Fernandes, M. A. Vieira, M. Vieira, P. Vieira, P. Louro, and M. Véstias, “Using visible light communication to implement intelligent traffic signals and cooperative trajectories at urban intersections,” in Light-Emitting Devices, Materials, and Applications XXVII, J. K. Kim, M. R. Krames, and M. Strassburg, Eds., SPIE, 2023, p. 124410G. doi: 10.1117/12.2647862.
7. M. A. Vieira, M. Vieira, P. Vieira, R. Fernandes, and P. Louro, “Dynamic vehicular visible light communication for traffic management,” in Next-Generation Optical Communication: Components, Sub-Systems, and Systems XII, G. Li, K. Nakajima, and A. K. Srivastava, Eds., SPIE, 2023, p. 124290O. doi: 10.1117/12.2647866.
8. M. A. Vieira, M. Vieira, P. Louro, P. Vieira, and A. Fantoni, “Vehicular Visible Light Communication for Intersection Management,” Signals, vol. 4, no. 2, pp. 457–477, 2023, doi: 10.3390/signals4020024.
9. G. Galvão, M. Vieira, P. Louro, M. Vieira, M. Véstias, and P. Vieira, “Visible Light Communication at Urban Intersections to Improve Traffic Signaling and Cooperative Trajectories,” Apr. 2023, pp. 60–65. doi: 10.1109/YEF-ECE58420.2023.10209320.
10. A. Vidali, L. Crociani, G. Vizzari, and S. Bandini, “A Deep Reinforcement Learning Approach to Adaptive Traffic Lights Management.”[Online].Available: https://population.un.org/wup/
11. G. Galvão, M. A. Vieira, M. Vieira, P. Vieira, P. Louro, M. Vestias, P. Lourenço, "Traffic signals and cooperative trajectories at urban intersections: leveraging visible light communication for implementation," Proc. SPIE 12906, Light-Emitting Devices, Materials, and Applications XXVIII, 129060O (13 March 2024); doi : 10.1117/12.3000529
12. A. Tigga, L. Hota, S. Patel, and A. Kumar, “A Deep Q-Learning-Based Adaptive Traffic Light Control System for Urban Safety,” in Proceedings - 2022 4th International Conference on Advances in Computing, Communication Control and Networking, ICAC3N 2022, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 2430–2435. doi: 10.1109/ICAC3N56670.2022.10074123.
13. M. A. Vieira, G. Galvão, M. Vieira, M. Véstias, P. Vieira, and P. Louro, “Traffic Signaling and Cooperative Trajectories based on Visible Light Communication,” Sensors and Electronic Instrumentation Advances, p. 23, 2023.

Acknowledgements

This work was sponsored by FCT – Fundação para a Ciência e a Tecnologia, within the Research Unit CTS – Center of Technology and Systems, reference UIDB/00066/2020.