Roads Infrastructure Digital Twin: A Step Toward Smarter Cities Realization

Oussama El Marai, Tarik Taleb, and JaeSeung Song

Abstract

Digital Twin is a new concept that consists of creating an up-to-date virtual asset in cyberspace which mimics the original physical asset in most of its aspects, ultimately to monitor, analyze, test, and optimize the physical asset. In this article, we investigate and discuss the use of the digital twin concept of the roads as a step toward realizing the dream of smart cities. To this end, we propose the deployment of a Digital Twin Box to the roads that is composed of a 360° camera and a set of IoT devices connected to a Single Onboard Computer. The Digital Twin Box creates a digital twin of the physical road asset by constantly sending real-time data to the edge/cloud, including the 360° live stream, GPS location, and measurements of the temperature and humidity. This data will be used for real-time monitoring and other purposes by displaying the live stream via head-mounted devices or using a 360° web-based player. Additionally, we perform an object detection process to extract all possible objects from the captured stream. For some specific objects (person and vehicle), an identification module and a tracking module are employed to identify the corresponding objects and keep track of all video frames where these objects appeared. The outcome of the latter step would be of utmost importance to many other services and domains such as national security. To show the viability of the proposed solution, we have implemented and conducted real-world experiments where we focus more on the detection and recognition processes. The achieved results show the effectiveness of the proposed solution in creating a digital twin of the roads, a step forward to enable self-driving vehicles as a crucial component of smart mobility, using the Digital Twin Box.

Introduction

Over the last decade, there have been many rapid technological advancements in various fields such as networking, cloud computing, computer vision, Internet of Things (IoT) and Artificial Intelligence (AI). These technologies are supporting each other in one way or another. For instance, the rapid expansion of IoT devices is expected to reach 5.8 billion endpoints by the end of 2020 [1]. These devices, among many others, are constantly generating and sending data over the network, which creates a bandwidth crunch for network providers. Consequently, the current network infrastructure will be incredibly overwhelmed, which desperately pushes toward enhancing the network's performance in terms of bandwidth and latency. Luckily, the 5G technology is already here, or to be deployed soon, to carry out the expected sheer volume of the exchanged data. Once the data arrives at its destination, it will be treated and processed to extract relevant information, such as detecting and tracking objects from a video stream, and generating useful knowledge. The latter task demands heavy computations that require powerful RAM, CPU and likely GPU capable machines which are nowadays easily accessible, thanks to cloud computing technologies. Alongside the hardware advances, the abundant amount of data we are witnessing in this era has essentially contributed to the mushrooming and matureness of AI techniques [2].

With continuous technology advances, consumers become more and more demanding and their satisfaction level is pushed further. A major paradigm we are witnessing nowadays is the shift toward automated systems as consumers are increasingly looking forward to a fully connected world that encompasses most of our life's fields including education, healthcare, industry, transportation, and social life. This interconnection promises a lavish lifestyle and enhances safety, efficiency, productivity, energy consumption, environmental protection, and sustainability.

All the aforementioned technological developments have actively contributed to the emergence of new concepts such as smart cities, ultimately to further improve the quality of life of people. Over the past few years, this concept has drawn much attention from many researchers due to its countless benefits. The motivation behind this is the upward population trend, scarce resources and the environmental damage caused by modern industries and resulting in climate change. These factors are primarily threatening the global food supply of the coming generations, which are expected to reach roughly 10 billion by 2050, according to a United Nations forecast [3]. This fact desperately urges for rethinking and redesigning our actual cities, that are deemed to be the main source for the aforementioned issues, to make them eco-friendly, optimized, smarter and safer. We mean by smarter, a city where every object (e.g., buildings, factories, and cars) has the capability of safely operating autonomously, taking decisions, adapting to changing conditions while being able to timely communicate and exchange information with the other entities. For instance, Smart Factories (SFs), Autonomous Vehicles (AVs) and Digital Twins (DTs) are new emerging and fascinating technologies toward realizing the dream of smart(er) cities.

Digital Object Identifier:
10.1109/MNET.011.2000398

Oussama El Marai is with Aalto University; Tarik Taleb is with Aalto University, Oulu University, and Sejong University; JaeSeung Song is with Sejong University.

Digital Twin is a technology that seems to drum up a lot of interest in the future. It is defined as "a virtual representation of what has been produced" [4]. So, we have:

- The physical asset
- The cyber representation of the physical asset
- The link between those entities which is the data gathered through different sensors and sent to the machine that processes the data and produces the virtual representation.

This concept is of paramount importance and could be used for many purposes such as monitoring, managing, maintaining, optimizing and forecasting. It could be effectively employed in numerous domains including teaching, healthcare consultancy and tourism. To this end, the physical asset should be equipped with relevant sensors (e.g., humidity and temperature sensors) to accomplish specific tasks. These sensors constantly collect data and send it in real-time over the network to either edge or cloud servers. Obviously, the more sensors are deployed to the physical asset, the more accurate the view we get. Once the data reaches its destination, it will be analyzed, synthesized, and eventually displayed to the users in an adequate representation. In this vein, DTs can benefit and leverage new emerging technologies such as IoT devices at the data acquisition phase, 5G at the transmission phase, AI and ML for data mining, analysis and prediction, and finally using video streaming, Virtual Reality (VR) and Augmented Reality (AR) for better data representation and immersive viewing experience.

In this article, we present and describe our Digital Twin Box (DTB) platform for the digital twinning of roads. DTB consists of a number of devices including 360° camera, GPS device and Internet dongle for connectivity. A number of DTBs are deployed to the roads to live stream the moving objects to the cloud servers. These streams are offered to the viewers in real-time (live streams) or shifted (VoD) way. Also, the video streams go through an Object Detection module to extract the different objects and save them to the database. Later, these objects could be used by several applications and domains such as national security, tourism, and Intelligent Transportation Systems (ITS). The use of 360° cameras is motivated by immersive user viewing, which enables a richer and more engaging experience. However, this comes at the expense of overwhelming the bandwidth since it requires higher bitrates [5], which would exert an additional burden on the underlying infrastructure and result in lower user QoE compared to legacy video streams. As a remediation, we propose to locally process the video streams and only transmit the outcome from the object detection and recognition processes.

The rest of this article is outlined as follows. We provide a brief review of previous work. We describe the proposed DT architecture and its different components and modules. The testbed setup as well as the experimental results from the object detection and recognition processes in terms of processing time and number of detected objects are provided. Finally, we summarize the article's contributions and shed light on future research directions.

The use of 360° cameras is motivated by immersive user viewing, which enables a richer and more engaging experience. However, this comes at the expense of overwhelming the bandwidth since it requires higher bitrates, which would exert an additional burden on the underlying infrastructure and result in lower user QoE compared to legacy video streams.

Related Work

In [6], the authors propose a way for generating knowledge as digital twins models from the huge amount of data generated from industrial production lines. To do so, they employ graph-based query language enriched with reasoning rules. The proposed solution aims to facilitate the understanding of complex generated data from production line management systems and automate the process of inferring important rules that help decision making. The proposed automation pipeline consists of four different stages, namely feature extraction, ontology creation, knowledge graph generation and semantic relation extraction. Borodulin *et al.* propose the concept of the Digital Twin-as-a-Service (DTaaS) model in [7]. The proposed model uses a cloud computing platform for the orchestration and simulation of industrial processes in smart factories. It considers the DT as a set of cloud services and permits dynamic resource allocation in the cloud.

The authors in [8] tackle the security aspect of cyber-physical systems (CPS) in the industrial context since any failure with this regard would have catastrophic outcomes for the organization assets and could harm human safety. To this aim, they propose a new framework called *CPS Twining*, that allows operators to create and maintain security-aware digital twins of CPS for monitoring and testing purposes in virtual isolated environments. The proposed framework allows the creation of the virtual environment solely from the specification languages (e.g., AML). Additionally, the generated virtual environments can be used by security experts for testing and validation without affecting the production environment. A prototype has been implemented to demonstrate the effectiveness of the proposed framework.

For the healthcare domain, particularly for elderly health management, a cloud-based digital twin framework, dubbed *CloudDTH*, has been proposed by Liu *et al.* in [9] that allows bridging the gap between the medical physical environment and its clone in cyberspace by designing real-time services for the monitoring and management of the entire lifecycle of elderly people. The proposed framework has been validated through application scenarios that include different factors such as weather (e.g., wind speed and temperature), real-time and recorded patient's physiological data, to demonstrate the feasibility of digital twins in the healthcare field.

In [10], the authors introduce a new digital twin model for vertical farming for sustainable agriculture. The proposed model enables the planning, monitoring and optimization of the operations of the farming process, ultimately improving the productivity of the farms and lowering the costs.

In the sports field, Barricelli *et al.* have proposed in [11] the integration of digital twin technology for the monitoring, assessment, prediction and behavioral suggestions of the athletes. To this end, they propose *SmartFit* framework that could be used by coaches and trainers to monitor in real-time the readiness of their athletes for competitions. *SmartFit* continuously gathers, through IoT sensors embedded in wearable devices, relevant data (e.g., mood and food income) from the athletes. This data is accumulated in the history of the athlete and used later for predicting their performance. Using machine learning techniques, the data history, including the collected measurements and the coach's feedback, is processed and used for suggesting optimized actions for the athlete.

In [12], the authors propose a decentralized approach, based on blockchain, for the creation of a secure digital twins process. The proposed approach permits the verification of data sources and uses only trusted data for the creation of a digital twin. Furthermore, it guarantees the traceability and accessibility of the different logs and transactions at the four different phases of digital twin creation. The proposed approach has been evaluated through different analyses including security, cost and digital twins requirements conformity.

In [13], El Saddik emphasizes the role of recent multimedia content and the Tactile Internet in providing a richer and more engaging user experience. Due to new technological advances in sensory devices, it is now possible to capture and save not only the sound and image but also haptics, olfaction and tastes sensing. These advances would extend the capability of digital twins technology to create a cyber clone very close to its physical or original copy, which certainly offers a richer experience in terms of interactivity and collaboration.

For the construction sector, the authors in [14] proposed exploring the digital twin technology for the management and optimization of the building's operations. To do so, they presented the different steps for the implementation of a case study consisting of a digital twin of a building facade. In this research, the authors discuss the practically-faced issues and limitations during their experiments, mainly related to the IoT devices.

Road's Infrastructure Digital Twin Use Cases and Potential Challenges

In this section, we showcase some highly important use cases and domain applications that essentially rely on the digital model of the roads' infrastructure and highlight some of its related endless benefits. Also, we discuss some of the important challenges related to this technology.

Use Cases

Self-Driving Vehicles: Self-driving vehicles, also known as autonomous vehicles, define a new flourishing technology that is gaining ample interest from researchers around the globe in both industry and academia. This technology is promising manifold benefits at different domain levels, such as economic and environmental, by providing the vehicles with the necessary intelligence to perform common maneuvers and take decisions without requiring human assistance. However, such advanced intelligence can never be achieved with traditional roads' infrastructure, since it heavily relies on the constant exchange of data with its surroundings, among which the road itself. This could be achieved by creating a digital twin of the roads, by deploying sensors into the physical world, collecting and sending data to the IT infrastructure (edge or cloud) to be saved and potentially processed to infer knowledge.

National Security: The creation of a digital twin of the roads would greatly contribute to the diminution of crimes (e.g., car theft) and help the authorities to catch and/or track suspicious persons and vehicles. This is achieved by employing object detection and recognition mechanisms to eventually identify both persons and vehicles and check, for instance, if a given person is wanted or not.

Insurance and Safety: Roads' digital twins would be of utmost importance for insurance companies to resolve accident conflicts using the recorded footage. Additionally, if the spot where an accident occurred is equipped with Internet of Things (IoT) devices that measure the ambient temperature and humidity, this data would be useful to construct a better understanding of the accident circumstances and would lead to accurate and fair decisions.

Challenges

In spite of the myriad benefits of the DT technology, there are a number of potential salient challenges. For instance, the presentation of the huge amount of generated data in a convenient way to the end-user based on their own customization and preferences is a hard task. In this vein, the use of the local dynamic maps concept could be a potential solution. Another possible challenge is to study the DTBs' locations in a way to minimize the number of deployed DTBs, which results in lowering the deployment and maintenance costs, while still ensuring better coverage of the roads. A third painstaking challenge that should be deeply investigated is the security side of the DT, especially when data originates from IoT devices. It is known that IoT devices are naturally vulnerable to security threats, which desperately calls to employ reliable and robust frameworks, such as the blockchain [12, 15], to protect data sources from cyber attacks as well as the data itself until it reaches its final destination. Any alteration or manipulation of the data would result in catastrophic consequences, especially for some domains such as autonomous driving. Indeed, the blockchain framework provides a fascinating distributed approach to preserve data integrity and traceability. The last challenge is the decrease of the detection accuracy during the night due to low contrast against the background. It is worth noting that all these important aspects are out of this article's scope.

System Architecture

In this section, we describe the proposed system architecture for the creation of a digital twin of the roads and its different components as a step toward the realization of the smart cities' dream. This step consists of creating a digital model of the infrastructure, among which the roads, and make them more intelligent. Crucially, this would underpin many other technologies, such as self-driving

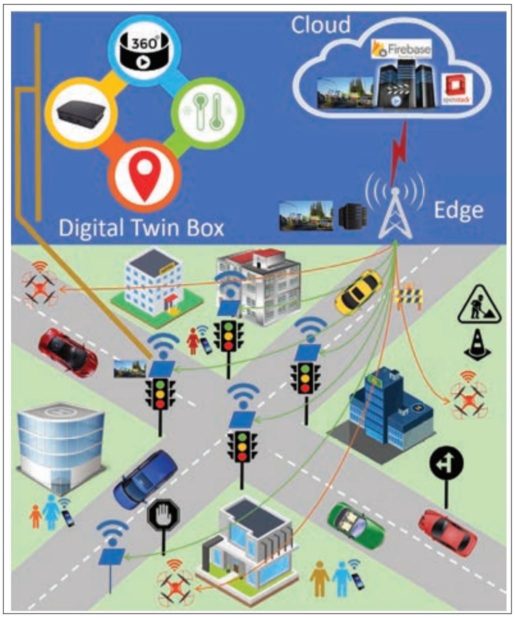

FIGURE 1. Global system architecture.

vehicles, that are deemed to be an indispensable part of the big dream of smart cities. Specifically, we introduce the concept of DTB and describe its interaction with the edge and cloud infrastructures. Ultimately, DTB aims to gather various data of a different nature that is eventually sent to the cloud servers. This data will be stored for ulterior exploitation, and potentially processed to extract and infer relevant information that could be used in many domains, among which security, tourism, and transportation. The collected data is mainly a 360° live stream using a 360° camera, GPS location and other relevant data such as temperature and humidity.

The global proposed system architecture is depicted in Fig. 1. This figure contains different components contributing altogether to the efficient creation of a digital twin of the roads' infrastructure. The proposed architecture allows the accommodation of various services with different levels of requirements. In the following, we provide a detailed description of each component, its role in the proposed architecture, as well as its interaction with the other components.

Digital Twin Box

DTB is a set of IoT devices connected to a Single Onboard Computer (SOC), such as a Raspberry Pi. Mainly, we find a 360° camera that is capable of delivering a 360° spherical video stream. This SOC keeps streaming live the roads. These streams could be accessed in real-time, e.g., for monitoring purposes. They are also temporarily stored at the cloud servers for shift viewing experience. More importantly, the received stream will go through an object detection and recognition process to extract all possible objects contained in the video. To do so, all video frames need to go through the detection process, where we first detect the different objects present in the frame along with their types (e.g., person, car, and traffic light). These objects are cropped using their coordinates within the frame. In this work, we are interested in the tracking of two types of objects, namely persons and vehicles (e.g., car and truck).

After detecting an object and identifying its class category, we perform an object recognition task (i.e., focusing only on the categories of persons and cars) to recognize and identify which person/car it is and if the system has already seen this object previously, in the same or different video, and accordingly register the detected object in the databases of the system. In the case of a new object, we create a new entry for it in the objects table, and save the video frames, where this object has been seen, in the corresponding tables. Otherwise, we retrieve the object ID and update its corresponding video frames in the database. The idea beneath is to keep track of all objects (currently of only person and car types) in the database. These data would be of great importance and could be used by many other services in various use cases as described in the previous section.

Although the global process of object detection and recognition looks similar for all objects, there are some differences in the recognition phase. For the person category, a face detection and recognition module would be required to first locate the face in the whole body image and then perform the matching process with previously recognized persons' faces existing in the database, respectively. As to the car identification process, it entails two separate modules, namely plate detection and plate recognition, to respectively locate the plate number, crop it, and read the corresponding alphanumeric string. For the rest of the objects (e.g., animal, traffic light, and traffic sign), there is no need to perform the recognition process because they are either static objects with no identity or simply not worthwhile (and perhaps not economical) to identify and track in the context of smart cities.

At the end of the object detection and recognition process, it is important to save all detected objects in the database as records as well as cropped images on the disk. This would be useful for inferring new information and making important recommendations. For instance, if a DTB is deployed in a residential area and a dangerous animal is detected, this would help to raise an alert to the authorities to take the relevant actions. It could also be used to generate automatic statistical reports regarding the frequency of human/car circulating in a specific area that could be used to improve the services provided in that area. Figure 2 illustrates the flowchart diagram for the whole object detection and recognition process.

It should be noted that the delivered stream is accompanied by other sensed data such as GPS location, measured ambient temperature, humidity and air quality. For the data coming from IoT devices, it could also be collected from sensors on board Unmanned Aerial Vehicles (UAVs) that are sent for a specific mission, which would provide more accurate measurements [16]. This data will also be saved in the database with its corresponding timestamp. The combination and fusion of all these data with the video stream feed would result in a more holistic view of the area and its environment at different points in time during the four seasons, which might be useful for tourism and insurance domains.

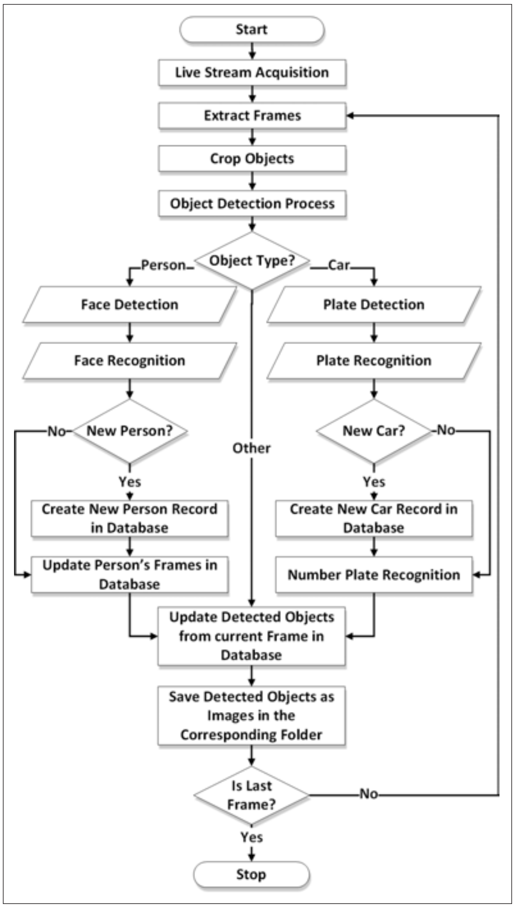

Figure 2. Object detection and recognition flowchart.

Where to Process in the Network?

The different aforementioned processes use Machine Learning (ML) models to accomplish their corresponding tasks. Generally speaking, object detection and recognition tasks are time-consuming and require high CPU and RAM resources. In many cases, it becomes necessary to carry them on GPU-capable machines to achieve reasonable response times. In this subsection, we discuss the different possible options to perform such heavy processing tasks.

Extreme Edge: Offloading the object detection and recognition tasks to the extreme edge (i.e., at the SOC) has the advantage of attenuating the cloud servers' workload and reducing the streaming latency [17]. The latter would be of vital importance for some other technologies such as autonomous driving whereby, for instance, vehicles need to instantly detect, recognize and interpret traffic signs to take some critical actions (e.g., braking and deceleration). Furthermore, processing tasks at the extreme edge would greatly improve the system's scalability, notably when the number of deployed DTBs is high. It also helps reduce pressure on the underlying network infrastructure, especially when the transmitted service is bandwidth-consuming such as video streams, and only the detected objects are accessed by the users. Self-driving vehicle technology would take advantage of it to increase their awareness of their surroundings, especially to detect objects around the corner, by communicating with other DTBs located at the same spot. However, this would increase the cost of the DTB since it requires a powerful GPU-capable SOC such as Jetson AGX Xavier.

Edge: Edge computing is also a good choice when keeping the latency at lower values is a requirement and the extreme edge device has very limited resources (e.g., Raspberry pi). Additionally, it helps to save the bandwidth utilization by sending the video streams and the sensed data to an edge server that is close to the DTB in question, which may offer better processing capabilities while keeping the deployment cost of the DTB as low as possible. The edge computing paradigm offers an ideal environment for many use cases where both the DTBs and the consumers of the live streams or the outcome of the object detection and recognition processes are in the vicinity of the edge servers. It is worth noting that we can still do some lightweight object detection processes on devices with limited resources using some models (e.g. Tensorflow Lite) that are specifically designed and optimized for IoT and mobile devices.

Cloud: Cloud computing may offer the most powerful and scalable configuration for handling heavy tasks due to its highly-available resources. However, sending live streams to the cloud for processing may adversely affect its scalability, notably when the number of deployed DTBs groes up for fully covering an area. Moreover, this would exert high pressure on the network, especially when streaming bandwidth-intensive videos, such as 360° videos.

Users

The outcome of the DTBs is the digital twin of the roads, consisting of 360° live streams and/or the detected objects from the live streams as well as the different sensed data sent by the IoT devices deployed within the DTBs. This result could be used by many consumers (e.g., individual users, corporate, or autonomous vehicles) for different purposes, among which monitoring, maintenance and safety, in various use cases. In the tourism domain, for instance, users can view 360° videos (either live or recorded) using head-mounted devices or HTML5 players to see a specific area they are interested in, along with the different sensed data at a specific period, before they travel. Also, the national security authorities can also use the platform to search for a suspicious person/car just by using their picture/its plate number to see the different places visited by that person/car, respectively.

Experimentation and Performance Results

This section describes the setup used during the conducted experiments and discusses its outcome. We first provide details on the device specifications used to accomplish this experiment, as well as the technologies used in our implementation. Then, we describe the platform setup in which we have run our implementation. Finally, we provide and discuss the different obtained results from the experiment. The conducted experiment has been performed using a video with 512 × 512 resolution.

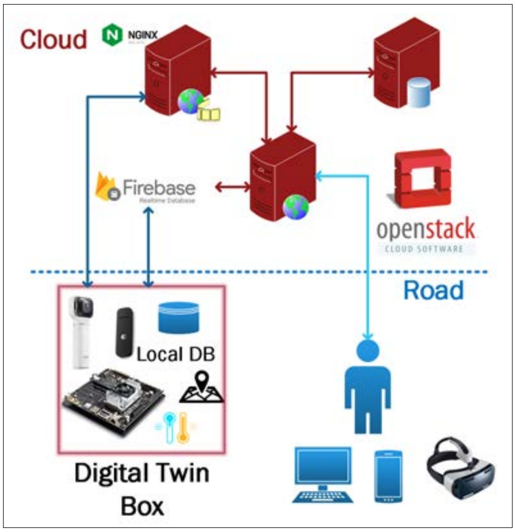

FIGURE 3. Platform setup.

Devices Specifications and Technologies Choice

In this experiment, we show the performance of the detection and recognition processes at the extreme edge. To this end, we have used a Jetson TX2 single onboard computer (SOC) from NVIDIA. The Jetson TX2 is a powerful SOC and is designed for computation-heavy tasks. It has a GPU architecture with 256 NVIDIA CUDA cores, a Quad-Core ARM® Cortex®-A57 MPCore, and 8GB 128-bit LPDDR4 memory. The operating system running on this Jetson TX2 is Ubuntu 18.04. To detect the different objects in video frames, we have used the Single-Shot Detector (SSD) mobile net v2 model that ships with Jetson SOCs. It is faster and provides more accuracy compared to other state of the art models such as YOLO and Faster R-CNN models.

For the face detection part, we have used a python library called *face_recognition* that uses deep learning models to locate the face within a person's body. The model used by this library provides very high accuracy of 99.38 percent. For the recognition phase, we perform a comparison of the new face with the existing faces in our database and we calculate the euclidean distance between the two faces. If the distance is above a certain threshold (0.7 in our experiments) we consider it the same person.

Regarding the vehicle's plate number, we have used Google AI vision API to detect and locate the plate number with an accuracy of 81.7 percent, whereas we perform text recognition on the text contained inside the plate box. If the text does not follow the general pattern of the plate numbers in the country, we reject it. Otherwise, it will be considered as a successfully recognized plate number.

Platform Setup

An overview of the platform setup is illustrated in Fig. 3. In our OpenStack cloud platform, we have three servers, namely web, streaming and database servers. The streaming server receives from the DTB the 360° video via the 4G LTE network. The SOC in the DTB captures the live stream and performs object detection and recognition and stores the results in a local database. Also, it constantly reads measurements from the IoT devices along with the GPS data and saves the measured values locally. A backup module is periodically executed to sync the local database at the SOC with the real-time Firebase and a MySQL database in OpenStack servers. It is worth noting that Firebase is used to store real-time data such as the detected objects and GPS data, while the relational database is used to store other data such as the user's info and the countries data. On the other hand, users can access the 360° streams for discovering or monitoring purposes via head-mounted devices or simply using the web-based player. They can also access the web platform to visualize the detected objects and track the recognized ones (i.e., persons and vehicles).

Performance Results

In this subsection, we show the performance results from the detection and recognition processes in terms of the time consumed per frame to detect and recognize all the objects contained in that frame, as well as the number of detected objects.

Figures 4a and 4b illustrate the per frame elapsed time to detect and recognize the objects at each frame, respectively. It is worth noting

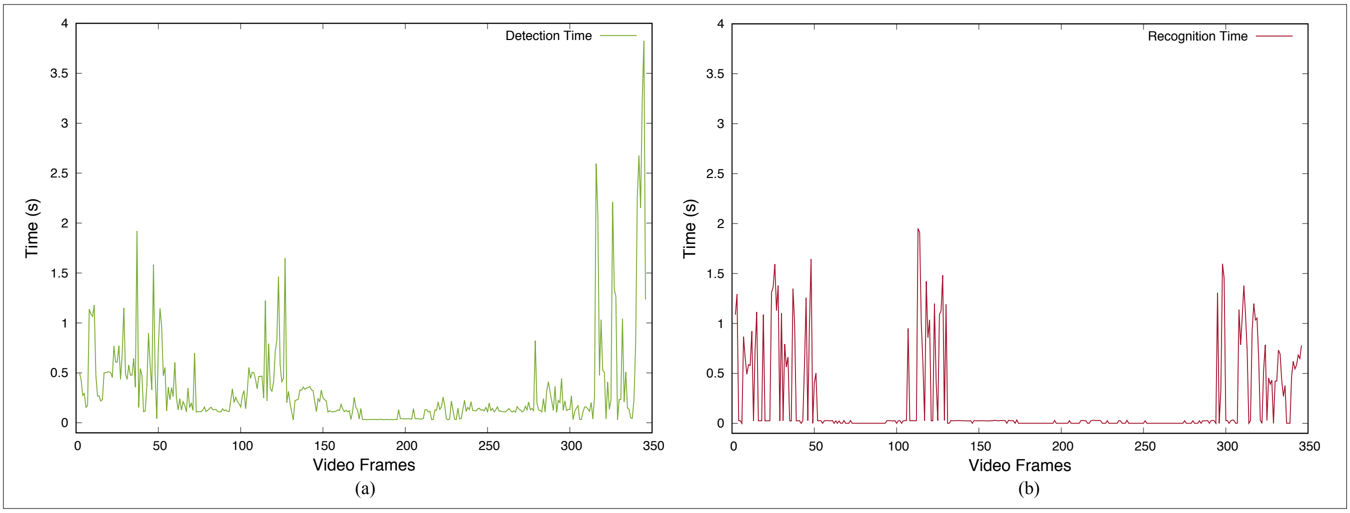

FIGURE 4. Object detection and recognition time per video frame: a) object detection time per video frame; b) object recognition time per video frame.

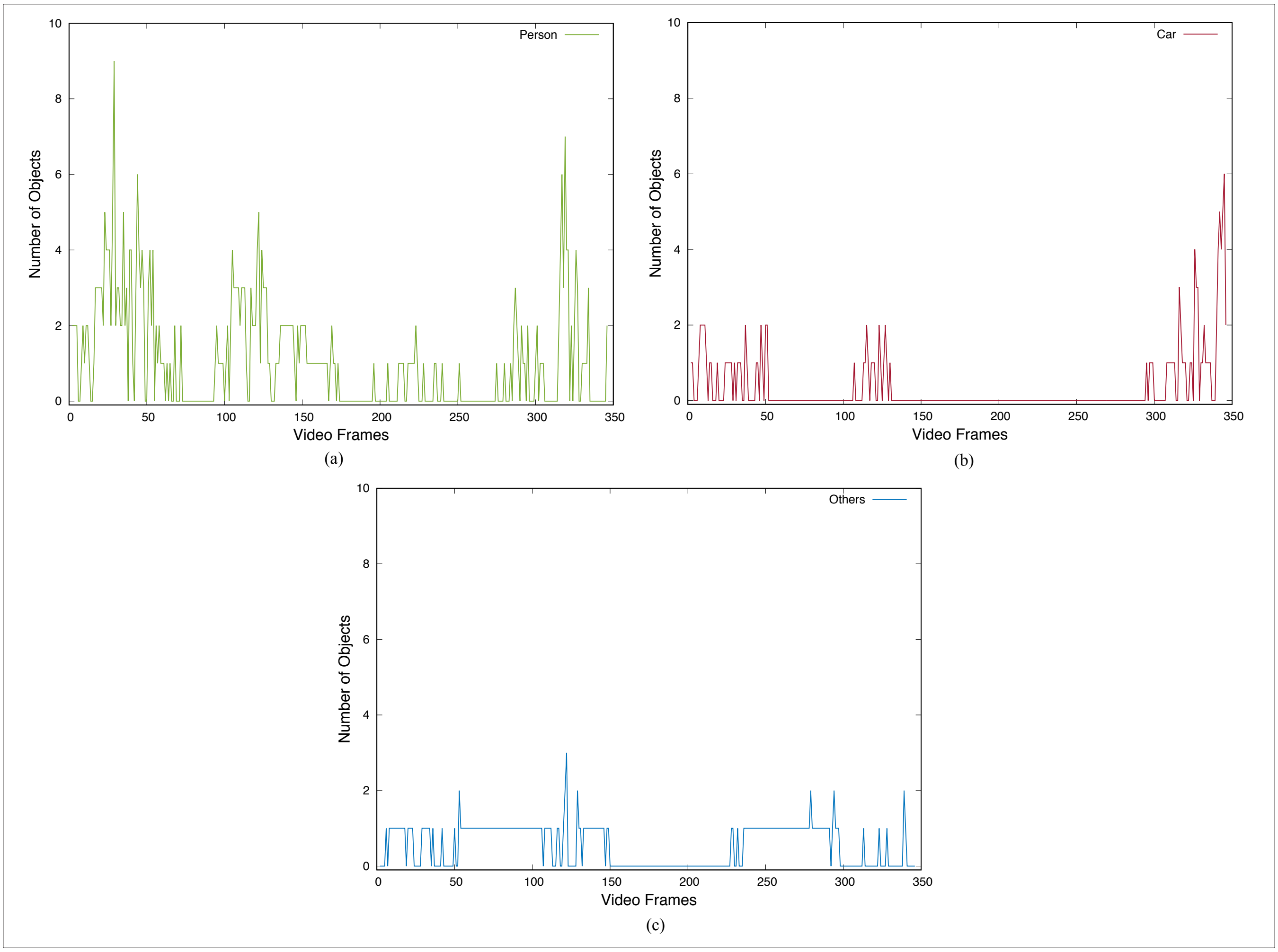

FIGURE 5. Number of detected objects per video frame: a) number of detected persons per video frame; b) number of detected cars per video frame; c) number of other detected objects per video frame.

that the recognition time, represented in Fig. 4b, includes the time consumed to recognize both persons and vehicles. As we can see, usually the process of detection does not take too much time, especially when there are no objects to recognize, for instance, between the interval [55, 100] frames and [135, 290] frames. This is mainly due to the speed of the SSD mobile net v2 model and its ability to detect multiple objects in one shot pass. We can also see that when the processing of one frame is relatively high, this is mainly caused by the recognition process (the red curve), as we can see during the intervals [0, 50], [110, 135] and [290, 340].

The number of detected objects per video frame is plotted in Figs. 5a–5c. These figures respectively correspond to three categories of detected objects, namely persons, cars and others which basically include the rest of the objects such as traffic lights. Aligned with the plots from Fig. 4, the number of detected objects is higher within the frames interval where the time is relatively high, which is quite intuitive since the detection and recognition of more objects incur more time. From Fig. 4 and Fig. 5, we also conclude that the recognition process takes more time than the detection as we observe time picks in Figs. 4a and 4b when a person and/or car objects are detected as per Figs. 5a and 5b.

In Figs. 6a and 6b, we show the number of detected persons and cars, respectively. These figures also show the frequency (orange bars) of the recognized objects throughout the video as well as the average recognition time. The average recognition time for cars is relatively high compared to persons.

Conclusion

In this article, we have proposed a methodology to create a digital twin of the roads' infrastructure, which is considered as a step toward enabling a gamut of essential technologies (e.g., self-driving vehicles) and services that are deemed to be a crucial part in the journey of realizing the dream of smart cities. It consists of deploying Digital Twin Boxes (DTB) composed of a 360° camera, GPS device and other IoT devices for sensing environmental measurements such as ambient temperature and humidity. Additionally, we perform object detection and recognition on video streams to extract all possible objects and save them to the database. The recognition process is performed on two types of objects, namely vehicles and persons, and entails both identification and tracking processes to keep track of when and where the selected objects appeared, along with other measured data received from IoT devices. The combination and fusion of all

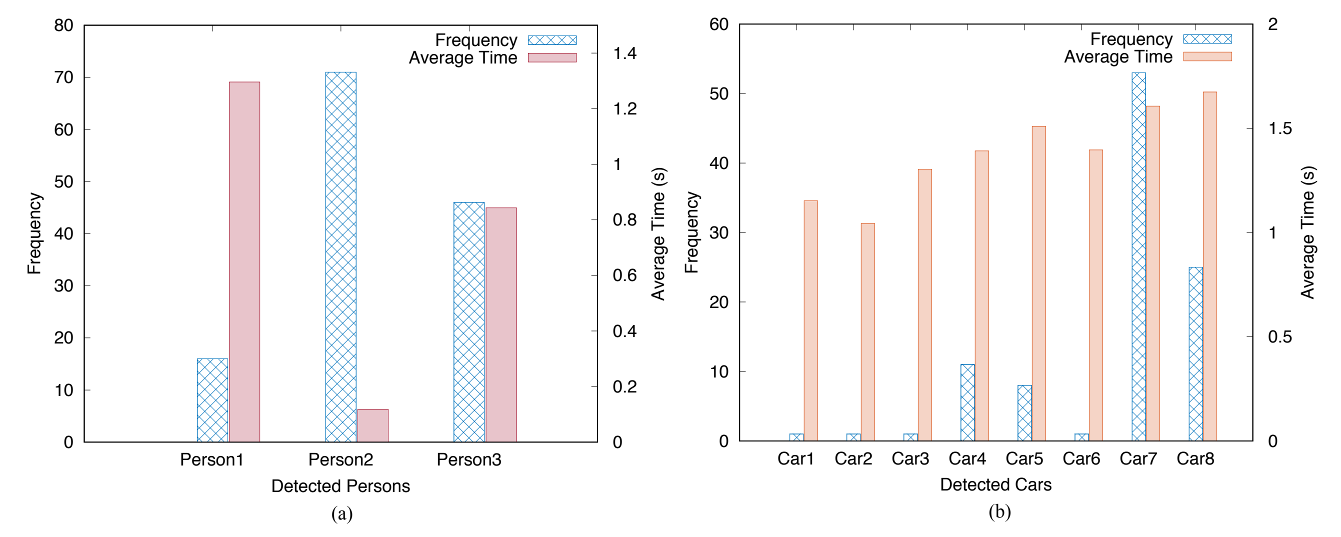

FIGURE 6. Persons and cars recognition frequency and average recognition time: a) persons recognition frequency and average recognition time per person.; b) cars recognition frequency and average recognition time per car.

gathered and processed data will give a better understanding of the contextual circumstances when accessing the data. The resulting database would be of great importance for many other services and domains such as tourism, insurance and national security.

In our next research plan, we endeavor to improve both the face and car plate number recognition to increase their accuracy. Also, we plan to optimize the placement of the DTBs in the roads to make the overall deployment cost-efficient.

Acknowledgment

This work is partially supported by the Academy of Finland Project CSN, under Grant Agreement 311654, and the 6Genesis project under Grant No. 318927. Prof. Song was supported by an Institute for Information & Communications Technology Promotion (IITP) (No. 2018-0-01456, the AutoMaTa project).

References

[1] "Gartner Says 5.8 Billion Enterprise and Automotive IoT Endpoints Will be in Use in 2020," https://www.gartner.com/en/newsroom/press-releases/ 2019-08-29-gartner-says-5-8-billion-enterprise-and-automotive-io, accessed: 2020-04-18.

[2] Y. Lu *et al.*, "Communication-Efficient Federated Learning for Digital Twin Edge Networks in Industrial IoT," *IEEE Trans. Industrial Informatics*, pp. 1–1, 2020.

[3] "United Nations: World Population Projected to Reach 9.8 Billion in 2050, and 11.2 Billion in 2100," https://www.un.org/development/desa/ en/news/population/world-population-prospects-2017.html, accessed: 2020-04-24.

[4] M. Grieves, "Digital Twin: Manufacturing Excellence Through Virtual Factory Replication," white paper, vol. 1, 2014, pp. 1–7.

[5] C.-L. Fan *et al.*, "A Survey on 360° Video Streaming: Acquisition, Transmission, and Display," *ACM Comput. Surv.*, vol. 52, no. 4, Aug. 2019, https://doi.org/10.1145/3329119.

[6] A. Banerjee *et al.*, "Generating Digital Twin Models Using Knowledge Graphs for Industrial Production Lines," *Proc. 2017 ACM on Web Science Conf.*, ser. WebSci '17. New York, NY, USA: Association for Computing Machinery, 2017, pp. 425–30, https://doi.org/10.1145/3091478.3162383.

[7] K. Borodulin *et al.*, "Towards Digital Twins Cloud Platform: Microservices and Computational Workflows to Rule a Smart Factory," *Proc. 10th Int'l. Conf. Utility and Cloud Computing*, ser. UCC '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 209–10. , https://doi.org/10.1145/3147213.3149234.

[8] M. Eckhart and A. Ekelhart, "Towards Security-Aware Virtual Environments for Digital Twins," *Proc. 4th ACM Wksp. Cyber-Physical System Security*, ser. CPSS '18. New York, NY, USA: Association for Computing Machinery, 2018, pp. 61–72, https://doi.org/10.1145/3198458.3198464

[9] Y. Liu *et al.*, "A Novel Cloud-Based Framework for the Elderly Healthcare Services Using Digital Twin," *IEEE Access*, vol. 7, 2019, pp. 49,088–101.

[10] J. Monteiro *et al.*, "Towards Sustainable Digital Twins for Vertical Farming," *2018 13th Int'l. Conf. Digital Information Management (ICDIM)*, 2018, pp. 234–39.

[11] B. R. Barricelli *et al.*, "Human Digital Twin for Fitness Management," *IEEE Access*, vol. 8, 2020, pp. 26,637–664.

[12] H. R. Hasan *et al.*, "A Blockchain-Based Approach for the Creation of Digital Twins," *IEEE Access*, vol. 8, 2020, pp. 34,113–126.

[13] A. E. Saddik, "Multimedia and the Tactile Internet," *IEEE MultiMedia*, vol. 27, no. 1, 2020, pp. 5–7.

[14] S. H. Khajavi *et al.*, "Digital Twin: Vision, Benefits, Boundaries, and Creation for Buildings," *IEEE Access*, vol. 7, 2019, pp. 147,406–419.

[15] R. Li and H. Asaeda, "A Blockchain-Based Data Life Cycle Protection Framework for Information-Centric Networks," *IEEE Commun. Mag.*, vol. 57, no. 6, 2019, pp. 20–25.

[16] N. H. Motlagh, M. Bagaa, and T. Taleb, "Energy and Delay Aware Task Assignment Mechanism for UAV-Based IoT Platform," *IEEE Internet of Things J.*, vol. 6, no. 4, 2019, pp. 6523–36.

[17] Y. Lu *et al.*, "Low-Latency Federated Learning and Blockchain for Edge Association in Digital Twin Empowered 6G Networks," *IEEE Trans. Industrial Informatics*, 2020, pp. 1–1.

Biographies

Oussama El Marai (oussama.elmarai@aalto.fi) received his engineering degree in computer science in 2005 from the University of Science and Technology Houari Boumediene (USTHB), Algiers, Algeria, the master degree from Ecole Nationale Superieure d'Informatique (ESI) in 2009, and he is pursuing his doctoral degree at the School of Electrical Engineering, Aalto University, Finland. His research interests include adaptive video delivery, video QoE optimization, digital twin and image processing.

Tarik Taleb [S'05, M'05, SM'10] (Tarik.Taleb@aalto.fi) received the B.E. degree (with distinction) in information engineering in 2001, and the M.Sc. and Ph.D. degrees in information sciences from Tohoku University, Sendai, Japan, in 2003 and 2005, respectively. He is currently a professor with the School of Electrical Engineering, Aalto University, Espoo, Finland. He is the founder and Director of the MOSA!C Lab (http://www.mosaic-lab.org/).

JaeSeung Song [M'09, SM'17] (jssong@sejong.ac.kr) received the B.S. and M.S. degrees in computer science from Sogang University, and the Ph.D. degree from the Department of Computing, Imperial College London, U.K., in 2013. He is currently an associate professor, leading the Software Engineering and Security Lab (SESLab), Computer and Information Security Department, Sejong University.