A joint Computer Vision and Reconfigurable Intelligent Meta-surface Approach for Interference Reduction in Beyond 5G Networks

Valeria Loscri, Anna Maria Vegni, Eros Innocenti, Romeo Giuliano, Franco Mazzenga

To cite this version:

Valeria Loscri, Anna Maria Vegni, Eros Innocenti, Romeo Giuliano, Franco Mazzenga. A joint Computer Vision and Reconfigurable Intelligent Meta-surface Approach for Interference Reduction in Beyond 5G Networks. HPSR 2021 - IEEE International Conference on High Performance Switching and Routing, Jun 2021, Paris, France. hal-03213861

HAL Id: hal-03213861
https://inria.hal.science/hal-03213861
Submitted on 30 Apr 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
A joint Computer Vision and Reconfigurable Intelligent Meta-surface Approach for Interference Reduction in Beyond 5G Networks

Valeria Loscrì  
FUN Research-lab  
INRIA  
Lille-Nord Europe  
valeria.loscri@inria.fr

Annamaria Vegni  
COMLAB Lab  
INRIA

Eros Innocenti, Romeo Giuliano  
Dept. of Engineering Science  
Guglielmo Marconi University  
eros@newtechweb.it  
r.giuliano@unimarconi.it

Franco Mazzenga  
Dept. of Enterprise Engineering  
University of Rome Tor Vergata  
mazzenga@ing.uniroma2.it

Abstract—Reconfigurable Intelligent Meta-surfaces (RIMs) are particular devices able to control and manipulate radio frequency wireless signals. This promising technology allows to improve the reliability of wireless networks, thanks to the capacity of reflecting the desired signals through appropriate phase shifts. The joint use of RIMs and Computer Vision (CV) technology is the main objective of this paper. This synergistic approach is used to correctly identify the specific configuration of a radiation pattern, to be used as input for computing optimal coding sequences of the RIM. Indeed, by the means of a CV algorithm it is possible to infer a connectivity graph related to a real scenario, where people is moving. The information about network nodes such as their distance, the relative position, etc. is used for feeding an intelligent logic, able to compute the optimal configuration for re-directing the signals towards a given receiver target node. Numerical results show the huge potentiality of this combined approach in terms of interference reduction. It has been observed that for high traffic load, it is possible to reduce the average interference in the network of 40%. Furthermore, an analysis including the positioning estimation error of the CV algorithm has been addressed, in order to consider how it affects the interference reduction. Results show that, even though there is an increasing effect of interference, when the error is accounted, the interference reduction impact is still important.

Index Terms—Reconfigurable Intelligent Meta-surface, Beam Steering, Computer Vision, B5G.

I. INTRODUCTION

Recently, the possibility to leverage visual data originated by different vision sensors permits to consider Computer Vision (CV) approaches as a key enabler technology for Beyond-5G wireless networks. One of the first contributions in the literature that proposed the combination of CV solutions with wireless networks, in particular with millimeter-wave radar, is the work in [2], where it was demonstrated how images can be leveraged for improving the distinction between lane lines at a short distance. Another work [4] proposes the integration of CV with wireless sensors to perform a high accurate indoor localization, through the use of Wi-Fi signals emitted by the smartphones for uniquely identifying the users. An interesting perspective on the CV exploitation is also presented in [5], where the authors introduce the concepts of View to Communicate (V2C) and Communicate to View (C2V).

In V2C, the visual data are used to enhance the prediction accuracy of a wireless channel, by accounting for blockages and for improving the localization in indoor environments. In C2V, the opposite approach is considered.

Another interesting concept, that has started to be considered by research community, is Reconfigurable Intelligent Meta-surface (RIM) [6]. RIM is a promising technology for improving the reliability of wireless networks, thanks to the capacity of reflecting the desired signals through appropriate phase shifts. Besides connectivity, RIM-based networks will allow to sense the environment to realize the vision of smart living in smart cities by providing context-awareness capabilities [7]. However, one of the most important factors hindering the development of RIM is represented by the fact that signals impinging on a RIM are often mixed with interfering signals, whose dynamic is unknown and unpredictable. Furthermore, the optimal configuration of a massive number of RIM units remains a big challenge. In [8] the authors propose the Env-Twin framework to enable automation of optimal control in Reconfigurable Intelligent Surfaces (RIS)-assisted wireless networks. Env-Twin is able to predict optimal RIM configuration for any new receiver location in the same wireless network.

Based on the recent advancements in the two domains i.e. (i) CV applied to wireless networks and (ii) RIM-based wireless communication systems, in this work we consider the combination of CV with RIM to be applied in a wireless context. Specifically, we consider a V2C approach based on the joint use of CV and RIMs. For the best of our knowledge, this is the first work presenting a synergistic approach of the CV technology with the RIM paradigm in a wireless context, aiming to improve the system performance, as well as to reduce interference on network nodes.

In a dynamic scenario the configuration of a meta-surface has to be computed and updated several times (e.g., in case the target receiver is moving or changing). Thus, we propose the design of a RIM-based network relying on a CV approach in order to reduce the interference of the wireless network. Interference, above all in Heterogeneous Networks (HetNets)
is a fundamental issue and how outlined by the authors in [3], the different techniques adopted right now, can improve the performance of the network, without mitigating dominant interference scenarios.

We exploit the CV to detect receivers, meta-surface and their inter-distance in a given area in order to properly feed RIMs allowing its fast configuration.

In order to be able to effectively “track” a receiver node, an efficient system has to be considered, that is not too much expensive in terms of information to be exchanged, while keeping a sufficient degree of accuracy, *i.e.*, estimation capability of the wireless channel. To the best of our knowledge, there are no similar approaches in literature. Results obtained show a drastic reduction of the interference in high traffic load scenarios, for increasing number of RIMs. The positioning error due to the CV sub-system is also integrated in the analysis, and even though there is a small deterioration of the performance, the effectiveness of the combined approach seems very promising for interference reduction.

The rest of the paper is organized as follows. In Section II we describe the specific assumptions and the model for the joint CV-RIM system. In Section III we detail the CV approach and describe a specific scenario considered. Section IV present the results of proposed approach, expressed in terms of interference reduction. Finally, conclusions are drawn at the end of this paper.

**II. SYSTEM MODEL**

In this paper, we consider an outdoor scenario with mobile nodes (*i.e.*, pedestrians) equipped with a wireless device acting either as a transmitter or receiver. We assume $N$ mobile transmitter nodes and $M$, with $M < N$, static nodes coated with RIM $^1$ (*i.e.*, RIM nodes), whose position is known. An example of meta-atom and metasurface operating at mmWave has been proposed in [1] and is shown in Figure 1.

We focus on beam steering applications and consider a mobile node to be tracked. The specific meta-surface connected with the potential receiver will be characterized with a new configuration, when the receiver will change its position, in order to track it.

**III. THE COMPUTER VISION APPROACH**

In order to provide data to RIMs, we considered a CV sub-system to compute distances between persons and metasurfaces in the area. It relies on FairMOT, which is a real-time multi-object tracking approach [10] that offers a high performance solution for the detection and the tracking of persons. As for the detection, a ResNet-34 backbone is adopted, combined with a more efficient version of Deep Layer Aggregation (DLA), which adds more skip connections.

---

$^1$We remind that a RIM is assumed as a matrix of meta-atoms, which are the basic units of the RIM.
between low and high level features. We selected FairMOT for this work because it is built on CenterNet, an anchor free model, which provides both good performances and low inference time. Since real-timeliness is required for feeding RIMs, two-stage detectors are not applicable. Indeed, DLA is used in order to obtain better accuracy with fewer trainable parameters, hence improving the overall speed of the network. The resulting model is then trained on the supervised CrowdHuman Dataset [11], which consists in a wide and meticulous annotated collection of 470k human instances with an average of 22.6 pedestrians per image. Each instance is characterized by a head bounding box and a human full-body bounding box. CrowdHuman is well suited for crowded scenarios where two or more persons may overlap each other causing malfunctioning during the detection phase. The entire dataset is split in 15k, 4.3k and 5k images for training, validation and test, respectively. We used the model offered by FairMOT, which took approximately 30 hours per training step using a pair of RTX 2080 Ti. The entire training phase consists of a total of 30 epochs using a batch size of 12, an input image size of 1088 × 608 and the learning rate value equal to $e^{-4}$ for the first 10 epochs and then $e^{-5}$. Standard data augmentation techniques have been used for training, such as scaling, rotation and colors manipulation.

Regarding the tracking stage, FairMOT uses an additional classification task, whose aim is to generate features that help to measure the affinity between persons. This goal is obtained adding an additional convolution layer made by 128 kernels on top of the backbone features. Video camera images are captured by the means of a streaming protocol such as Real Time Streaming Protocol (RTSP) or Web Real-Time Communication (WebRTC). The latter enables a latency lower than 500 ms.

Finally, the video stream is fed as an input of the described neural network which, at the end of a forward pass, outputs a list of detections including the associated tracking ID and the respective bounding box coordinates. The selected scenario has to be calibrated before starting the process. A minimum of four key points –the more, the better– must be chosen and referenced measuring the exact distance among them. An homography [12] matrix is then calculated based on the correspondence between these key points in the camera plane and the respective ones in a flat 2D plane. Assuming that the calibration is correctly performed, we are now able to precisely calculate the distance between pairs of bounding boxes or between meta-surfaces and bounding boxes, using the view projection and the exact distance from the ground level.

In In Fig. 4 it is reported how the CV video process works in a real environment. We considered a real video in Duc de Richelieu Monument in Odessa, Ukraine, recorded by a local webcam. The CV processor is able to identify several persons and to assess the distances between them and the RIMs. In Fig. 4, RIMs are supposed to be inserted in some elements present in the square e.g., the statue, the street lights, etc., whose positions are known.

### IV. RESULTS OF THE JOINT CV-RIM BASED SYSTEM

In this section, we evaluate the performance of the specific CV algorithm described in the previous section, by deriving the accuracy in detecting persons collected from a real video and the error in the estimation of the distances among persons and between persons and RIMs. We then study the behavior of the joint CV-RIM based system model and its impact in terms of interference reduction.

#### A. CV performance

Performance of the CV sub-system has been evaluated in terms of the Recall parameter, $\mathcal{R}$, which is defined as [13]:

$$\mathcal{R} = \frac{T_p}{T_p + F_n},$$

where $T_p$ is the number of correct (true) positive (i.e., a person has been correctly identified by the CV processor), and $F_n$ represents the number of cases in which the system failed to recognize a person (i.e., the human shape has not been identified as a person). During the CV evaluation process we focused on the recall parameter because it resulted more significant for this application, rather than the typical ones such as accuracy, precision and F1 score [13]. In fact, the correspondence between these key points in the camera plane and the respective ones in a flat 2D plane. Assuming that the calibration is correctly performed, we are now able to precisely calculate the distance between pairs of bounding boxes or between meta-surfaces and bounding boxes, using the view projection and the exact distance from the ground level.

#### A. CV performance

Performance of the CV sub-system has been evaluated in terms of the Recall parameter, $\mathcal{R}$, which is defined as [13]:

$$\mathcal{R} = \frac{T_p}{T_p + F_n},$$

where $T_p$ is the number of correct (true) positive (i.e., a person has been correctly identified by the CV processor), and $F_n$ represents the number of cases in which the system failed to recognize a person (i.e., the human shape has not been identified as a person). During the CV evaluation process we focused on the recall parameter because it resulted more significant for this application, rather than the typical ones such as accuracy, precision and F1 score [13]. In fact, the
most interesting cases are when a person is identified respect to when it is not identified by the CV model.

To evaluate the performance of the CV sub-system, we consider a video of about 2 minutes recorded by a webcam in Odessa at a rate of 25 fps. The processor analyzes one frame every 50 (i.e., every 2 s). During this time, it is able to detect persons and compute the distances from the RIMs. We obtain $R = 94.2\%$, showing the good behavior of this technique.

Concerning the precision in the distance computation, we suppose to have an error percentage $\Delta$ in measuring the distance between reference points for the considered environment. Then, we compare a grid of real points in the square with respect to the points estimated by the CV sub-system. The Probability Density Function (PDF) of the distance error $E_{\text{dist}}$ (i.e., the Euclidean distance between the real point with respect to the corresponding estimated point by CV) are reported in Fig. 5, for different values of the error percentage $\Delta$ w.r.t the reference points. Results show the accuracy in the distance computation of CV sub-system, providing an average error of 11 cm and a standard deviation of 5 cm for $\Delta = 1\%$. To show the sensitivity of the CV, we also report the unlikely case of $\Delta = 3\%$, which has an average error of 34 cm and a standard deviation of 15 cm.

### B. Connectivity Graphs and Interference Reduction

In this section we evaluate the interference reduction when the CV sub-system is combined with RIMs. In Fig. 6 and Fig. 7, we show the graphs of connected nodes, obtained with our CV sub-system explained above. The weights associated to the edges represent the interference a node exerts on the other nodes. It is worth to notice that we do not derive the weight information with the CV approach, but on the position of RIM in respect of the target receiver.

Such as an example, in Fig. 8 (a) the input graph, derived with the CV sub-system, is exploited by the processing unit managing the RIM for deriving the specific orientation of the radiation pattern. Specifically, node 11 is a RIM node tracking node 5, according to a radiation pattern as depicted in Fig. 8 (b). In this work, the design of the intelligent logic is out of scope, but we assume that it is capable to adaptively determine the coding sequences of the meta-surface in order to obtain the correct radiation pattern scheme as shown in Fig. 8 (b). In practice, the information gathered with the graph are used to infer the correct radiation pattern, in terms of specific parameters, such as the orientation of the main beam and its width. These parameters will change dynamically based on the variations of the scenario (e.g., the change of the receiver position). From Fig. 6 (a), we can redefine the interference as shown in Fig. 6 (b). Similar results are obtained for the interference in Fig. 7 (b), obtained from Fig. 7 (a), in case of low traffic load.

The main parameters considered for the numerical evaluation are summarized in the Table I.

We have considered two different types of evaluations. In the first approach depicted in Fig. 9, we consider a constant graph as derived in a specific time from the CV algorithm. The graph is constituted by 21 nodes and a variable number of meta-surfaces (ranging from 1 to 6). Two different situations are considered, the first one with a high traffic load (HTL, $I_{ht}$) from which we derive the normalized interference comprised between 0 and 1. The other situation considers a case of
low traffic load (LTL, $I_{lt}$) with a correspondent average low interference. The second type of evaluation refers to different graphs, with a fixed number of nodes $N$ and variable number of meta-surfaces $M$. We then define the ratio $\rho = M/N$. More specifically, we want to capture the different situations that can occur in a real world, where the mobile nodes can be deployed by occupying different positions, while the meta-surface position is kept fixed. In practice, we vary the number of receivers and their positions and we evaluate the average normalized interference reduction in this case.

In Fig. 9 we show the effect of an increasing number of meta-surfaces on the wireless system. The interference is computed in a normalized way, in order to evaluate the impact regardless to the specific communication technology. We evaluate the impact of the number of meta-surfaces on two different scenarios, one defined as High Interference and the other Low Interference. In the case of High Interference, we consider that nodes in the networks transmit with high probability and generate an interference close to 1. For “High Interference” and “Low Interference”, we assume that the CV approach is ideal, namely not affected by error. In this case, we can notice how for the high interference scenario, the increasing of the meta-surfaces objects, by re-directing the radiation pattern towards the target receiver, can drastically reduce the normalized interference from 0.85 to 0.45.

In order to evaluate the impact of the CV error in the position estimation on the interference mitigation, we have considered the results obtained in the previous section. In particular, $E_{dist}$ represents the error in position calculated by the CV sub-system with respect to the real user position. This error is proportional to the real distance $R$ [m] between the meta-surface and the target receiver, and the angle $\theta$ that the radiation pattern of the real position will form with the one of the estimated position, causing a radiation pattern generated by the meta-surface with a non-perfectly aligned steering. This can be mathematically expressed as

$$E_{dist} = R \cdot \theta.$$  \hspace{1cm} (2)

In order to account for error in distance computing by the CV sub-system, we have considered the initial error of reference points $\Delta = 1\%$, which causes the corresponding distance error $E_{dist}$ in Fig. 5. Fig. 9 depicts the interference behavior for different RIMs within the simulated graph. We
A reconfigurable meta-surface and the CV algorithm allows for the optimization of the interference reduction. In particular, we have assumed that certain nodes are coated with a RIM algorithm combined with a CV-based network. In this way, we have presented a framework constituted of the RIM. In order to evaluate the impact on the network, we have considered two different scenarios characterized with “high traffic load” and “low traffic load”. The results are very interesting in terms of potential reduction of the interference. In particular, by increasing the number of coated nodes, and in the “high traffic load” conditions, there is a potential reduction of 40%. Even in presence of an error estimation for the receiver position, the impact of the interference reductions is important.

In the future, we aim to integrate this framework with a real unit logic in order to evaluate the robustness in face of the error estimation of the CV.

ACKNOWLEDGEMENT

This work is partially supported by the Exploratory Action ETHICAM.

REFERENCES


