

# Toward Global Sensing Quality Maximization: A Configuration Optimization Scheme for Camera Networks

Xuechao Zhang\*, Xuda Ding\*, Yi Ren<sup>†</sup>, Yu Zheng<sup>†</sup>, Chongrong Fang\*, and Jianping He\*

**Abstract**—The performance of a camera network monitoring a set of targets depends crucially on the configuration of the cameras. In this paper, we investigate the reconfiguration strategy for the parameterized camera network model, with which the sensing qualities of the multiple targets can be optimized globally and simultaneously. We first propose to use the number of pixels occupied by a unit-length object in image as a metric of the sensing quality of the object, which is determined by the parameters of the camera, such as intrinsic, extrinsic, and distortion coefficients. Then, we form a single quantity that measures the sensing quality of the targets by the camera network. This quantity further serves as the objective function of our optimization problem to obtain the optimal camera configuration. We verify the effectiveness of our approach through extensive simulations and experiments, and the results reveal its improved performance on the AprilTag detection tasks. Codes and related utilities for this work are open-sourced and available at <https://github.com/sszxc/MultiCam-Simulation>.

## I. INTRODUCTION

### A. Background and motivation

Cameras play important roles in society surveillance [1], environmental monitoring [2], and scene perception, localization and mapping in robotics [3], etc. Compared to single-camera systems, camera networks tend to have a larger field of view and more viewing angles for each object, leading to stronger anti-interference ability and better detection performance [4], [5]. Fig. 1 shows a typical application for camera networks as drones tracking ground vehicles with cameras. A typical task of camera networks is to provide adequate coverage for the targets [6], [7]. However, meeting the qualitative coverage requirement does not necessarily ensure the desirable sensing quality of the targets in the scene. The distortion inevitably introduced in the imaging process of cameras depends on the lens's physical properties, the camera's parameters, viewing angle and the objects' relative position. Different configurations of the camera network may result in different sensing quality of the targets.

Configuring a camera network properly to cover multiple targets while ensuring desirable sensing quality is challenging due to the vast amount of camera parameters involved and a lack of widely accepted sensing performance metric [8]. In this paper, we propose a straightforward metric that captures the coverage and sensing quality of objects for a

This work is supported by the CIE-Tencent Robotics X Rhino-Bird Focused Research Program. The authors would like to thank Dr. Xiaoming Duan, Han Wang and Hao Jiang for their helpful advice on technical issues.

\*: the Dept. of Automation, Shanghai Jiao Tong University, Shanghai, China. email: {zhang.xc, dingxuda, crfang, jphe}@sjtu.edu.cn.

<sup>†</sup>: the Tencent Robotics X Lab, Shenzhen, China. email: {evanyren, petezheng}@tencent.com

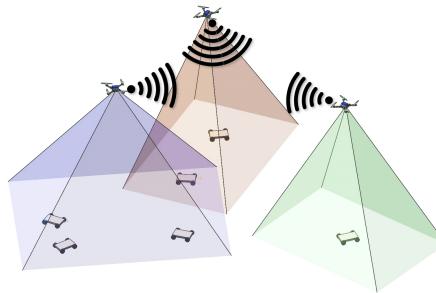


Fig. 1: A network of three cameras mounted on the drones for tracking autonomous ground vehicles.

camera network, based on which we automate the search for the configuration that optimizes the proposed metric.

### B. Related Work

Recent surveys [9] [10] summarize the historical developments in the area of camera network coverage. Globally optimal configuration for such systems, similar to the *Art Gallery Problem* [11] and *Watchmen Tour Problem* [12], has been proved to be NP-hard, and using general optimization methods could lead to sub-optimal solutions. Recent literature reports several approaches that have been proposed to solve this problem.

In [13] [14], the authors focus on complex indoor environments and present a method for pan-tilt-zoom (PTZ) camera reconfiguration. The proposed method considers the targets, camera distortion, and environment illumination, and a particle swarm optimizer gives the solution. However, the performance of the method relies on complete modeling of the environment, which is generally very difficult to obtain. The authors in [15] take into consideration the realistic constraints of computer vision and aim to achieve a balance between coverage and resolution for multiple cameras. They split the 2D plane into polygons with equal areas and use a greedy algorithm to segment the polygons for a simplified camera model and determine the cameras' configuration.

Recently, a class of gradient methods based on Voronoi diagrams has been widely used to solve the problem of sensor allocation. In [16], the authors focus on a 2D convex environment and propose a sensing quality metric using prior knowledge of the imaging process. They successfully use a provably correct greedy algorithm to configure cameras for given event distribution.

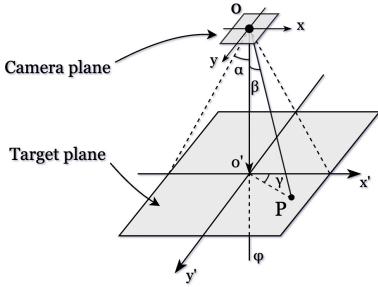


Fig. 2: Camera symbol definition.

### C. Contributions and Organization

In this paper, we build an automatic reconfiguration system for camera networks that significantly improves collective sensing performance. The main contributions are as follows.

- We propose a new metric that measures the coverage and sensing quality of the targets for a camera. This metric is easy to compute and is shown to be effective in simulation and experiments.
- We propose a novel model to quantitatively describe the effect of the camera distortion on sensing quality, which is suitable for any pre-calibrated cameras and targets in 3D space.
- We verify the reliability and usability of the strategy through sufficient experiments both in virtual and physical environments.

The rest of the paper is organized as follows. Section II explains the relationship between sensing quality and the number of pixels occupied by the target, and presents the proposed sensing quality model for a multi-camera system. In Section III, we illustrate the effectiveness of the proposed reconfiguration model using simulations and experimental results, respectively. Section IV concludes with a summary of our contributions and a brief discussion of future directions.

## II. PROBLEM STATEMENT AND FORMULATION

A general camera network consists of several heterogeneous cameras that may have distinct properties. Dynamic configuration adjustment of the camera network must be performed collectively to ensure that the entire network of cameras is globally optimized concerning the specified task or performance requirements. The sensing process is camera-specific and mutually exclusive, but when the cameras' fields of view overlap the fusion results should be calculated jointly.

As shown in Fig. 2, we consider a model for a camera  $C_i \in \mathbb{C}$  for  $i \in \{1, 2, \dots, N_c\}$  with a conic field of view, whose configuration is specified by its fixed position  $o_i \in \mathbb{R}^3$ , variable optical-axis direction  $\phi_i \in \mathbb{R}^3$ , angle of view  $2\alpha_i \in (0, \pi)$  and maximum radial resolution of  $w_i$  pixels. The position of the target  $T_j \in \mathbb{T}$  for  $j \in \{1, 2, \dots, N_t\}$  in the sensing task is denoted by  $p_j \in \mathbb{R}^3$ .

Let  $o'_i$  be the point such that  $\overrightarrow{o'_i p_j} \perp \phi_i$ . The plane that is parallel to the camera plane and contains  $o'_i$  is the target plane. Let  $\beta_j^i$  be the angle between  $\overrightarrow{o'_i p_j}$  and  $\phi_i$ , and  $\gamma_j^i$  be the

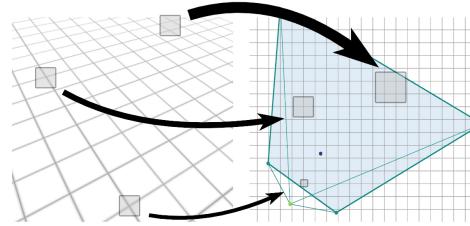


Fig. 3: Perspective comparison: (left) image space and (right) world space. The farther away from the lens, the larger the area described by a single pixel, and vice versa.

angle between  $\overrightarrow{o'_i p_j}$  and  $\overrightarrow{o'_i x'}$ . If  $\hat{p}_j$  is the estimated position for  $T_j$ , then the sensing error is defined by

$$\epsilon_j = \|p_j - \hat{p}_j\|_2. \quad (1)$$

Counting the number of pixels a target occupies in the image before correcting distortion is an intuitive description of the sensing quality. To illustrate the idea, suppose the target with the length of  $l$  occupies  $n$  pixels in the image. Moreover, let the actual length represented by each pixel on the target plane be  $k$ . Then the measurement  $l'$  has an error  $|l - l'| < 2k$ . The reason is that each pixel on complementary metal-oxide-semiconductor (CMOS) measures the average brightness information through integration [17], and the error caused by sampling will make both ends of the target captured as an uncertain pixel. It shows that when the same target occupies more pixels, the sensor has richer information to describe the target, making the downstream object detection more accurate. In other words, a shorter actual length represented by each pixel results in a smaller relative error. The error of positioning  $\epsilon_j$  is bounded by

$$\epsilon_j < k. \quad (2)$$

### A. Single-Camera Model

This section discusses how the number of pixels occupied by the target as a metric of sensing quality is related to two critical parameters, i.e., the camera's perspective factor and the distortion factor.

1) *Perspective*: Perspective primarily depends on the angle of view and focal length which considerably deteriorates the sensing quality. Objects appear smaller as their distance from the observer increases. In imaging process, the actual distance between the camera and the object has a great influence on the number of pixels that the object occupies, illustrated in Fig. 3. It is shown that the number of pixels occupied by an object with unit length is inversely proportional to the distance between the camera and the object. To improve sensing quality, the camera should be placed closer to the target.

Suppose that camera  $C_i$  performs the task of tracking target  $T_j$  and assume only perspective factor is considered. According to [18], the sensing quality here can be denote as

$$q_p \triangleq \frac{2|\overrightarrow{o'_i p_j}| \cos \beta_j^i \tan \alpha_i}{w_i}. \quad (3)$$

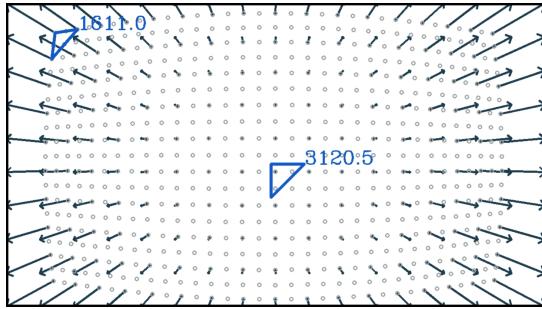


Fig. 4: The arrows indicate the mapping relationship of the pixels during the distortion correction, and the blue triangles represent the origin areas on the sensor occupied by two triangles of same size after correction. In this example, there is 48.4% less information at the edge of the image than in the center.

2) *Distortion*: Distortions can be introduced at the image acquisition and processing stages. With a specific lens and perspective, distortions are inherent in the captured videos which can seriously affect the achievable results [19]. In the pinhole camera model system, a widely-used mathematical formulation to correct camera distortions is the Brown-Conrady model [20] which includes radial and slight tangential distortion.

$$\begin{aligned} x' &= xK_r + 2s_1xy + s_2(r^2 + 2x^2), \\ y' &= yK_r + s_1(r^2 + 2y^2) + 2s_2xy, \end{aligned} \quad (4)$$

where  $r^2 = x^2 + y^2$  and

$$K_r = \frac{1 + k_1r^2 + k_2r^4 + k_3r^6}{1 + k_4r^2 + k_5r^4 + k_6r^6}. \quad (5)$$

In (4),  $(x, y)$  and  $(x', y')$  stands for the points before and after the distortion, respectively.  $k_1, k_2, k_3, k_4, k_5$ , and  $k_6$  are radial distortion coefficients.  $s_1$  and  $s_2$  are tangential distortion coefficients. These coefficients can be easily calculated by using classical calibration algorithms [21].

Distortion correction is necessary for performing detection, positioning, and recognition tasks. The distortion correction process of an image is a non-linear mapping, which does not bring any new information about the targets as the differences in pixel density are compensated by interpolation algorithms. After the stretching, the effective information at

the edge and the center of the corrected image has a huge gap, as shown in Fig. 4.

Furthermore, the model coefficients for distortion correction can also be used to evaluate the effective pixel density obtained by the sensor during the imaging process. The pixels are stretched in different proportions in the directions of x and y. Denote  $x'_t, x'_{t+1}$  as two adjacent pixels after distortion correction, then  $|x_t - x_{t+1}|$  represents the original information provided by the sensor on x direction. According to (4), the influence of distortion factor can be written as the derivative of  $x', y'$ . The pixel density is the product of the distortion factor in the directions of x and y. Therefore, for camera  $C_i$  and target  $T_j$ , we use  $q_d$  to denote the influence factor of distortion on pixel density.

$$\begin{aligned} x' &= \frac{\tan \beta_j^i \cos \gamma_j^i}{\tan \alpha_i}, \\ y' &= \frac{\tan \beta_j^i \sin \gamma_j^i}{\tan \alpha_i}, \\ q_d &\triangleq \frac{\partial x'}{\partial x} \frac{\partial y'}{\partial y}. \end{aligned} \quad (6)$$

Thus, the influence of different types of lenses can be described uniformly, no matter how they distort lights. This constraint indicates that the camera should be configured to align its optical axis with the object.

Finally, combining (3) and (6) yields a comprehensive description of the sensing quality for a single camera as

$$Q_j = q_p q_d > \epsilon_j. \quad (7)$$

An illustration of this model is shown in Fig. 5.

### B. Multi-Camera Model

To extend the single-camera model to multiple cameras, some modifications need to be made to take into account the information fusion. Now consider two cameras  $C_a, C_b$  observe the same target, for example, and  $Q_a, Q_b$  (let  $Q_a < Q_b$ ) denote the actual length represented by the unit pixel at the target location of the two cameras. Fusing data from the two cameras allows for more precise measurements than sensing with a single camera. More specifically, the fused result is the intersection of the origin range, that is  $\epsilon_j \in [0, Q_a]$  considering all possibilities of each camera's relative

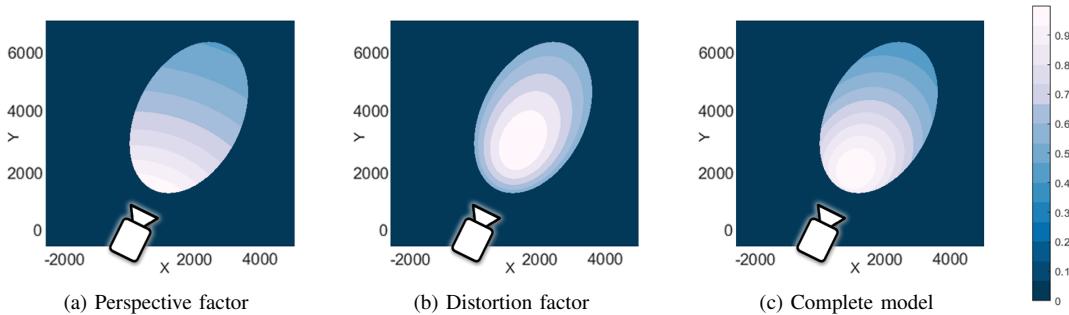


Fig. 5: The sensing quality on the ground plane by a certain camera mounted at  $(0, 0, 3000)$ mm. The XY axis represents the position, and the lighter the color, the higher the quality.

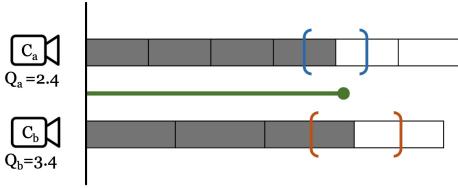


Fig. 6: A one-dimensional example for fusion: The task is to estimate the length of the green line, and the fused result is the intersection of two cameras with different  $Q_j$ .

position and focal length. A one-dimensional example for intuitive demonstration is given in Fig. 6. Suppose that the overlap between cameras occurs randomly. The expectation of the error bound can be written as

$$\bar{Q}_j = \frac{1}{1/Q_a + 1/Q_b}. \quad (8)$$

Similarly, it can be extended to the camera network of arbitrary scale. For a camera  $C_i$  and a target  $j$ , use  $V_{ij}$  to denote the visibility.

$$V_j^i = \begin{cases} 1, & \beta_j^i \leq \alpha_i \\ 0, & \beta_j^i > \alpha_i. \end{cases} \quad (9)$$

Then, the fusion of every cameras' estimation is written as

$$\bar{Q}_j = \frac{1}{\sum_{C_i \in \mathbb{C}} V_j^i / Q_j^i} > \epsilon_j. \quad (10)$$

### C. Optimization

The tasks of a camera network in the real life vary greatly, resulting in different optimization problems. One of the most common goals is that each target is covered by the field of view of at least one camera and the average sensing quality

---

#### Algorithm 1: Camera Network Configuration

---

**Input:** camera angle of view  $\alpha_i$ , camera distortion coefficients, camera resolution  $w_i$  pixels, user constraints on  $o_i, \phi_i$ .

**Output:** optimal camera configuration, targets' position  $p_j$ .

```

1 Initial scan to get targets' position  $p_j$ ;
2 while true do
3   Solve the optimization problem (11) and get
    $o_i, \phi_i$ ;
4   for  $C_i \in \mathbb{C}$  do
5     Capture the image of  $C_i$ ;
6     Run AprilTag detection algorithm to estimate
      $p_j$ ;
7   end
8   for  $T_j \in \mathbb{T}$  do
9     if target  $T_j$  gets more than one estimation
     from the camera network then
       | Update  $p_j$  to the fusion of estimations;
10    end
11  end
12 end
13 end
```

---

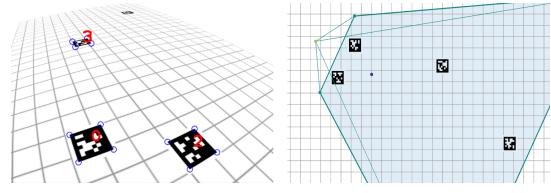


Fig. 7: AprilTag detection task in the virtual environment, (left) image space and (right) world space. The farthest tag was not detected successfully because it occupied few pixels on the image.

of all targets is maximized. The objective function based on the proposed sensing quality metric (10) can be written as

$$\begin{aligned} \text{Minimize} \quad & \frac{1}{N_t} \sum_{T_j \in \mathbb{T}} \bar{Q}_j \\ \text{subject to} \quad & \sum_{C_i \in \mathbb{C}} V_j^i \geq 1, \forall T_j \in \mathbb{T}. \end{aligned} \quad (11)$$

Further, with the formulation of sensing quality, different optimization problems can be constructed to deal with other application requirements. For example, we can optimize the worst-case sensing quality as

$$\begin{aligned} \text{Minimize} \quad & \max_{T_j \in \mathbb{T}} \bar{Q}_j \\ \text{subject to} \quad & \sum_{C_i \in \mathbb{C}} V_j^i \geq 1, \forall T_j \in \mathbb{T}. \end{aligned} \quad (12)$$

Based on (11), Algorithm 1 is proposed for camera network configuration systems. Apparently, Algorithm 1 can be used to solve (12) with the modification of Line 3.

## III. PERFORMANCE EVALUATION

### A. Simulation

In this section, we provide simulation experiments demonstrating the claimed strength of proposed sensing quality model for the reconfiguration of heterogeneous camera networks, e.g., PTZ camera networks or drone mode (cameras carried by unmanned aerial vehicles). In order to evaluate the sensing quality, we build a virtual environment to simulate all the camera views. In the designed experiment, several robots move randomly on the  $5m \times 3m$  floor [22], [23]. Each robot is attached with an individual visual fiducial marker AprilTag [24], which facilitates the vision-based six degrees of freedom robot localization. The task of the camera network is to obtain the positions of all the robots as accurately as possible. Specifically, the robots' position  $p_j$  are randomly given and cameras have fixed mounting position  $o_i$  and flexible optical-axis direction  $\phi_i$  in the PTZ camera mode, while flexible position  $o_i$  (minimum height is limited to 1m) but fixed downward optical-axis direction  $\phi_i$  in the drone mode. All the cameras' coefficients are calibrated from real cameras. Our virtual environment generates images corresponding to the given camera configuration, shown in Fig. 7. AprilTag detection algorithm is used to estimate the position of robots according to the images. The optimization problem (11) is solved by `fmincon` in MATLAB 2021, which runs on an AMD Ryzen 5 3600 CPU, 3.60GHz processor and 32GB

TABLE I: Simulation Results on Multi-Robot Detection

| $N_c$ | $N_t$ | PTZ      |                  |          | Drone    |                  |          |
|-------|-------|----------|------------------|----------|----------|------------------|----------|
|       |       | Method   | Computing Time/s | Error/mm | Method   | Computing Time/s | Error/mm |
| 3     | 3     | IPOPT    | 0.684            | 11.97    | IPOPT    | 0.534            | 5.13     |
|       |       | SQP      | 0.223            | 11.85    | SQP      | 0.120            | 5.67     |
|       |       | Arslan's | -                | 13.56    | Arslan's | -                | 9.71     |
| 5     | 10    | IPOPT    | 1.065            | 21.96    | IPOPT    | 1.321            | 14.62    |
|       |       | SQP      | 0.490            | 25.69    | SQP      | 0.586            | 17.51    |
|       |       | Arslan's | -                | 47.11    | Arslan's | -                | 22.18    |
| 7     | 20    | IPOPT    | 1.492            | 23.55    | IPOPT    | 1.915            | 17.44    |
|       |       | SQP      | 0.720            | 27.23    | SQP      | 0.746            | 21.99    |
|       |       | Arslan's | -                | 51.02    | Arslan's | -                | 22.93    |

\* Arslan's method [16] did not give comparisons on computing time.

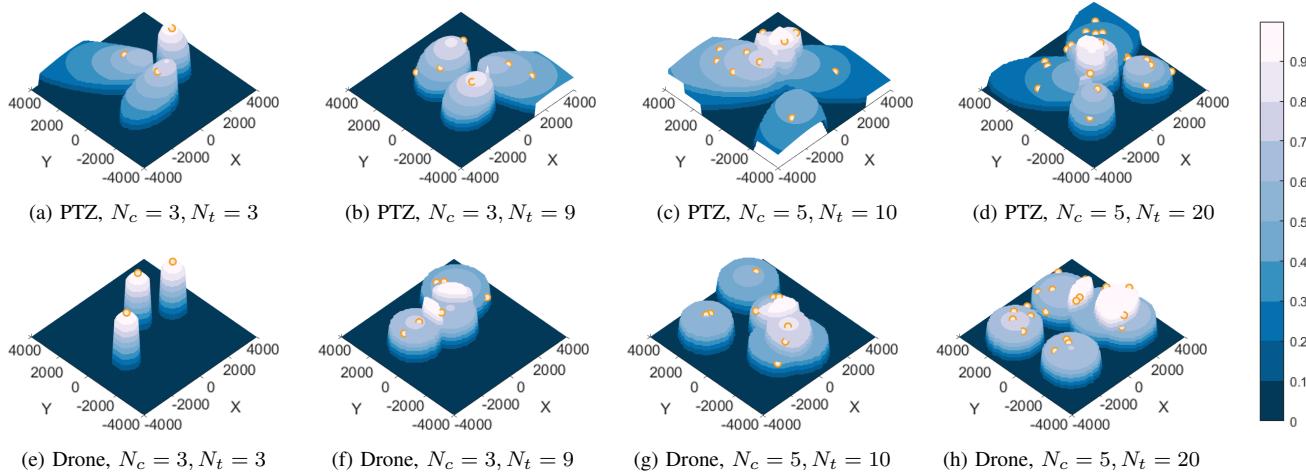


Fig. 8: Simulations for different network scales. The XY axis represents the experiment plane. The Z axis represents the sensing quality, which has been normalized for better presentation. The targets' positions are marked with circles.

RAM. We generate 12 different simulation scenarios and run 10 experiments in each scenario. The results are shown in Fig. 8 and Table I.

First, the average errors of the PTZ camera systems are larger than that of the drone camera systems. This is because the fixed positions of the PTZ cameras lead to more serious distortion of the sight when monitoring robots at a large relative angle, while drones can fly directly above the robots to get a sight with better sensing quality. Besides, we compare the performance of the two algorithms, i.e., Interior Point OPTimizer (IPOPT) and Sequential Quadratic Programming (SQP). The results show that SQP computes faster with larger error. Moreover, the sensing quality model in [16] is used for comparison. It is obvious that our model achieves more accurate localization since our model incorporates the effects of distortion. Lastly, as the numbers of cameras and targets increase, the average computing time increases too due to the increasing complexity of the optimization problem, and it becomes harder to find a global optimal solution.

### B. Experiment

To experimentally validate the proposed model, we set up a 3m (width)  $\times$  5m (length)  $\times$  2.5m (height) rectangular box-shaped environment with three self-build interchangeable-lens PTZ cameras mounted on the ceiling, the intrinsic and

extrinsic coefficients of which are pre-calibrated. We set several Omni-directional robots attached with AprilTags moving on the ground as sensing targets. The cameras' positions are fixed, and the optical-axis directions perform as optimization variables, limited by PTZ cameras' physical properties. Since the estimation of the depth depends on the size of the tags in the image, this experiment can examine both the positioning and measurement ability simultaneously. The ground truth of the robots' positions is given by manual measurement with laser rangefinder. Our algorithm will give an optimal camera configuration for sensing quality, as shown in Fig. 9 and video (<https://youtu.be/5UiEs1RrC-0>).

We first use manual configuration of the cameras to cover all the robots and get the average error. Then, the configuration are set according to Algorithm 1. The average  $\bar{Q}$  is 1.165 and average error decreased by 37.0%, down to 35.3. Finally, we use sensing model proposed by [16] as another comparison. The results are shown in Table II.

TABLE II: Experimental Results on Multi-Robot Detection

| Method         | Average $\bar{Q}$ | Average Error/mm |
|----------------|-------------------|------------------|
| Manual Setting | 1.610             | 56.0             |
| Our Method     | <b>1.165</b>      | <b>35.3</b>      |
| Arslan's [16]  | 1.482             | 37.2             |

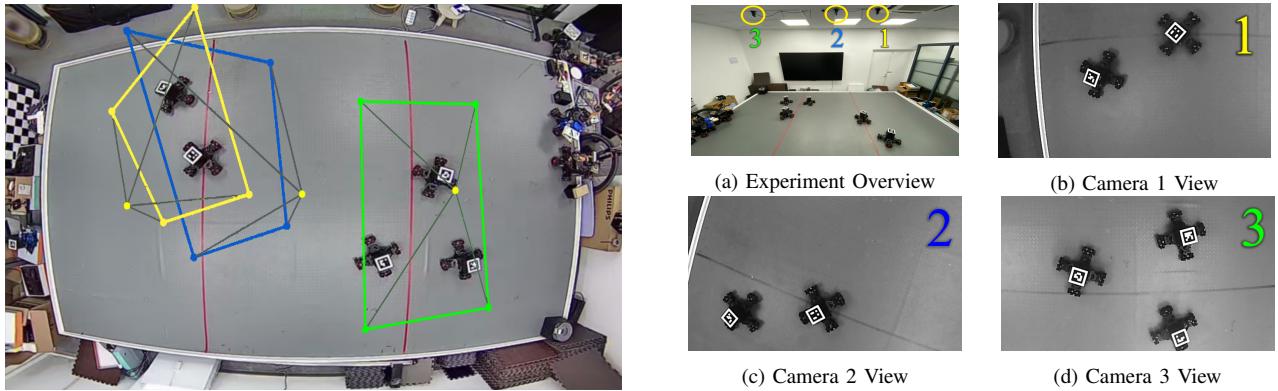


Fig. 9: Experimental demonstration of camera network configuration for optimal sensing quality. (left) Top view of the environment and visualization of an optimal configuration for certain situation. (a) Three PTZ cameras mounted on the ceiling. (b)(c)(d) The camera view with the optimized configuration.

#### IV. CONCLUSION

In this paper, we considered the balance between the coverage for all targets and sensing quality of each single one for a camera network. The problem has been formulated as an optimization problem which takes into account the perspective, distortion factor and multi-camera fusion results. Optimal configuration for camera network can be achieved without modeling the environment, suitable for any pre-calibrated cameras and targets anywhere in 3D space. The proposed method demonstrates superior performance as compared to previous approaches, increasing the average accuracy in AprilTag detection task by 38.9%.

We would like to introduce some potential research directions: i) developing optimization approach for online tracking and large-scale problems, ii) optimizing sensing quality while taking the surveillance of unknown areas into account, and iii) developing a refined model for objects which need to be tracked from a certain direction.

#### REFERENCES

- [1] C. Chen, R. Surette, and M. Shah, "Automated monitoring for security camera networks: promise from computer vision labs," *Security Journal*, vol. 34, no. 3, pp. 389–409, 2021.
- [2] C. Bernhard, R. Reinhard, M. Kleer, and H. Hecht, "A case for raising the camera: a driving simulator test of camera-monitor systems," *Human Factors*, p. 00187208211010941, 2021.
- [3] Y. Wu, F. Tang, and H. Li, "Image-based camera localization: an overview," *Visual Computing for Industry, Biomedicine, and Art*, vol. 1, no. 1, pp. 1–13, 2018.
- [4] Y. Yang, D. Tang, D. Wang, W. Song, J. Wang, and M. Fu, "Multi-camera visual slam for off-road navigation," *Robotics and Autonomous Systems*, vol. 128, p. 103505, 2020.
- [5] S. Yang, S. A. Scherer, X. Yi, and A. Zell, "Multi-camera visual slam for autonomous navigation of micro aerial vehicles," *Robotics and Autonomous Systems*, vol. 93, pp. 116–134, 2017.
- [6] X. Zhang, X. Chen, F. Farzadpour, and Y. Fang, "A visual distance approach for multicamera deployment with coverage optimization," *IEEE/ASME Transactions on Mechatronics*, vol. 23, no. 3, pp. 1007–1018, 2018.
- [7] M. S. Suresh, A. Narayanan, and V. Menon, "Maximizing camera coverage in multicamera surveillance networks," *IEEE Sensors Journal*, vol. 20, no. 17, pp. 10170–10178, 2020.
- [8] E. Hörster and R. Lienhart, "On the optimal placement of multiple visual sensors," in *Proceedings of the 4th ACM International Workshop on Video Surveillance and Sensor Networks*, pp. 111–120, 2006.
- [9] A. Mavrinac and X. Chen, "Modeling Coverage in Camera Networks: A Survey," *International Journal of Computer Vision*, vol. 101, pp. 205–226, Jan. 2013.
- [10] C. Piciarelli, L. Esterle, A. Khan, B. Rinner, and G. L. Foresti, "Dynamic reconfiguration in camera networks: A short survey," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 5, pp. 965–977, 2016.
- [11] J. O'rourke, *Art gallery theorems and algorithms*, vol. 57. Oxford New York, NY, USA, 1987.
- [12] A. Efrat, L. J. Guibas, S. Har-Peled, D. C. Lin, J. S. Mitchell, and T. Murali, "Sweeping simple polygons with a chain of guards," in *Proceedings of the Eleventh Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 927–936, 2000.
- [13] K. R. Konda and N. Conci, "Optimal configuration of PTZ camera networks based on visual quality assessment and coverage maximization," in *2013 Seventh International Conference on Distributed Smart Cameras (ICDSC)*, pp. 1–8, IEEE, Oct. 2013.
- [14] K. R. Konda, N. Conci, and F. De Natale, "Global Coverage Maximization in PTZ-Camera Networks Based on Visual Quality Assessment," *IEEE Sensors Journal*, vol. 16, pp. 6317–6332, Aug. 2016.
- [15] S. Aghajanzadeh, R. Naidu, S.-H. Chen, C. Tung, A. Goel, Y.-H. Lu, and G. K. Thiruvathukal, "Camera Placement Meeting Restrictions of Computer Vision," in *2020 IEEE International Conference on Image Processing (ICIP)*, pp. 3254–3258, IEEE, Oct. 2020.
- [16] O. Arslan, H. Min, and D. E. Koditschek, "Voronoi-based coverage control of pan/tilt/zoom camera networks," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5062–5069, IEEE.
- [17] T. Kuroda, *Essential principles of image sensors*. CRC press, 2017.
- [18] I. Robinson and A. Trautman, "Optical geometry," in *New Theories in Physics*, 1988.
- [19] N. Bisagno and N. Conci, "Virtual camera modeling for multi-view simulation of surveillance scenes," in *2018 26th European Signal Processing Conference (EUSIPCO)*, pp. 2170–2174, IEEE, Sept. 2018.
- [20] J. G. Fryer and D. C. Brown, "Lens distortion for close-range photogrammetry," *Photogrammetric Engineering and Remote Sensing*, vol. 52, pp. 51–58, 1986.
- [21] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 1, pp. 666–673, IEEE, 1999.
- [22] X. Ding, H. Wang, H. Li, H. Jiang, and J. He, "Robopheus: A virtual-physical interactive mobile robotic testbed," *arXiv preprint arXiv:2103.04391*, 2021.
- [23] H. Wang, X. Ding, J. He, K. Margellos, and A. Papachristodoulou, "Safety-aware optimal control in motion planning," *arXiv preprint arXiv:2204.13380*, 2022.
- [24] E. Olson, "AprilTag: A robust and flexible visual fiducial system," in *2011 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3400–3407, IEEE, May 2011.