Machine Vision Aided Adaptive Beamforming Decision for IRS-Assisted Wireless Networks

M. Munawar*◊+, M. Guenach*+, and I. Moerman*◊+, *Senior Member, IEEE*
*IMEC Leuven, ◊IDLab, +Ghent University
Email: Muteen.Munawar@imec.be

***Abstract*—This study leverages machine vision to assist communication in wireless networks, with a specific focus on intelligent reflecting surface (IRS)-assisted wireless networks. Instead of depending on traditional schemes such as alternating optimization or semidefinite relaxation to maximize signal strength in an IRS-assisted network, which are computationally expensive and impractical, we use visual data to make low-complexity beamforming decisions for users. Our approach involves employing a ceiling camera with a fish-eye view, covering a wide communication area. The user within the network is initially detected using the YOLOv2 object detection method. Subsequently, we propose closed-form analytical expressions to determine the distances between the access point (AP), user, and IRS. Accounting for the non-uniform nature of the fish-eye image, we introduce a novel method to determine non-uniform pixel weightages using trigonometric techniques. Based on the calculated distances, we make beamforming decisions depending on the user's proximity to the AP or IRS. The proposed method significantly reduces computational complexity, making it nearly independent of the number of reflecting elements at the IRS. Simulation results indicate that the proposed approach exhibits extremely lower computational costs compared not only to conventional schemes such as alternating optimization and semidefinite relaxation-based convex solvers but also to low-complexity heuristic schemes.**

I. Introduction

A. Intelligent Reflecting Surface

The intelligent reflecting surface (IRS), a promising technology, is envisioned to contribute controllable reflecting channels to facilitate spectrum- and energy-efficient communication in next-generation wireless communication systems [1]. Specifically, the IRS is a metasurface containing several meta-elements that allow it to change its material properties in real-time, resulting in controllable reflecting channels [1]–[4]. To maximize the signal strength at the receiver in an IRS-assisted wireless network, a large number of reflecting elements at the IRS demand appropriate phase shift values, which is a computationally expensive task from the optimization point of view [1].

The literature is filled with beamforming schemes, i.e., the optimization of phase shifts, for IRS-assisted wireless networks [5]–[9]. Commonly used methods to optimize IRS phases can be divided into three categories: 1) iterative methods, i.e., alternating optimization (AO) [6], [7], 2) general-purpose convex solvers, i.e., CVX, [8] and 3) heuristic schemes [9]. Among these methods, AO and general-purpose CVX [10] methods provide better beamforming solutions, but the computational complexity of both AO and CVX, i.e., semidefinite relaxation (SDR), is very high [9], especially SDR-based solutions, and hence not suitable for the practical implementation of IRSs, which are supposed to be very low cost, low complexity, and low energy-consuming.

Recently, a heuristic approach named adaptive selection beamforming (ASB) [9] was proposed in the literature that is 1) non-iterative and 2) requires no general-purpose convex solvers, providing near AO and SDR performance with up to 70% lower computational cost than AO [1]. The scheme we propose [2] in this paper is further less computationally expensive than conventional ASB and becomes possible with the integration of machine vision and communication, as detailed in the following subsection.

B. Machine Vision and Communication

The integration of machine vision and communication theory holds the potential to enhance the performance of various applications [11]. This synergy can be explored in two main directions. The first one, known as communicate-to-view (C2V), involves leveraging RF signals-based data to enhance visual applications. While traditional computer vision relies on visible light data, utilizing RF signals with various frequencies offers distinct advantages, such as increased diffraction, which can enhance the resolution of imagery for distorted or occluded objects [12].

The second direction, view-to-communicate (V2C), focuses on utilizing visual data to improve communication networks. Given the high data speeds demanded by next-generation communication systems, there is a pressing need for computationally efficient algorithms that can rapidly estimate channels and process data. Machine vision emerges as a promising tool for achieving these goals. For instance, visual data can be employed to predict future channel behaviors and patterns of mobile blockages, thereby enhancing decision-making processes [13].

The proposed work in this paper primarily aligns with the V2C direction. Specifically, in this work, we exploit visual information to make a low-complexity beamforming decision for IRS-assisted wireless networks. The meaning of the beamforming decision is elaborated in later sections.

The research leading to these results has received funding from the European Community's Research Foundation Flanders (FWO) under grant agreement no A2582˙00˙01˙01 Project Strip-link MIMO FWO - 97707. (Corresponding author: Muteen Munawar.)

[1]The computational complexity of SDR is very high; therefore, it is not compared in numbers.

[2]In Section III, before we discuss the proposed scheme, we quickly explain what AO, SDR, and ASB methods are.

The remainder of the paper is structured as follows: Section II presents the mathematical model of our proposed system and outlines the problem formulation. In Section III, we discuss the existing methods and detail the proposed scheme. Additionally, Section IV evaluates the performance of the proposed scheme and compares it with closely related works. Finally, Section V concludes the paper.

Notation: Scalars are represented by italic letters, while vectors and matrices are denoted by boldface lowercase and uppercase letters, respectively. For a complex-valued vector $\mathbf{v}$, $(\mathbf{v})^H$ signifies the conjugate transpose, $|\mathbf{v}|$ represents the Euclidean norm, and $\operatorname{diag}(\mathbf{v})$ indicates a diagonal matrix with each diagonal element corresponding to the elements in $\mathbf{v}$. Regarding a complex number x, $|x|$ and $\arg(x)$ denote the absolute value and phase of x, respectively.

II. SYSTEM MODEL AND PROBLEM FORMULATION

The system model considered in this work is illustrated in Fig. 1. This image is captured by a ceiling pin-hole camera with a fish-eye view [14]. The IRS and access point (AP) are fixed at known locations. A user, i.e., a robot, moves within the area and communicates with the AP with the assistance of the IRS. The area in the image is partitioned into two regions. The center region comprises pixels with nearly uniform distance weights corresponding to the actual region. The non-uniform region consists of pixels with higher weight compared to the center pixels, attributed to the non-uniform view of the camera. The channels among the AP, IRS, and User are depicted in Fig. 2.

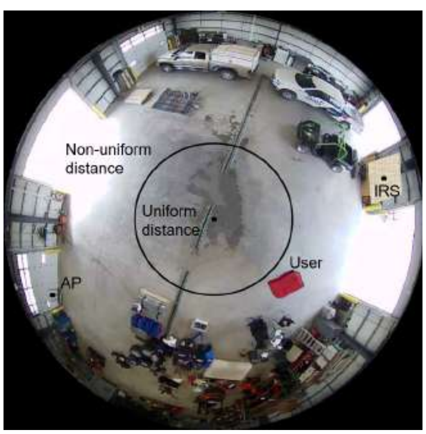

Fig. 1: System model (fish-eye view).

To formulate the problem, the communication network shown in Fig. 2 is simplified in Fig. 3. Specifically, we consider a communication network where an IRS, with N reflecting elements, aids a multi-antenna AP, having N_t transmit antennas, in communicating with a single-antenna user. Let $\mathbf{h}_d^H \in \mathbb{C}^{1\times N_t}$, $\mathbf{h}_r^H \in \mathbb{C}^{1\times N}$, and $\mathbf{G} \in \mathbb{C}^{N\times N_t}$ denote the baseband equivalent channels for the AP-User link, IRS-User link, and AP-IRS link, respectively [3]. The received signal at

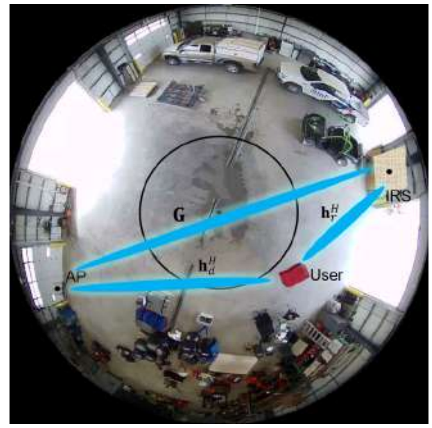

Fig. 2: System model with channels.

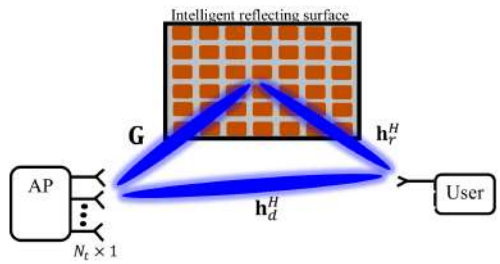

Fig. 3: Simplified system model.

the user is expressed as:

$$\begin{aligned} y &= \mathbf{h}_r^H \mathbf{\Phi} \mathbf{G} \mathbf{v} s + \mathbf{h}_d^H \mathbf{v} s + n \\ &= \left(\mathbf{h}_r^H \mathbf{\Phi} \mathbf{G} + \mathbf{h}_d^H\right) \mathbf{v} s + n, \end{aligned} \tag{1}$$

where n, $\mathbf{v}$, and s denote additive white Gaussian noise at the receiver with zero mean and σ^2 variance, the beamforming vector at AP, and the transmitted information symbol, respectively, whereas $\mathbf{\Phi}$ represents a diagonal matrix with the operations of reflecting elements on the diagonal, i.e., $\mathbf{\Phi} = \operatorname{diag}(\alpha_1 e^{j\phi_1},\, \alpha_2 e^{j\phi_2},\, \cdots,\, \alpha_N e^{j\phi_N})$, with $\alpha \in [0, 1]$ and $\phi \in [0, 2\pi]$ being the amplitude and phase changes of reflecting signals, respectively. The expressions $\mathbf{h}_r^H \mathbf{\Phi} \mathbf{G} \mathbf{v} s$ and $\mathbf{h}_d^H \mathbf{v} s$ in (1) indicate the signal received though AP-User link and AP-IRS-User link, respectively. More details on the system model can be found in [9], and are omitted here for brevity.

The received signal strength at the user, denoted by γ, is given by:

$$\gamma = \left|\left(\mathbf{h}_r^H \mathbf{\Phi} \mathbf{G} + \mathbf{h}_d^H\right) \mathbf{v}\right|^2. \tag{2}$$

To maximize (2), we need to optimize $\mathbf{v}$ and $\mathbf{\Phi}$ at the AP and IRS, respectively. Accordingly, the optimization problem is formulated as

$$\begin{aligned} \max_{\mathbf{v},\mathbf{\Phi}} \quad & \left|\left(\mathbf{h}_r^H \mathbf{\Phi} \mathbf{G} + \mathbf{h}_d^H\right) \mathbf{v}\right|^2 \\ \text{s.t.} \quad & \|\mathbf{v}\|^2 \le p, \\ & 0 \le \phi_n \le 2\pi, \ \ n = 1, \cdots, N, \end{aligned} \tag{P1}$$

where p represents the maximum transmission power budget at the AP. This optimization problem has a non-convex objective function concerning $\mathbf{\Phi}$ and $\mathbf{v}$, resulting in a non-convex problem. Common approaches, such as AO [7] and SDR [8],

[3]The discussion on the acquisition of channel state information is not within the scope of this work; it can be acquired using existing conventional schemes [15]–[17].

are employed in the literature. Although AO and SDR offer good performance, they come with higher computational costs [9]. In this work, we initially explore two simple beamforming solutions based on the heuristic ASB scheme [9] and subsequently use visual data to select one of them.

III. Machine Vision aided Adaptive Beamforming Selection

To initiate the detailed algorithm, we begin by discussing the conventional AO, SDR, and ASB schemes.

In AO-based schemes [7], the optimal $\mathbf{v}$ for the AP is selected using maximum ratio transmission (MRT):

$$\mathbf{v} = \sqrt{p}\frac{\left(\mathbf{h}_r^H\mathbf{\Phi}\mathbf{G} + \mathbf{h}_d^H\right)^H}{\left\|\mathbf{h}_r^H\mathbf{\Phi}\mathbf{G} + \mathbf{h}_d^H\right\|}, \tag{3}$$

whereas the optimal phase shifts for IRS, i.e., $\mathbf{\Phi}$, are selected such that the signal from the direct path and reflecting path coherently combine at the receiver:

$$\phi_n = \arg\left(\mathbf{h}_d^H\mathbf{v}\right) - \arg\left(h_{n,r}^H\right) - \arg\left(g_n^H\mathbf{v}\right),\ n = 1, \cdots, N, \tag{4}$$

where $h_{n,r}$, ϕ_n, and $\mathbf{g}_n^H$ denote the nth element of $\mathbf{h}_r$, nth element of $\mathbf{\Phi}$, and nth row of $\mathbf{G}$, respectively. Note that $\mathbf{v}$ and $\mathbf{\Phi}$ from (3) and (4) are dependent on each other, and hence solving them requires an iterative method known as AO (more details on AO in [7]).

In SDR-based schemes, the optimal $\mathbf{v}$ for the AP, i.e., (3), is inserted into the objective function of (P1), and the resulting expression is converted into an SDR program, which is further solved using a general-purpose convex solver, i.e., CVX. More details on such methods can be found in [6], [8].

In the ASB scheme, two simple solutions for $\mathbf{v}$ and $\mathbf{\Phi}$, denoted as $(\mathbf{v}_1, \mathbf{\Phi}_1)$ and $(\mathbf{v}_2, \mathbf{\Phi}_2)$, are derived as follows [4]:

$$\mathbf{v}_1 = \sqrt{p}\frac{\mathbf{h}_d}{\|\mathbf{h}_d\|} \tag{5}$$

$$\phi_{1n} = -\arg\left(h_{n,r}^H\right) - \arg\left(g_n^H\sqrt{p}\frac{\mathbf{h}_d}{\|\mathbf{h}_d\|}\right),\ n = 1, \cdots, N, \tag{6}$$

and

$$\mathbf{v}_2 = \sqrt{p}\frac{\mathbf{g}_n}{\|\mathbf{g}_n\|} \tag{7}$$

$$\phi_{2n} = \arg\left(\mathbf{h}_d^H\sqrt{p}\frac{\mathbf{g}_n}{\|\mathbf{g}_n\|}\right) - \arg\left(h_{n,r}^H\right),\ n = 1, \cdots, N. \tag{8}$$

Following the computation of $(\mathbf{v}_1, \mathbf{\Phi}_1)$ and $(\mathbf{v}_2, \mathbf{\Phi}_2)$ using ((5), (6)) and ((7), (8)), respectively, the ASB scheme calculates the corresponding total channel gains:

$$a_1 = \left|\left(\mathbf{h}_r^H\mathbf{\Phi}_1\mathbf{G} + \mathbf{h}_d^H\right)\mathbf{v}_1\right| \tag{9}$$

and

$$a_2 = \left|\left(\mathbf{h}_r^H\mathbf{\Phi}_2\mathbf{G} + \mathbf{h}_d^H\right)\mathbf{v}_2\right|. \tag{10}$$

Finally, the scheme compares a_1 and a_2, selecting $(\mathbf{v}_1, \mathbf{\Phi}_1)$ as the final solution if $a_1 \geq a_2$, and similarly chooses $(\mathbf{v}_2, \mathbf{\Phi}_2)$ if $a_2 > a_1$. This selection of $(\mathbf{v}_1, \mathbf{\Phi}_1)$ or $(\mathbf{v}_2, \mathbf{\Phi}_2)$ is known as ASB or beamforming decision. Note that the calculation of a_1 and a_2 is computationally expensive due to the large, complex matrix multiplications. Imagine a large value of N or N_t, leading to computationally expensive matrix-vector multiplications (details on the computational complexity of AO, SDR, and ASB are provided in [9]).

In this paper, we eliminate the computational cost associated with the calculation of a_1 and a_2. Specifically, we leverage visual data for this decision-making process, i.e., the selection of $(\mathbf{v}_1, \mathbf{\Phi}_1)$ or $(\mathbf{v}_2, \mathbf{\Phi}_2)$. We compare the SNR performance and computational cost with AO [8], SDR [7], and ASB [9]. The detailed work is divided into the following parts: 1) Detect the user's presence in an image and retrieve pixel coordinates. 2) Based on the non-uniform distances in the image, assign each pixel the corresponding actual distance of the field. 3) Based on the detection, find the distance from the person to the AP and IRS. 4) Based on the calculated distance, find a beamforming decision for the AP and IRS. 5) Evaluate the performance and computational complexity of the proposed work, comparing it with closely related works in the literature. The following subsections explain these parts in detail.

A. User detection using YOLOv2

First of all, we detect the user and obtain its location coordinates in the image. For this purpose, we utilize YOLOv2 [18], a single-stage real-time object detection model. To train YOLOv2, we create a dataset comprising a total of 160 images with scaled and rotated versions of the robot within them.

We use MATLAB to conduct the YOLOv2 training. Specifically, the training involves 30 epochs and a minibatch size of 6, and the training outcomes are illustrated in Fig. 4. The average precision is 0.96, as indicated in Fig. 5.

Training a YOLO v2 Object Detector for the following object classes:

* vehicle

Training on single CPU.
Initializing input data normalization.

Epoch	Iteration	Time Elapsed (hh:mm:ss)	Mini-batch RMSE	Validation RMSE	Mini-batch Loss	Validation Loss	Base Learning Rate
1	1	00:00:13	0.98	1.77	80.5857	3.1362	0.0010
4	50	00:04:32	1.63	1.47	2.6454	2.1564	0.0010
7	100	00:09:14	0.47	0.50	0.2104	0.2550	0.0010
10	150	00:13:58	0.32	0.44	0.1011	0.1939	0.0010
13	200	00:18:14	0.26	0.40	0.0672	0.1614	0.0010
16	250	00:22:36	0.22	0.33	0.0500	0.1076	0.0010
19	300	00:26:58	0.22	0.27	0.0492	0.0704	0.0010
22	350	00:31:18	0.18	0.34	0.0315	0.1167	0.0010
25	400	00:35:37	0.19	0.23	0.0368	0.0537	0.0010
29	450	00:40:04	0.16	0.23	0.0271	0.0515	0.0010
30	480	00:42:36	0.13	0.22	0.0171	0.0464	0.0010

Training finished: Max epochs completed.
Detector training complete.

Fig. 4: YOLOv2 Training.

To test the trained neural network, the outcome corresponding to an unseen sample image is depicted in Fig. 6, where the model detected the robot with a confidence score of 81.

B. Pixels weight and distance calculation

To calculate the distance, we first need to determine the actual weights of pixels in the image. To explain the proposed algorithm for determining the non-uniform pixel weights, for simplicity, we divide the image into two regions, as shown in Fig. 7, where (x_{IRS}, y_{IRS}), (x_{AP}, y_{AP}), (x_U, y_U), and (x_c, y_c) denote the locations of IRS, AP, user, and center, respectively. The center region outlined by a circle contains pixels with different weights compared to the remaining region. Pixels in the non-uniform region cover more distance

[4]Details are available in [9] and are omitted here for brevity.

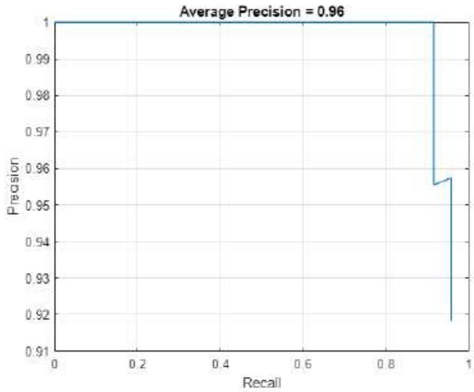

Fig. 5: Average precision.

Fig. 6: Trained neural network: unseen sample test.

than the center region due to the nature of the fish-eye view. To calculate the number of meters per pixel in the

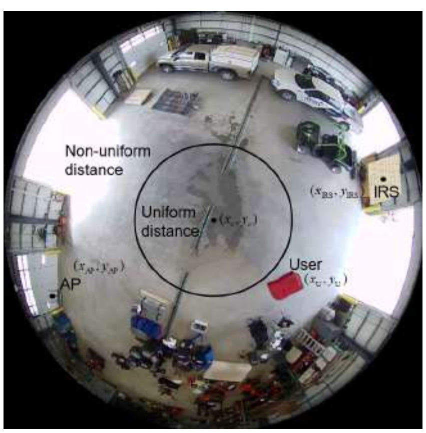

Fig. 7: Known pixel coordinates of AP and IRS.

uniform region, d_{uniform}, we use $\frac{d_{kl}}{\sqrt{(x_k-x_l)^2+(y_k-y_l)^2}}$. Here, (x_k, y_k) and (x_l, y_l) denote any two known locations (pixel coordinates) in the circle region with a corresponding known distance, d_{kl}. Using a similar method, we calculate the number of meters per pixel for non-uniform regions, $d_{\text{non-uniform}}$.

To calculate the total distance between AP and user, we need to know the number of pixels in the uniform and non-uniform regions. The key idea is to determine the intersection points of the center region (circle) and the straight line connecting AP-User.

The function for the uniform region can be determined by the following circle equation:

$$(x - x_c)^2 + (y - y_c)^2 = \text{radius}, \tag{11}$$

and the line function of the AP-User can be calculated as:

$$y - y_{\text{U}} = \frac{(y_{\text{AP}} - y_{\text{U}})}{(x_{\text{AP}} - x_{\text{U}})}(x - x_{\text{U}}). \tag{12}$$

Next, we solve (11) and (12) to find their intersection points. Depending on the number of intersection points, we can observe the following cases:

1) Two intersection points: Such a case is depicted in Fig. 8, where the user location results in two cross-section points. Correspondingly, the total distance d

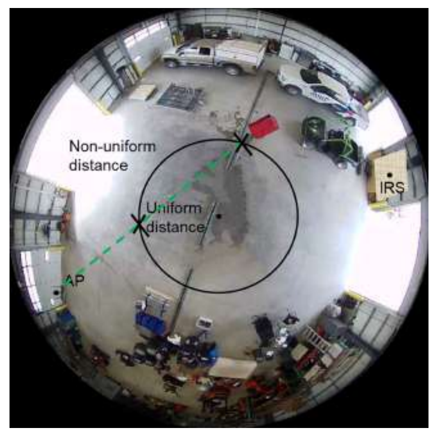

Fig. 8: Two intersection points.

can be calculated as follows:

$$\begin{aligned} d = &\{(q_{\text{A,U}} - q_{\text{P,P}}) \times d_{\text{non-uniform}}\} \\ &+ (q_{\text{A,U}} \times d_{\text{uniform}}), \end{aligned} \tag{13}$$

where $q_{\text{A,U}}$ and $q_{\text{P,P}}$ denote the number of pixels between AP and user, and the number of pixels between two cross-section points, respectively, whereas d_{uniform} and $d_{\text{non-uniform}}$ denote distance per pixel in uniform and non-uniform region of images, respectively.

2) One intersection point: This case results in two sub-cases. If the user is in the circle (Fig. 9), the distance is calculated by

$$\begin{aligned} d = &\{(q_{\text{A,U}} - q_{\text{U,P}}) \times d_{\text{non-uniform}}\} \\ &+ (q_{\text{U,P}} \times d_{\text{uniform}}), \end{aligned} \tag{14}$$

where $q_{\text{U,P}}$ denotes the number of pixels between the user and the cross-section point.

Otherwise, if there is only one pixel with non-uniform weight, i.e., a tangent cross-section, as shown in Fig. 10, the distance is given by

$$d = (q_{\text{A,U}} \times d_{\text{non-uniform}}) + d_{\text{uniform}} \tag{15}$$

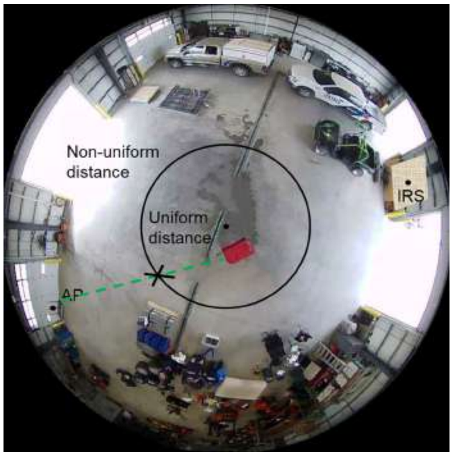

Fig. 9: One intersection point: within the circle.

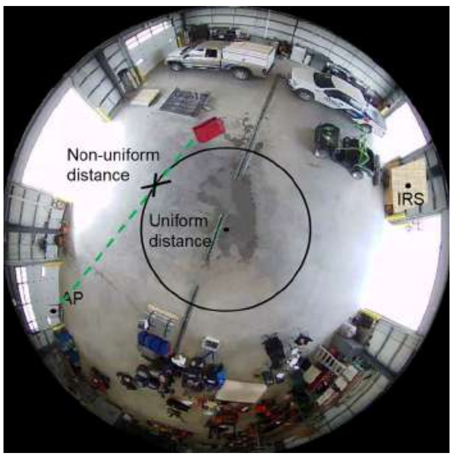

Fig. 10: One intersection point: tangent to the circle.

3) No intersection point: Such a case is shown in Fig. 11 and the distance is given by

$$d = q_{\mathrm{A,U}} \times d_{\mathrm{non-uniform}}. \tag{16}$$

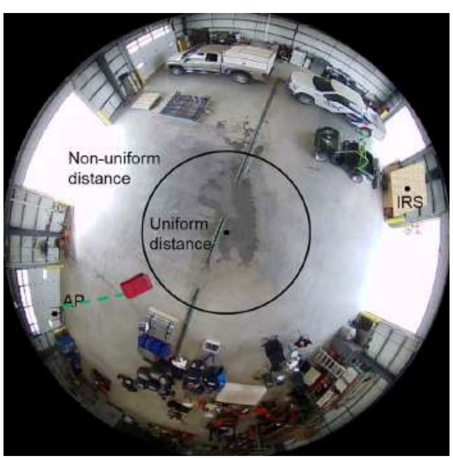

Fig. 11: No intersection point.

C. Beamforming Decision

After calculating the distance, we can make the adaptive beamforming decision using the following condition: If $d < d_i$, select $\mathbf{v}_1$ and $\mathbf{\Phi}_1$ using (5) and (6), respectively. Otherwise, choose $\mathbf{v}_2$ and $\mathbf{\Phi}_2$ using (7) and (8), respectively, where d_i is learned during YOLOv2 training.

IV. Simulation results

A. Simulation Setup

The simulation setup is depicted in Fig. 7, where the AP-IRS distance is 50 meters, the total transmit power p at the AP is 36 dBm, and the noise power (σ^2) at the user is -96 dBm. Further details on simulation parameters are provided in Section IV of [9]; hence, omitted here for brevity.

B. SNR Performance

For $N_t = 8$ and $N = 100$, Fig. 12 illustrates the real-time evaluation of SNR performance for 50 images, where each image represents a geographically random location of the user (i.e., x-axis of Fig. 12). In most cases, the SNR performance is nearly identical to AO, SDR, and ASB. However, at a few locations where the detector could not detect the object, the proposed scheme shows negligible performance degradation compared to the other schemes.

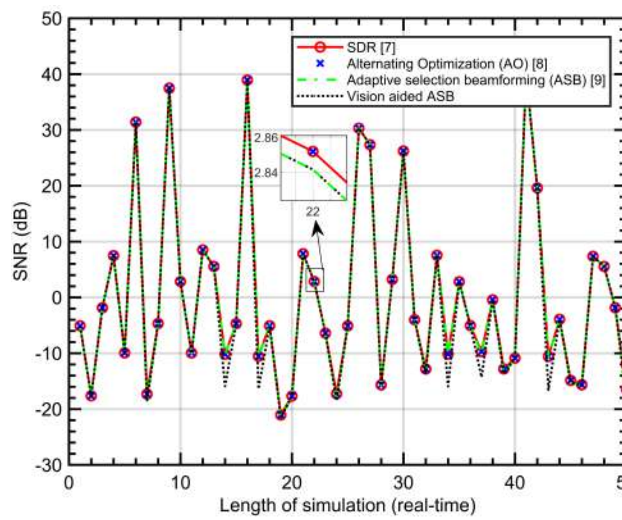

Fig. 12: SNR performance for various benchmark schemes.

C. Computational Cost

Next, in Fig. 13, we plot the computational cost[5] versus the number of reflecting elements. It can be observed that the proposed scheme's computational cost is almost independent of the size of the IRS. Specifically, for $N = 200$, $N = 900$, and 3000, the proposed scheme achieves 64%, 91%, and 99% lower computational cost, respectively, compared to AO. Similar reductions can be seen in comparison with ASB. The computational complexity of SDR is much higher, i.e., several

[5]The computational cost is calculated using MATLAB R2022a in terms of the average CPU running time on an AMD Ryzen R7-5800H CPU @ 3.20GHz and 16 GB of RAM.

seconds, even for small values of N; hence, it is not plotted in Fig. 13.

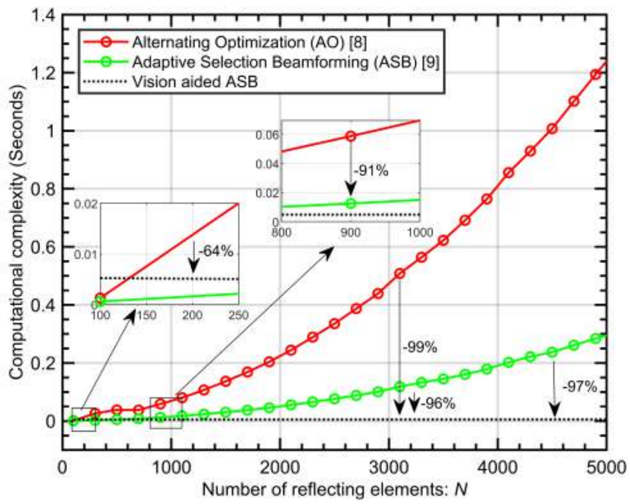

Fig. 13: Computational complexity versus the number of reflecting elements.

It's essential to acknowledge the possibility of more efficient and effective approaches for implementing the proposed scheme at various stages. For instance, in user detection, utilizing a custom neural network designed specifically for the application's requirements could potentially achieve greater accuracy while reducing computational complexity. Similarly, exploring alternative methods, such as angle-based calculations for distance estimation, might present advantages compared to the method detailed in Section III-B.

Nevertheless, irrespective of the methodologies chosen, the proposed scheme demonstrates considerable potential for practical applications, as depicted in Figure 13.

V. Conclusion

The intelligent reflecting surface is a potential candidate to realize the next generation of communication networks. However, due to the large number of reflecting elements, the computational cost to calculate the optimal solution for all elements is very high. In this work, we use visual data to make a low-cost beamforming decision for such networks. Specifically, we determine two low-cost and suboptimal solutions. Then, we use the visual data to determine the location of the user with respect to the AP and IRS. Based on the distance, we select one of the simple solutions as the optimal solution.

Despite a minor performance degradation that is deemed negligible, the proposed scheme significantly reduces computational costs compared to state-of-the-art existing methods, namely AO, ASB, and SDR, particularly when the number of reflecting elements is large, i.e., when $N > 100$. For instance, when $N = 200$, 900, and 3000, the proposed scheme achieves 64%, 91%, and 99% lower computational cost, respectively, compared to AO.

The proposed scheme is readily applicable to detect multiple users in the network and make informed decisions based on visual data. For example, in addition to the beamforming decision, visual data can aid in predicting channel behaviors and signal blockage patterns.

References

[1] B. Zheng, C. You, W. Mei, and R. Zhang, "A survey on channel estimation and practical passive beamforming design for intelligent reflecting surface aided wireless communications," *IEEE Commun. Surv. Tutor.*, vol. 24, no. 2, pp. 1035–1071, 2022.

[2] S. Abeywickrama, R. Zhang, Q. Wu, and C. Yuen, "Intelligent reflecting surface: Practical phase shift model and beamforming optimization," *IEEE Trans. Commun.*, vol. 68, no. 9, pp. 5849–5863, 2020.

[3] J. Y. Dai, W. Tang, M. Z. Chen, C. H. Chan, Q. Cheng, S. Jin, and T. J. Cui, "Wireless Communication Based on Information Metasurfaces," *IEEE Trans. Microw. Theory Techn.*, vol. 69, no. 3, pp. 1493–1510, 2021.

[4] S. Sun, W. Jiang, S. Gong, and T. Hong, "Reconfigurable linear-to-linear polarization conversion metasurface based on PIN diodes," *IEEE Antennas Wirel. Propag. Lett.*, vol. 17, no. 9, pp. 1722–1726, 2018.

[5] S. Gong, X. Lu, D. T. Hoang, D. Niyato, L. Shu, D. I. Kim, and Y.-C. Liang, "Toward smart wireless communications via intelligent reflecting surfaces: A contemporary survey," *IEEE Commun. Surv. Tutor.*, vol. 22, no. 4, pp. 2283–2314, 2020.

[6] Q. Wu and R. Zhang, "Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase shifts," *IEEE Trans. on Commun.*, vol. 68, no. 3, pp. 1838–1851, 2019.

[7] ——, "Intelligent reflecting surface enhanced wireless network: Joint active and passive beamforming design," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2018, pp. 1–6.

[8] ——, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5394–5409, 2019.

[9] M. Munawar and K. Lee, "Low-Complexity Adaptive Selection Beamforming for IRS-Assisted Single-User Wireless Networks," *IEEE Trans. Veh. Technol.*, vol. 72, no. 4, pp. 5458–5462, 2023.

[10] M. Grant and S. Boyd, "CVX: Matlab Software for Disciplined Convex Programming," [online] Available:http://cvxr.com/cvx.

[11] T. Nishio, Y. Koda, J. Park, M. Bennis, and K. Doppler, "When wireless communications meet computer vision in beyond 5G," *IEEE Communications Standards Magazine*, vol. 5, no. 2, pp. 76–83, 2021.

[12] H. H. Zhang and R. S. Chen, "Coherent processing and superresolution technique of multi-band radar data based on fast sparse bayesian learning algorithm," *IEEE Transactions on Antennas and Propagation*, vol. 62, no. 12, pp. 6217–6227, 2014.

[13] Y. Tian, G. Pan, and M.-S. Alouini, "Applying deep-learning-based computer vision to wireless communications: Methodologies, opportunities, and challenges," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 132–143, 2020.

[14] Westman Security, "Westman Security - Fisheye Camera," https://www.westmansecurity.ca/fisheye, accessed: February 1, 2024.

[15] D. Mishra and H. Johansson, "Channel estimation and low-complexity beamforming design for passive intelligent surface assisted MISO wireless energy transfer," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process. (ICASSP)*, 2019, pp. 4659–4663.

[16] X. Ma, Z. Chen, W. Chen, Z. Li, Y. Chi, C. Han, and S. Li, "Joint channel estimation and data rate maximization for intelligent reflecting surface assisted terahertz MIMO communication systems," *IEEE Access*, vol. 8, pp. 99 565–99 581, 2020.

[17] Z. Wang, L. Liu, and S. Cui, "Channel estimation for intelligent reflecting surface assisted multiuser communications: Framework, algorithms, and analysis," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6607–6620, 2020.

[18] J. Sang, Z. Wu, P. Guo, H. Hu, H. Xiang, Q. Zhang, and B. Cai, "An improved YOLOv2 for vehicle detection," *Sensors*, vol. 18, no. 12, p. 4272, 2018.