Federated Edge for Tracking Mobile Targets on Video Surveillance Streams in Smart Cities

Francesco Martella*, Maria Fazio*, Antonio Celesti*, Valeria Lukaj*, Antonino Quattrocchi*, Massimo Di Gangi*, Massimo Villari*†
* University of Messina, Italy - { fmartella, mfazio, acelesti, vlukaj, aquattrocchi, mdigangi, mvillari}@unime.it
† ALMA Digit S.R.L., Messina, Italy - m.villari@almadigit.com

***Abstract*— Nowadays, video surveillance is a very common practice in Smart Cities. There are public and private video surveillance systems, and very often different systems or single devices frame the same area. However, when a target needs to be identified or needs to be tracked in real-time, such solutions typically require human intervention to configure the devices in the best possible way (e.g., choosing the optimal cameras, setting up their focus, and so on). To address such a problem, in this paper, we define a new interrogation method based on a Federated Edge approach. This approach addresses the problem from the point of view of both camera hardware and shooting angle associated with it. According to the presented approach, it is possible to understand which the best camera to identify a target and possibly tracking it in a specific area is. A case study is defined in the context of urban mobility management.**

***Index Terms*—Smart City, Urban Mobility, Solid Angle, Video Analysis, Edge Computing, Federated Edge.**

I. Introduction

The Smart City model is now a constant in continuous growth all over the world. This sector is the subject of several funding programs in Europe, e.g., Next Generation EU and Horizon Europe make available hundreds of millions of euros for the digitization of services and territories. Particular importance in this digitization process is given to the safety of citizens and the protection of the environment. To this end, the massive deployment of city monitoring cameras is assuming considerable importance. Typically, cameras in Smart Cities are used for video surveillance or traffic control and over the years they have been installed in specific sites for the acquisition of images and videos on a specific road/area. However, modern cameras have network and computation capabilities that make them smart devices able to perform analysis on the Edge in a collaborative way. To better understand the new challenging scenarios for smart city video monitoring, let us consider a group of cameras that film the same place and the same scenes from different perspectives and a reference target that is of interest for the monitoring system (e.g., a vehicle entering a limited-traffic zone, a crowd of people, a too fast mobile object, etc.). The cameras can provide information on the same target with different quality levels that depend on several factors, such as the distance of the target from the camera lens, the camera resolution, the focus angle, the type of target, the velocity of the target, etc., and it is not possible to know in advance which the camera that better suits the monitoring/tracking of the target is. For this reason, we can no more analyze video flows coming from different cameras independently of each other, but we should start thinking about connected systems where a federation of cameras that acquire data on the same target cooperate to provide the highest quality information to the application layer (e.g., video surveillance or traffic control). This paper intends to introduce the concept of *Federate Edge*, enabling a dynamic set up of smart cameras on the basis of their ability to recognize a target in a common scene. We start from the characterization of the target itself and the ability of cameras to get information on it. To this aim, we investigate the solid angle construction around the target to normalize data and estimate information quality. Then, we provide design strategies and enabling technologies for the establishment of the Federated Edge. Inside the federation, the devices will collaboratively identify the one that is most suitable for recognizing the target of interest. The system thus designed is able to respond to target-based queries optimizing the results on the basis of data from the best device. In the presentation of the concepts, we will refer to two use cases as concrete examples about the possible benefits of the proposed solution, which are video surveillance and traffic monitoring. They are analyzed with reference to the real deployment of traffic cameras in the city of Messina, which has hundreds of cameras managed by a centralized Video Management System (VMS) and is currently involved in two project on urban mobility, which are the MeSmart project, funded by the Italian PON Metro 2014-2020, and URBANITE funded by the European Horizon 2020. The paper is organized as follows. A comparison between related initiatives available in literature and our solution is presented in Section II. Objectives and benefits of our work are reported Section III. The key concepts on Federeted Edge is presented in the Section IV, whereas in Section V, we present our reference architecture and design strategies. Final remarks and possible future developments are summarized in Section VI.

II. State of the Art

When different cameras film the same scene, its elements can be observed from different angles, and therefore a camera can recognize the same element or not. In [1] the authors address these problems by basing their study on the recognition through zooming of objects at a great distance. However, this method has limitations in adapting to the scene changes that our approach aims to address. An interesting approach to

978-1-6654-9792-3/22/$31.00 ©2022 IEEE

target recognition in an area is described in [2]. The authors demonstrate that they achieve high accuracy in detecting and recognizing a static target, while this is not the case for moving targets. Our work aims at solving this problem taking into account the devices "close" to the Edge device that locks the target. In [3] the authors introduce an image analysis method that maintains high detection performance by reducing the number of pixels processed by about 70% and the detection time by more than 50%. In our solution we propose how it is possible to carry out analyzes and queries on the Edge reducing the number of pixel acquired. [4] contains important considerations regarding the hardware used in the cameras. In our work we want to propose an Edge-Based model [5] considering the NGSI-LD standard that allows us to choose the camera to be used also considering the on-board hardware. The use of a database to be queried to search for the best camera for the required need also implies the introduction of security concepts [6] and unique geo-referencing of both the camera and the image [7]. Several approaches also based on neural networks for target recognition have been used in the development of computer vision. In [8] some representative target detection algorithms are analyzed considering the problems of algorithms. The study we are proposing aims at finding a solution that can be applied on the Edge by solving the various problems faced in the literature. An interesting study on the recognition of targets in low resolution conditions is reported in [9]. The authors achieved good results. However, it is interesting to compare this method in terms of computing power used with the one proposed in this paper. In [10] it is highlighted how it is possible to use customized generic Edge devices to carry out multiple activities simultaneously as a solution to lighten the work of Cloud infrastructures. The paper merely shows how a target identification algorithm can be run on an Edge computing device. In our work we want to introduce the concept linked to this type of device of being able to be interrogated to understand which targets "see best". Collaborative Cloud and object tracking Edge are presented in [11], in which Machine Learning (ML) algorithms are described in the approach of a partial processing of the video capture on the Edge. Perimeter networks created between devices are used in cooperative cache and video features in [12] where opportunistic algorithms for sharing video portions are taken into consideration. Edge Computing-based adaptive wireless video streaming mentioned in [13] is based on the idea of adopting Dynamic Adaptive Streaming over HTTP (DASH) for perimeter transcoding by cooperating with Edge device and backend. Our goal is to use the Edge approach integrated with NGSI-LD to characterize the device and understand in real time which device is better than another in recognizing and searching for targets even in real-time.

III. Smart Camera Data Acquisition and Target Identification

The identification of a target in a VMS is a well-known problem in the field of computer vision. The images registered by a camera have several features that could be gathered and analyzed, such as:

- *Scene*: it is the image that represents the whole environment at a specific instant of time;
- *Object(s)*: it is an element in the scene (e.g., a vehicle, a person, a tree, ...) that can be characterized by specific properties (e.g., size, color, movement,...);
- *Target*: it is the "interesting" feature of the scene. The adjective "interesting" depends on the specific use case and, hence, the target is specified by the application/end user and can change during the time. The target can coincide with an object of the scene or can be an information get from a part of the scene (e.g., the movement of an object).

The installation of the cameras depends on the width of the area to monitor and the perspective. Thus, often cameras film the same scene but with different visions. Also, their ability to identify targets is very different due to their specific technological features. For this reason we define *the ability to recognize a specific target of an Edge Device* as the set of its hardware and software properties and its relative position with respect to the target that make it able to provide information on the target. The *quality of information* provided by and Edge device on a target is the probability that the Edge device recognizes the target in one point of the scene is greater than the probability that another Edge device that frames the same scene recognizes the same target. Moreover, the target characterizes the smart camera processing at the Edge. For example, in the management of limited access traffic zones, the target could be the vehicles' license plate and smart cameras have to perform license plate recognition. For this reason, in this paper, we refer to smart cameras also as Edge devices.

In the field of smart cities, where there are often hundreds of video cameras deployed in the environment, it may be useful to understand:

1) given a uniquely identified area: which camera has the best capacity to recognize a specific target?
2) given a target, which is the best camera able to recognize it and at which quality level?

The proposed work aims at identifying the methods and technologies useful for setting up an Information and Communication Technology (ICT) system that can answer these questions. In particular, we aim at designing a digital solution which allows querying the system providing as input a target or an area of the city or a specific camera feature, and receives as response data from the camera(s) that best meets the input requirement(s).

A. Video Surveillance Scenarios

In smart cities, video surveillance is often exploited for urban security and for traffic control. Compared to the possibility of having different cameras that film the same area, there are therefore 2 scenarios.

The first scenario concerned the urban security. As can be seen in Figure 1 in a square, or in an area closed to traffic, it

Fig. 1. Cameras that frame overlapping areas in a square (Piazza Cairoli, Messina-Italy).

is possible that several cameras film common areas. After the position or even the hardware characteristics it is possible that the analysis capacity of the video is different and therefore a target can be recognized better from one shooting point than another. In Figure 1, it is evident how the two cameras partially film the same area with a person crossing the square. If, for example, the person is a target depending from his movements, from the light and also from the presence of any other object in the scene, it may happen that the same algorithm can have different results on one video instead of the other.

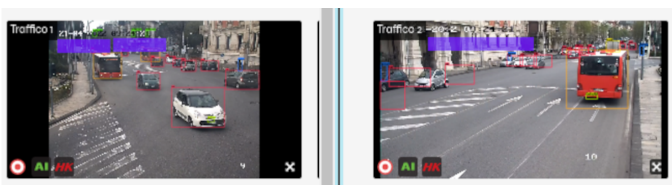

Fig. 2. Cameras that film overlapping areas in Street Viale Garibaldi, Messina-Italy.

The second scenario of interest concerns traffic monitoring. Also in this case there are situations in which several cameras view the same area. Figure 2 shows an example of this scenario. The framed area is located in street Garibaldi, Messina (Italy) and represents a critical artery for city traffic. The 2 cameras frame a large common area from 2 different perspectives, and it is clear that, if the target is the bus, the details captured from the 2 cameras are different. Being a trafficked area it is likely that in some cases it will not be possible to read a vehicle license plate or to clearly understand the details of an accident. If, on the other hand, a system is defined that can modify the type of shot on a point on the basis of a dynamically specification, it is possible to overcome the problems relating to these types of situations.

Another example of the second scenario described is shown in Figure 3. In this case even 3 cameras frame a vast common scene located in Vittorio Emanuele II street, Messina (Italy) The framed area is close to that shown in Figure 2 and is also a critical area from the point of view of city traffic. It is evident from the scenes in Figure 3 that the common area for the 3 cameras is very large. The box identified as "Traffico 7" and "Traffico 6" clearly frame the same truck in the foreground. In "Traffico 7" a vehicle is highlighted in red in the background that you see in the "Traffico 5" box. The same tram stop is evident in the different shots. In traffic conditions, the specific functions of a camera could also be affected here by unforeseen obstacles, particular lighting or

Fig. 3. Cameras that frame overlapping areas in Street Vittorio Emanuele II, Messina-Italy (Traffico 5, Traffico 6, Traffico 7)

traffic conditions. Therefore, the importance of an adaptive system that is able to modify itself according to the requests or specific needs is always greater as the number of devices that frame a common area increases.

IV. Target Tracking for Smart Video Surveillance

Target tracking is an issue related to the effective adoption of image recognition methods. Examples of such interesting methods are described in [14] and [15]. In particular in [15], the parameters that can be used are associated with the solid angle that subtends at the camera lens by the frame of the image.

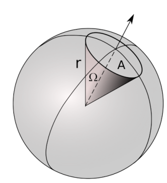

Fig. 4. Solid Angle representation ("http://pngegg.com")

Solid angle is an extension in three-dimensional space of the plane angle. If we consider Figure 4 we can define it as:

$$\Omega = A/R^2 (SolidAngle)$$

In this formula, *A* represents the area of the spherical portion of radius *R* seen under the angle. The ratio, even in the three dimensions, between the portion of circumference, the radius and the subtended angle is maintained. As in the planar angle, the solid angle can be defined how the ratio between the area of the spherical surface and the radius of the sphere considered. For better understand the concept, we can image a light bulb in the centre of a sphere in Figure 4. For the whole sphere the solid angle through which the light rays pass is valid: If we consider the only part of the spherical surface crossed by the light rays, using the differentials we obtain:

$$d\Omega = dS/(R^2) = (R^2 \theta d\theta d\phi)/R^2 = sin\theta d\theta d\phi$$

where θ is the *latitude* (angle from the north pole) and ϕ is the *longitude*. This value represent the portion of surface that a given camera can be able to frame given a specific angle.

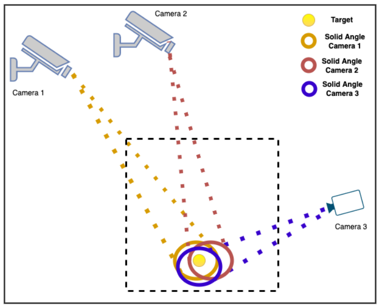

Fig. 5. Example of Federated Edge with Solid Angle Application

Figure 5 shows a descriptive picture of the reference scenarios. The diagram shows 3 cameras that film the same area identified by the dotted square with their respective solid angles on the reference target. The first parameters that influence the "goodness" of the shot and therefore the ability of the single device to recognize a target are the hardware characteristics and the distance of the Edge device from the target. Each device is able to calculate the distance from the target starting from its position, subsequently the federated devices can elect the "closest" device to the target to be recognized. However, this approach is not sufficient to solve the problems identified. Another parameter that becomes fundamental for our purpose is the solid angle. In [15] the mathematical laws to calculate it are reported, and it is explained how in combination with the other parameters it can be used to better recognize an image. In Figure 5 the solid angle is shown with different colors for each camera. The resulting "cone" identifies a particular area of the image on which the device can work. The fact that a target is in the solid angle of the 3 devices allows us to make an exhaustive comparison between their inability to identify it. This comparison can be made at the Edge because analyzing only the pixels in the solid angle greatly reduces the complexity of the image to be analyzed. Therefore, based on the data acquired from the backend system, the devices will know which of them can best identify the target audience. The fact that the devices are federated, and when they talk to each other, still allows for a noticeable improvement. A moving target can in fact change its exposure to the camera. This implies that for various reasons the device defined as "master" at a certain instant of time t understands that one of the devices of the federation can identify the target in a better way and therefore can pass it the title of master.

This approach also introduces the possibility of tracking a target within the Smart City. In fact, once the target has been identified and the system knows its position univocally, it is possible that the federation master assigns it a unique label. The consequence of this action is that if the target moves in another group of Federated Edge (but also in the area observed by a single device) it is clearly identified. The situation described is schematized in Figure 6. The diagram

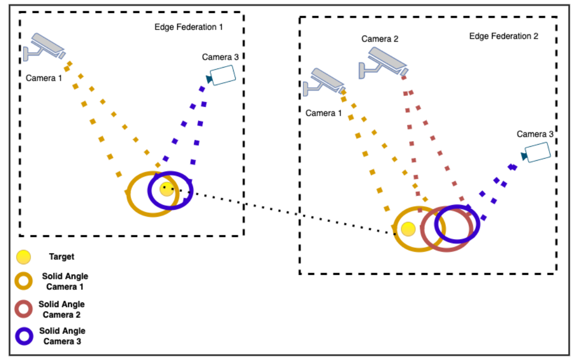

Fig. 6. Example of Target that move between 2 Federated Edge Environment.

shows a target that is initially identified in the solid angle of an Edge device that is part of a federation identified as "Edge Federation 1". Subsequently the target moves to the solid angle of another camera located in another federation of Edge identified as "Edge Federation 2". The ICT system described in Section V is able to keep track of the identified target and therefore allow the querying for its identification both in real time and in deferred time. This observation automates the video analysis processes carried out by humans to follow the movements of a target. Furthermore, the data collection means that the querying, for example of a license plate, allows to immediately obtain the positions and/or the routes taken without the need for long video analyzes to be carried out with human intervention. It is evident that the system described is linked to variable parameters. Furthermore, constant communication is required between the different federated devices and with the backend system. To manage this need it is necessary to clearly define the technology and the data model to be used. In this sense, a study has already been carried out in [5]. The NGSI-LD model that is described lends itself to configurations that can be defined for specific use cases and above all allows dynamic modification of the software on Edge devices as well. The definition of the models can be made preliminary and therefore the various needs encountered can be codified by making use of the NGSI-LD to find practical application.

V. Design Strategies for Federated Edge

Figure 7 shows a reference architecture for the implementation of the Federated Edge concept we presented in this paper. The fundamental elements of the system are the ***Edge devices***. They are computational nodes able to process video frames according to the specific requirements of the application in execution for the end user. They can correspond to smart cameras, if these are available and equipped with the necessary software for data analysis, or they can be implemented physically coupling a surveillance camera with an embedded device for data processing at the Edge. A group of Edge Devices involved in the tracking of the same target forms a ***Federated Edge***. Each Federated Edge is configured dynamically on the basis of the target inputs received from the application and Edge Nodes in the Federated Edge cooperate

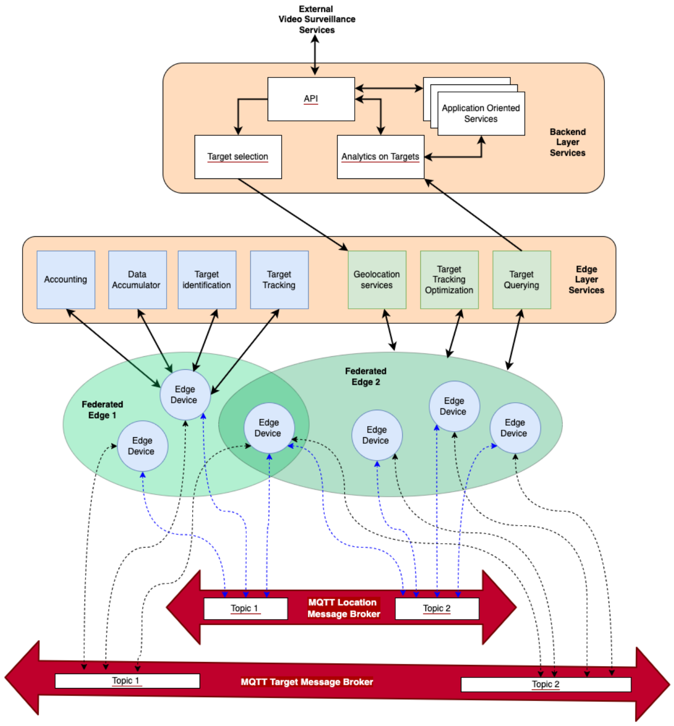

Fig. 7. General System Architecture

to provide the more significant information to the application. This means that not necessarily all the Edge Devices are part of a Federated Edge and, vice versa, one or more Edge Devices can take part to different Federated Edges at the same time if they are involved in the tracking of different targets in the same scene.

In the Federated Edge, it is necessary that each device can know (if it exists) "who is" the close device which shares the same scene and target. So, the device must be able to communicate with its neighbors [16]. To allow the Edge Devices to collaborate in the federation, it is necessary enabling a scalable and secure communication among them and exchange of data. To this aim, the Federated Edge exploit the Message Queuing Telemetry Transport (MQTT) standard, where topics are set up to allows devices to share information. In order to be part of a Federation, a device must:

- Physically share, through localization, the scene in which the target must be identified and then subscribe to a topic of the MQTT Location Message Broker (blue dotted arrow in Figure 7);
- Have the ability to recognize the target sought and then subscribe to a topic related to that target in the MQTT Target Message Broker (black dotted arrow in Figure 7)

The reference architecture is composed of 2 main layers of services: the ***Backend Layer Services*** and ***Edge Layer Services***.

The ***Backend Layer Services*** include the high level services to support external components (e.g., applications for Smart Cities or end-user data access requests) and they rely on the information provided by the Federated Edge. In particular, they include:

- *API*: it defines the API specifications for communicating with the external components, such as *External Video Surveillance Services* or proprietary video surveillance systems;
- *Target Selection*: it elaborates the target that is requested by the system together with the area of interest where the target has to be analyzed. These information are the input for the activities of the Federated Edge.
- *Analytics on Targets*: it collects data from the Federated Edge and implements specific analytics to create added value for the external components asking for target information.
- *Application Oriented Services*: they are additional services for enriching the value of data coming from the Federated Edge, such as classification of targets within the system as well as cross-relation of data on different targets.

The ***Edge Layer Services*** are software components executed at the Edge. They can be distinguished in services executed independently by each Edge Device and in services executed by the Edge Federation in a distributed and collaborative fashion. Services executed by each Edge Device in an independent way with respect to the other Edge nodes are:

- *Accounting*: it represents the component that manages the authentication of Edge devices and users who use the system;
- *Data Accumulator*: it is a component that takes care of the storage of data and/or information that the Edge devices eventually transmit. This component also has the function of interacting with any software modules that need to access the data history;
- *Target Identification*: it represents the target identification system on the Edge;
- *Target Tracking*: it defines the tracking methods of an Edge device within the federation.

Services executed by the Edge Federation in a distributed and collaborative fashion are:

- *Geo-location Services*: it uses the services described in [7] to uniquely identify a point on the earth if this has not already been identified within the system;
- *Target Tracking Optimization*: it optimizes target tracking between different federations. This function is delegated to the Federation Edge device which is in charge of the target;
- *Target Querying*: it searches for devices that can identify the target or that are tracking it at the time of the request;

Within the Federated Edge the device with the "ability to recognize a specific target of an Edge Device" with the higher "quality of information" is then identified. This device can be defined as Edge Master for federation. Edge Master is the device that is presumed to be able to identify and track the target and therefore can also decide if during the movement he loses his ability and can appoint another Edge Master.

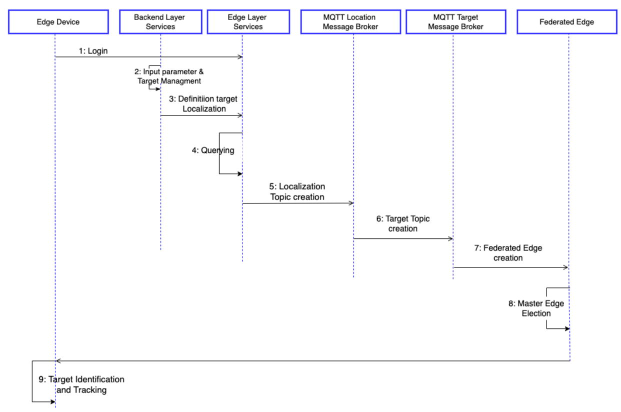

Fig. 8. Sequence Diagram.

For a better description of the system described, we refer to Figure 8. Each device that is part of the described system must be identified through accounting (Step 1 - Figure 8). Verification takes place in Edge Layer Services. The input of the target to be searched in the system and of the position takes place in the Backend Layer Services (Step 2 - Figure 8). The target model is identified in this component. The identification of the position takes place in the Edge Layer Services (Step 3 - Figure 8). This component has the ability to understand which cameras are suitable for target identification (Step 4 - Figure 8). At this point, the topics are created in the MQTT Server to put the Edge Devices in contact that are notified of the request. The Edge Devices must first subscribe to the topic related to the position (Step 5 - Figure 8) and only after, based on their ability to identify the target, can they subscribe to the topic concerning the target (Step 6 - Figure 8). The Federated Edge is defined among the Edge Devices subscribed to the topic of the MQTT Target Message Broker (Step 7 - Figure 8). Among the Edge Devices of the Federated Edge, the Master Edge is chosen on the basis of the definition of ”ability to recognize a specific target of an Edge Device” with the higher ”quality of information” (Step 8 - Figure 8). The chosen device will perform the required functions (Step 9 - Figure 8).

VI. Conclusion

The paper introduces the concept of Federated Edge defined as a solution to a real problem within Smart Cities. In particular, the system described defines the design principles of an ICT system useful for solving the following problems:

- given an uniquely identified area: understand which camera has the best ability to recognize a specific target;
- given a target: which camera is able to recognize it and at what quality level.

As a next phase of the proposed work, it will be necessary to implement the architecture described and validate the technologies proposed for its operation. For the evaluation of the system, its ability to create a Federated Edge in a time useful for identifying a target will be relevant. In addition, the ability of the devices to be part of multiple Federated Edges in parallel will be considered.

Acknowledgment

This work was co-funded by the Italian FSE, PON Research and Innovation 2014-2020 Axis I - Action I.1 ”Dottorati innovativi con caratterizzazione industriale”, the Italian MESMART project, CUP F41I18000230006 and by EU Horizon 2020 URBANITE project - grant agreement N° 870338. The authors thank Dr. Placido Accolla, Administrative Manager of the MESMART and URBANITE projects for the Municipality of Messina for his great support in our research activity.

References

[1] M. Mirmehdi, P. L. Palmer, and J. Kittler, “Towards optimal zoom for automatic target recognition,” GBR, Tech. Rep., 1997.

[2] H. Duan, Y. Zhang, and W. Sheng, “Image digital zoom based single target apriltag recognition algorithm in large scale changes on the distance,” in *2019 1st International Conference on Industrial Artificial Intelligence (IAI)*, 2019, pp. 1–6.

[3] M. Gao, R. Yu, A. Li, V. I. Morariu, and L. S. Davis, “Dynamic zoom-in network for fast object detection in large images,” 2017.

[4] A. Kolb, E. Barth, R. Koch, and R. Larsen, “Time-of-flight cameras in computer graphics,” *Computer Graphics Forum*, vol. 29, pp. 141 – 159, 03 2010.

[5] F. Martella, G. Parrino, G. Ciulla, R. D. Bernardo, A. Celesti, M. Fazio, and M. Villari, “Virtual device model extending ngsi-ld for faas at the edge,” in *2021 IEEE/ACM 21st International Symposium on Cluster, Cloud and Internet Computing (CCGrid)*, 2021, pp. 660–667.

[6] V. Lukaj, F. Martella, M. Fazio, A. Celesti, and M. Villari, “Trusted ecosystem for iot service provisioning based on brokering,” in *2021 IEEE/ACM 21st International Symposium on Cluster, Cloud and Internet Computing (CCGrid)*, 2021, pp. 746–753.

[7] Sicari, C. and Galletta, A. and Celesti, A. and Fazio, M. and Villari, M., “An Osmotic Computing Enabled Domain Naming System (OCE-DNS) for distributed service relocation between cloud and edge,” *Computers and Electrical Engineering*, vol. 96, 2021.

[8] W. Li, X. S. Feng, K. Zha, S. Li, and H. S. Zhu, vol. 1757, no. 1, p. 012003, jan 2021. [Online]. Available: https://doi.org/10.1088/1742-6596/1757/1/012003

[9] Y. Wang, G. Sun, and S. Guo, “Target detection method for low-resolution remote sensing image based on esrgan and redet,” *Photonics*, vol. 8, no. 10, 2021.

[10] A. Galletta, A. Ruggeri, M. Fazio, G. Dini, and M. Villari, “Mesmart-pro: Advanced processing at the edge for smart urban monitoring and reconfigurable services,” *Journal of Sensor and Actuator Networks*, vol. 9, no. 4, 2020.

[11] A. Mehrabi, M. Siekkinen, and A. Ylä-Jaaski, “Qoe-traffic optimization through collaborative edge caching in adaptive mobile video streaming,” *IEEE Access*, vol. 6, pp. 52 261–52 276, 2018.

[12] S. S. Kafıloğlu, G. Gür, and F. Alagöz, “Cooperative caching and video characteristics in d2d edge networks,” *IEEE Communications Letters*, vol. 24, no. 11, pp. 2647–2651, 2020.

[13] D. Wang, Y. Peng, X. Ma, W. Ding, H. Jiang, F. Chen, and J. Liu, “Adaptive wireless video streaming based on edge computing: Opportunities and approaches,” *IEEE Transactions on Services Computing*, vol. 12, no. 5, pp. 685–697, 2019.

[14] Y. Lamdan and H. Wolfson, “Geometric hashing: A general and efficient model-based recognition scheme,” in *[1988 Proceedings] Second International Conference on Computer Vision*, 1988, pp. 238–249.

[15] H.-S. Kim, R. Jain, and R. Volz, “Object recognition using multiple views,” in *Proceedings. 1985 IEEE International Conference on Robotics and Automation*, vol. 2, 1985, pp. 28–33.

[16] L. Carnevale, A. Ruggeri, F. Martella, A. Celesti, M. Fazio, and M. Villari, “Multi hop reconfiguration of end-devices in heterogeneous edge-iot mesh networks,” in *2021 IEEE Symposium on Computers and Communications (ISCC)*, 2021, pp. 1–6.