Dynamic Camera Network Reconfiguration for Crowd Surveillance

Niccoló Bisagno
Department of Information Engineering and Computer Science (DISI)
University of Trento
niccolo.bisagno@unitn.it

Nicola Conci
Department of Information Engineering and Computer Science (DISI)
University of Trento
nicola.conci@unitn.it

Bernhard Rinner
Institute of Networked and Embedded Systems (NES)
Alpen-Adria-Universität Klagenfurt
bernhard.rinner@aau.at

ABSTRACT

Crowd surveillance will play a fundamental role in the coming generation of video surveillance systems, in particular for improving public safety and security. However, traditional camera networks are mostly not able to closely survey the entire monitoring area due to limitations in coverage, resolution and analytics performance. A smart camera network, on the other hand, offers the ability to reconfigure the sensing infrastructure by incorporating active devices such as pan-tilt-zoom (PTZ) cameras and UAV-based cameras, which enable the adaptation of coverage and target resolution over time. This paper proposes a novel decentralized approach for dynamic network reconfiguration, where cameras locally control their PTZ parameters and position, to optimally cover the entire scene. For crowded scenes, cameras must deal with a trade-off among global coverage and target resolution to effectively perform crowd analysis. We evaluate our approach in a simulated environment surveyed with fixed, PTZ, and UAV-based cameras.

KEYWORDS

Smart camera network; crowd surveillance; camera control; UAV; PTZ

ACM Reference format:
Niccoló Bisagno, Nicola Conci, and Bernhard Rinner. 2018. Dynamic Camera Network Reconfiguration for Crowd Surveillance. In *Proceedings of International Conference on Distributed Smart Cameras, Eindhoven, Netherlands, September 3–4, 2018 (ICDSC '18),* 6 pages.
DOI: 10.1145/3243394.3243704

1 INTRODUCTION

Surveillance of crowded scenes is a key issue for public safety in indoor and outdoor environments. Various factors influence the development of a critical situation of crowds, hence a camera network must be able to capture local events as well as guarantee a global coverage of the whole area. Covering the entire monitoring area while maintaining a sufficient resolution of the (moving) individuals might be challenging with fixed cameras. A costly camera infrastructure is necessary to provide a sufficient target resolution in every part of the monitoring area to perform common tasks such as person identification. Consequently, video footage of potentially empty parts would also be captured with such static camera network.

An alternative approach is to deploy reconfigurable cameras, which can dynamically adapt their field of view (FoV), resolution and position. In this case, the goal is to optimize coverage and target resolution depending on the current state of the crowded scene. Such camera networks aim to focus the attention on critical areas of the crowd, but ensuring an acceptable level of attention also on less critical areas. In this paper, we propose a novel network control approach to explore the trade-off between target resolution and coverage in heterogeneous networks consisting of fixed, PTZ, and UAV-based cameras. In our approach, we model the crowd scene and the camera network in a simulation environment, we estimate the state of the crowd by merging the contributions of the individual cameras' FOVs and we let cameras locally decide on their next PTZ or position parameters.

Our contribution can be summarized as (1) a policy to trade-off between global coverage and crowd coverage, (2) a new metric to evaluate the performances of the surveillance task, (3) a framework to track the crowd flow based on the coverage maps, and (4) a 3D simulator of crowd behaviors based on [4] and heterogeneous camera networks.[1]

The remainder of this paper is organized as follows: Section 2 briefly discusses related work. Section 3 describes the key components of the proposed approach along with the evaluation metric. Section 4 presents the results of our simulation study, and Section 5 provides some concluding remarks together with a discussion about potential future work.

2 RELATED WORK

Automated video surveillance systems have been studied with the goal of reducing the human intervention while operating a control room [3, 11, 16]. In such frameworks, cameras need to be aware of the network configuration sharing the necessary information to improve events capturing and global coverage of the scene [9, 10, 13]. Due to the dynamic nature of the events and the corresponding need for reconfiguring the camera network layout, research in the field has to deal with a limited amount of annotated data. This also makes each event unique and difficult to reproduce.

Relying on virtual environments and simulation tools can help to partially address these issues. Virtualization has been widely

ICDSC '18, Eindhoven, Netherlands
© 2018 ACM. 978-1-4503-6511-6/18/09…$15.00
DOI: 10.1145/3243394.3243704

[1] Simulator available at https://github.com/nick1392/HeterogenousCameraNetwork

adopted in research, both in the community of camera networks [12, 17] and crowd analysis [6].

Pan, tilt and zoom (PTZ) cameras have been deployed to survey crowded scenes [1, 12, 17]. PTZ cameras can be reconfigured to increase coverage of certain areas, either by progressively scanning the environment, or zooming in to specific locations in presence of events of interest. In a cooperative camera network, PTZ cameras can be effectively used to track targets of interest [1, 2, 7, 14].

Unmanned Aerial Vehicles (UAVs), or drones, have been adopted for different services and purposes, both in civil and military applications including environmental pollution monitoring, agriculture monitoring, and management of natural disaster rescue operations [8, 15, 18].

Yao et al. [19] identify the key features of a distributed network for crowd surveillance, i.e., to (1) locate and re-identify a person across the network, (2) track persons, (3) recognize and detect local and global crowd behavior, (4) cluster and recognize actions, and (5) detect abnormal behaviors. To achieve these goals, issues like how to fuse information coming from multiple cameras performing crowd behavior analysis tasks, how to learn crowd behavior patterns, and how to cover an area with particular focus on key events, are among a variety of challenges to be tackled.

3 DYNAMIC CAMERA NETWORK RECONFIGURATION

Our approach is based on a set of fixed, PTZ, and UAV-based cameras with different characteristics and capabilities for the surveillance of crowded scenes. Multiple cameras provide diversity by observing and sensing an area of interest from different points of view, which further increases the reliability of the sensed data. Our framework for camera network reconfiguration is suitable for both static and dynamic scenarios.

In this section we introduce the key components of our proposal. In particular, we first introduce the observation model for the environment, which describes the relation between the observation and its confidence. We then describe, how each type of camera is modeled in the simulation environment, and formalize the reconfiguration objective. Next, we describe our reconfiguration policy that allows the network focus to be tuned in order to achieve a suitable trade-off between global coverage and crowd resolution.

3.1 Observation Model

The region of interest C, which has to be surveyed is divided in a uniform grid of $I \times J$ cells where the indexes $i \in \{1, 2, \ldots, I-1\}$ and $j \in \{1, 2, \ldots, J-1\}$ of each cell $c_{i,j} \in C$ represent the position of the cell in the grid. We assume a scenario evolving at discrete time steps $t = 0, 1, 2, \cdots, t_{end}$. At each time step, the network is able to gather the observation over the scene to be monitored, process it, and share it with the other camera nodes in order to plan the next set of actions to be taken. For this purpose we define

- an observations vector $O_{i,j}$, which represents the number of pedestrians detected for each cell $c_{i,j} \in C$;
- a spatial confidence vector $S_{i,j}$, which describes the confidence of the measures for each cell $c_{i,j} \in C$. The value only depends on the relative geometric position between the observing camera and the observed cell;
- a time confidence vector $L^t_{i,j}$, which depends on the time passed since the cell has last been observed;
- an overall confidence vector $F^t_{i,j}$, which depends on the temporal and spatial confidences.

The observations vector is defined as

$$O_{i,j} = \{o_{1,1}, o_{1,2}, \cdots, o_{i,j}, \cdots, o_{I,J}\} \tag{1}$$

The value $o_{i,j}$ for each cell $c_{i,j}$ is given as

$$o_{i,j} = \begin{cases} \frac{ped}{ped_{max}} & \text{if } ped \leq ped_{max} \\ 1 & \text{if } ped > ped_{max} \end{cases} \tag{2}$$

where ped is the number of pedestrians detected within the cell by a given camera, and ped_{max} is the maximum number of pedestrian for a cell to be considered as crowded. Crowded cells should be monitored with a higher resolution.

Occlusion of targets is one of the main challenges in crowded scenarios. We assume that our camera network is able to robustly detect a pedestrian when its head is captured with a resolution of at least 24 × 24 pixels, in line with the smaller bound for common face detection algorithms [5].

For each cell a spatial confidence vector is defined as

$$S_{i,j} = \{s_{1,1}, s_{1,2}, \cdots, s_{i,j}, \cdots, s_{I,J}\} \tag{3}$$

where the value $0 < s_{i,j} \leq 1$ is bounded, and decreases as the distance between the observing camera and the cell of interest $c_{i,j}$ increases. The actual value of a cell depends on the type of observing camera and is described in Section 3.2.

Similarly, a time confidence vector is defined as

$$L_{i,j} = \{l^t_{1,1}, l^t_{1,2}, \cdots, l^t_{i,j}, \cdots, l^t_{I,J}\}. \tag{4}$$

Each value $l^t_{i,j}$ is defined as

$$l^t_{i,j} = \begin{cases} 1 - \frac{t - t^0_{i,j}}{T_{MAX}} & \text{if } t - t^0_{i,j} \leq TMAX \\ 0 & \text{if } t - t^0_{i,j} > T_{MAX} \end{cases} \tag{5}$$

where $t^0_{i,j}$ is the most recent time instant, in which cell $c_{i,j}$ was observed, and T_{MAX} represents the time instant, after which the confidence drops to zero. The value $l^t_{i,j}$ decays over time if no new observation $o_{i,j}$ on cell $c_{i,j}$ become available.

Given the spatial and temporal confidence metrics, the overall confidence vector is defined as

$$F^t = \{f^t_{1,1}, f^t_{1,2}, \cdots, f^t_{i,j}, \cdots, f^t_{I,J}\} \tag{6}$$

with

$$f^t_{i,j} = s_{i,j} * l^t_{i,j}. \tag{7}$$

Thus, for each cell $c_{i,j}$ we have an observation $o_{i,j}$ with an overall confidence $f^t_{i,j}$. The confidence value varies between 0 and 1, where 1 represents the highest possible confidence. If a sufficient number of cameras is available for covering all cells concurrently, the overall confidence vector is given as $F^I = \{1, \cdots, 1\}$.

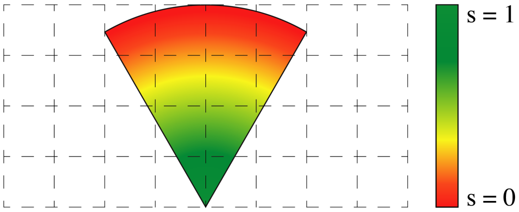

Figure 1: A fixed camera observes the environment without varying the spatial confidence for each cell at each time step.

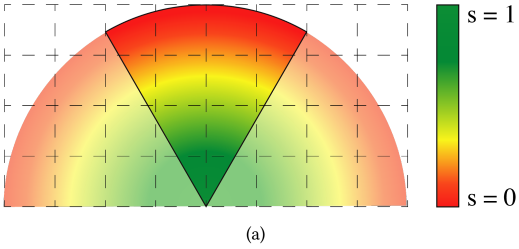

Figure 2: At each time step, a PTZ camera can pan its FOV in the range of 180° given a fixed initial position.

3.2 Camera Models

We briefly describe the models adopted for the three different camera types: fixed cameras, PTZ cameras, and UAV-based cameras. We assume that all fixed and PTZ cameras are mounted at a fixed height, such that their spatial confidence metric depends only on the distance from the cell. All UAV-based cameras fly at a fixed altitude.

3.2.1 Fixed Cameras. Fixed cameras (see Fig. 1) provide a confidence matrix, which gradually decreases as the distance from the camera increases. Being (x, y) a point in the space at a distance d from a fixed camera, the value of the spatial confidence $s(x, y)$ is defined as

$$s(x,y) = \begin{cases} -\frac{1}{d_{max}} * d + 1 & \text{if } d < d_{max} \\ 0 & \text{if } d \geq d_{max} \end{cases} \tag{8}$$

with d_{max} being the distance from the camera, over which the spatial confidence is zero. Thus, the confidence value $s_{i,j}$ of cell $c_{i,j}$ is defined as

$$s_{i,j} = \max\{s(x,y)\}_{\forall (x,y) \in c_{i,j}}. \tag{9}$$

3.2.2 PTZ Cameras. PTZ cameras are modeled similarly to fixed cameras, with the additional capability to dynamically change the field of view (see Fig. 2).

3.2.3 UAV-based Cameras. For UAV-based cameras the FOV projection on the ground plane is different with respect to the previous models, as shown in Fig. 3. The spatial confidence of point (x, y) at a distance d from the UAV is computed as

$$s(x,y) = \begin{cases} -\frac{1}{d_{uav}} * d + 1 & \text{if } d < d_{uav} \\ 0 & \text{if } d \geq d_{uav}. \end{cases} \tag{10}$$

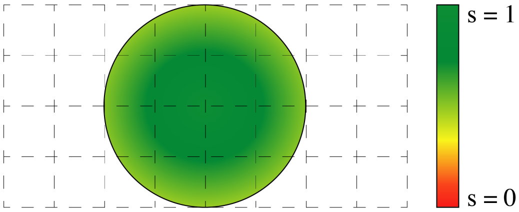

Figure 3: Example of the distribution of the spatial confidence in the area surveyed by an UAV.

3.3 Reconfiguration Objective

The objective of the heterogeneous camera network is to guarantee the coverage of the scene while focusing on more densely populated areas. The priority metric defines the importance of each cell to be observed. A high value indicates that the cell is crowded or that we have a low confidence on its current state, thus requiring an action.

In order to formalize the reconfiguration objective, a priority vector P is defined as

$$P^t = \{p^t_{1,1}, p^t_{1,2}, \cdots, p^t_{i,j}, \cdots, p^t_{I,J}\}. \tag{11}$$

The priority for each cell is defined as

$$p^t_{i,j} = \alpha * o^t_{i,j} + (1-\alpha)(1 - f^I_{i,j}) \tag{12}$$

where $0 \leq \alpha \leq 1$ represents a weighting factor to tune the configuration and $f^I_{i,j}$ represents the pre-defined ideal confidence for the cell.

The objective G of each camera, given its possible set of action, is to minimize the distance between the confidence vector and the priority vector

$$G = \min\{||F^{t+1} - P^t||\} \tag{13}$$

$$\begin{cases} \min\{F^{t+1} - F^I\} & \text{if } \alpha = 0 \\ \min\{F^{t+1} - O^t\} & \text{if } \alpha = 1 \end{cases} \tag{14}$$

Setting $\alpha = 1$ causes the network to focus on observing more densely populated areas with no incentive to explore unknown cells. In contrast, $\alpha = 0$ causes the network to focus on global coverage only without distinguishing on the crowd density of the cells.

3.4 Update Function

At each time step t, the network has knowledge about the current observation vector O^t, the spatial confidence vector S^t, the time confidence vector L^t, and the overall confidence vector F^t. In order to progress to the next time step $t + 1$, an update function for these vectors is required.

The temporary spatial confidence vector S^{t+1}_{temp} is determined by the geometry of cameras at time $t + 1$. For each cell, the value $s^{t+1}_{temp_{i,j}}$ is the maximum spatial confidence value of all cameras observing the cell (i, j). Cells that are not covered by any camera will have a spatial confidence value of 0.

We estimate the time confidence vector as follows. L^{t+1}_{time} is computed by applying Eq. 5 to each element of L^t. Another temporary time confidence vector L^{t+1}_{new} is computed setting to 1 the value of all cells currently observed, and setting to 0 all other cells.

With the estimated vectors we compute two estimations of the overall confidence vector such that:

$$F^{t+1}_{time} = S^t * L^{t+1}_{time} \tag{15}$$

$$F^{t+1}_{new} = S^{t+1}_{temp} * L^{t+1}_{new} \tag{16}$$

The new overall confidence vector is then computed as

$$F^{t+1} = \max\{F^{t+1}_{new}, F^{t+1}_{time}\}_{\forall (i,j)}. \tag{17}$$

For each cell (i, j) in which $f^{t+1}_{new} > f^{t+1}_{time}$, we also need to update the last time the cell has been observed $t^0(i, j) = t + 1$, and the observation vector $o^t(i, j)$.

3.5 Local Camera Decision

In our approach all the information vectors described in Section 3.1 are shared and known to all cameras. Each camera locally decides its next position using a greedy approach to minimize the cost defined in Eq. 13 in its neighborhood.

At each time step, each mobile, PTZ and UAV-mounted camera selects a neighborhood that can be explored. The UAV's neighborhood is defined as a square centered at the cell where the drone is currently placed (see Fig. 3). The PTZ neighborhood is a rectangle which covers the space in front of the camera as shown in Fig. 2.

For each cell in the neighborhood, we center a window W of size $N_w \times N_w$ on each cell $c_W \in W$ and we store in the cell the value

$$c_W = \sum ||f^{t+1}_{i,j} - p^t_{i,j}||. \tag{18}$$

The UAV will then move toward the cell in its neighborhood with the largest c_W, and the PTZ steers its FOV to be centered on that cell. If two or more cells have the same value of c_W, the camera selects one of them randomly.

3.6 Evaluation Metrics

We define the Global Coverage Metric (GCM) for evaluating the network coverage capability as

$$GCM(t) = \frac{\sum\limits_{\forall c_{i,j} | f^t_{i,j} > g} 1}{I * J} \tag{19}$$

with g being the threshold above which we consider the cell covered. We then average the results for the whole duration of the observation as follows:

$$GCM_{avg} = \sum_{t=0,\cdots,t_{end}} \frac{GCM(t)}{t+1} \tag{20}$$

We define the People Coverage Metric (PCM) for evaluating the network capability to cover pedestrian in the scene as

$$PCM_{tot} = \frac{\sum\limits_{\forall person \in c_{i,j} | f^t_{i,j} > p} 1}{totalPeople} \tag{21}$$

with p being the threshold above which we consider the cell covered.

ID	g and p	α	GCM	PCM
1	0.2	0	12.4 %	17.4 %
2	0.2	0.5	14.3 %	20.5 %
3	0.2	1	10.4 %	13.5 %
4	0.01	0	42.9 %	47.6 %
5	0.01	0.5	30.3 %	33.1 %
6	0.01	1	22.9 %	28.2 %
7	0.01	0	43.1 %	45.6 %
8	0.01	0.5	28.7 %	54.4 %
9	0.01	1	26.1 %	61.2 %

Table 1: Simulation experiments. Legend: ID–experiment; g,p–cell coverage thresholds; GCM–global coverage metric; PCM–people coverage metric. Experiments 1-6 refer to a uniformly distributed crowd, experiments 7-9 refer to a crowd with directional motion properties.

4 EXPERIMENTAL RESULTS

For the experiments we define an environment of size 60×60 meters. The scene is square-shaped exhibiting people passing by, cars, and vegetation. Pedestrians can enter and exit the scene from any point around the square. Each cell $c_{i,j}$ is a square of 1 × 1 meter. In this environment 2 fixed cameras, 2 UAVs and 2 PTZs are positioned as shown in Fig. 4(a). Sample images of the environment from a PTZ and a UAV-based camera are shown in Figures 4(b) and 4(c), respectively. For our experiments we simulate the movement of 400 pedestrians crossing the scene with the following parameters :

- T_{max} = 3 seconds
- ped_{max} = 2
- d_{max} = 10 meters
- fixed and PTZ cameras height = 5 meters
- UAV cameras height = 7 meters

4.1 Quantitative Results

In this section we present the quantitative results obtained with our model in the simulated environment. The goal is to evaluate the capabilities of the system to survey a crowded scene using the metrics defined in Sec. 3.6. We run 9 different simulation experiments with varying values of g, p, and α.

The values for g and p indicate how reliable the information is about position in space and pedestrians, respectively. A threshold of 0.2 indicates that our observation is at most 2.4 seconds old, when taken with a spatial confidence equal to 1. A threshold of 0.01 represents the cells and pedestrian about which we have a minimum level of information.

As a reference if all 6 cameras remain fixed, they are able to cover 6 % of the entire area with $g = 0.2$ and 12 % with $g = 0.01$. In experiments (3) and (6), α is set to 1, causing our camera network to focus only on observing pedestrians with no incentive to explore new areas in the environment. In experiments (1) and (4), α is set to 0 resulting in maximizing the coverage regardless of the position of pedestrians. In experiments (2) and (5), α is set to 0.5 aiming for balancing coverage and pedestrian tracking in crowded areas. We can observe that in experiments (1) and (4) we obtain the lowest values of GCM, which is expected since we are focusing

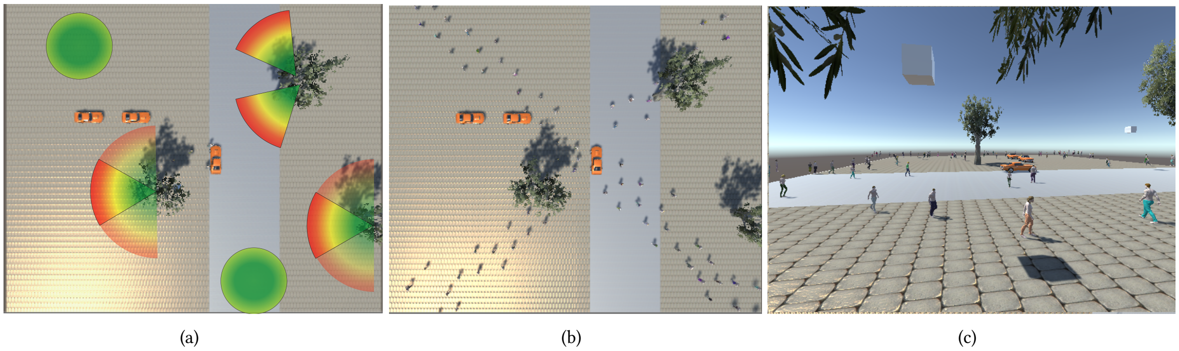

Figure 4: (a) Top view of the simulation environment including the camera positions. (b) Sample image from a UAV-based camera. (c) Sample image from a PTZ camera.

on pedestrians. We also achieve the lowest scores in term of PCM because cameras have no incentive in exploring new areas.

Experiments (7), (8), and (9) are conducted using a directional crowd (Fig. 4(b)). When the network focuses only on observation in (9), it obtains the best results in term of PCM and the worst one in terms of global coverage GCM. As expected, we obtain the best results in terms of coverage of the environment (GCM) in experiments (3) and (6). Since the crowd is uniformly distributed in the space, we also obtain the best results in terms of PCM. In experiments (2) and (5), the network combines global coverage and crowd monitoring, the system under performs compared with the scenes where $\alpha = 0$ and $\alpha = 1$.

4.2 Qualitative Results

In this section we present the qualitative results obtained with our model in the simulated environment. The goal is to demonstrate, how our system is able to follow the crowd.

For this purposes, we simulate a single group of five pedestrians crossing the scene from the bottom left to the top right as shown in the sequence depicted in Fig. 5. The UAV is able to closely follow the pedestrians in the environment, scoring a $PCM = 70.4\,\%$ and $GCM = 3.2\,\%$, as shown in Fig. 6. Fig. 7 shows how observation, priority and confidences maps are updated over time in order to guide the UAV in the tracking scenario.

5 CONCLUSION

In this paper we have presented a novel camera reconfiguration approach for crowd monitoring. Our approach allows heterogeneous camera networks to focus on high target resolution or on wide coverage. Although based on simplified assumptions for camera modeling and control, our approach is able to trade-off coverage and resolution of the network in a resource-effective way. In future research, network coordination will be improved relying on cooperative decision-making between cameras and assigning different polices (e.g., values of α) to different camera types.

REFERENCES

[1] Pietro Azzari, Luigi Di Stefano, and Alessandro Bevilacqua. 2005. An effective real-time mosaicing algorithm apt to detect motion through background subtraction using a PTZ camera. In *Proc. IEEE Conference on Advanced Video and Signal Based Surveillance.* 511–516.

[2] Alessandro Bevilacqua and Pietro Azzari. 2006. High-quality real time motion detection using ptz cameras. In *Proc. IEEE International Conference on Video and Signal Based Surveillance.* 23–23.

[3] Gian Luca Foresti, Petri Mähönen, and Carlo S Regazzoni. 2012. *Multimedia video-based surveillance systems: Requirements, Issues and Solutions.* Vol. 573. Springer Science & Business Media.

[4] Dirk Helbing and Peter Molnar. 1995. Social force model for pedestrian dynamics. *Physical review E* 51, 5 (1995), 4282.

[5] Michael Jones and Paul Viola. 2003. Fast multi-view face detection. *Mitsubishi Electric Research Lab TR-20003-96* 3, 14 (2003), 2.

[6] Julio Cezar Silveira Jacques Junior, Soraia Raupp Musse, and Claudio Rosito Jung. 2010. Crowd analysis using computer vision techniques. *Signal Processing Magazine* 27, 5 (2010), 66–77.

[7] Sangkyu Kang, Joon-Ki Paik, Andreas Koschan, Besma R Abidi, and Mongi A Abidi. 2003. Real-time video tracking using PTZ cameras. In *Proc. International Conference on Quality Control by Artificial Vision*, Vol. 5132. International Society for Optics and Photonics, 103–112.

[8] Asif Khan, Bernhard Rinner, and Andrea Cavallaro. 2018. Cooperative Robots to Observe Moving Targets: A Review. *IEEE Transactions on Cybernetics* 48, 1 (2018), 187–198.

[9] Krishna Reddy Konda and Nicola Conci. 2013. Optimal configuration of PTZ camera networks based on visual quality assessment and coverage maximization. In *Proc. International Conference on Distributed Smart Cameras.* IEEE, 1–8.

[10] Peter Lewis, Lukas Esterle, Arjun Chandra, Bernhard Rinner, and Xin Yao. 2013. Learning to be different: Heterogeneity and efficiency in distributed smart camera networks. In *Proc. IEEE 7th International Conference on Self-Adaptive and Self-Organizing Systems.* 209–218.

[11] Christian Micheloni, Bernhard Rinner, and Gian Luca Foresti. 2010. Video Analysis in PTZ Camera Networks - From master-slave to cooperative smart cameras. *IEEE Signal Processing Magazine* 27, 5 (2010), 78–90.

[12] Faisal Z Qureshi and Demetri Terzopoulos. 2007. Surveillance in virtual reality: System design and multi-camera control. In *Proc. Conference on Computer Vision and Pattern Recognition.* IEEE, 1–8.

[13] Martin Reisslein, Bernhard Rinner, and Amit Roy-Chowdhury. 2014. Smart Camera Networks. *IEEE Computer* 47, 5 (2014), 23–25.

[14] Bernhard Rinner, Lukas Esterle, Jennifer Simonjan, Georg Nebehay, Roman Pflugfelder, Gustavo Fernandez Dominguez, and Peter R Lewis. 2014. Self-aware and self-expressive camera networks. *IEEE Computer* 48, 7 (2014), 21–28.

[15] Allison Ryan, Marco Zennaro, Adam Howell, Raja Sengupta, and J Karl Hedrick. 2004. An overview of emerging results in cooperative UAV control. In *Proc. 43rd IEEE Conference on Decision and Control*, Vol. 1. 602–607.

[16] Mubarak Shah, Omar Javed, and Khurram Shafique. 2007. Automated visual surveillance in realistic scenarios. *IEEE MultiMedia* 14, 1 (2007).

[17] Geoffrey R Taylor, Andrew J Chosak, and Paul C Brewer. 2007. Ovvv: Using virtual worlds to design and evaluate surveillance systems. In *Proc. Conference on Computer Vision and Pattern Recognition.* IEEE, 1–8.

[18] Evşen Yanmaz, Saeed Yahyanejad, Bernhard Rinner, Hermann Hellwagner, and Christian Bettstetter. 2018. Drone networks: Communications, coordination, and sensing. *Ad Hoc Networks* 68 (2018), 1–15.

[19] Hongxun Yao, Andrea Cavallaro, Thierry Bouwmans, and Zhengyou Zhang. 2017. Guest Editorial Introduction to the Special Issue on Group and Crowd Behavior Analysis for Intelligent Multicamera Video Surveillance. *IEEE Transactions on Circuits and Systems for Video Technology* 27, 3 (2017), 405–408.

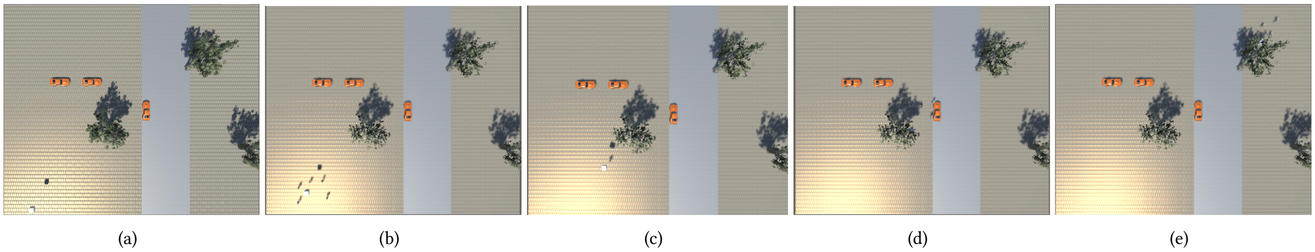

Figure 5: Image sequence of a group of pedestrian moving from the bottom left of the environment (a) to the top right (c). The image is captured by a top view camera during the simulation to demonstrate the tracking behavior of our network.

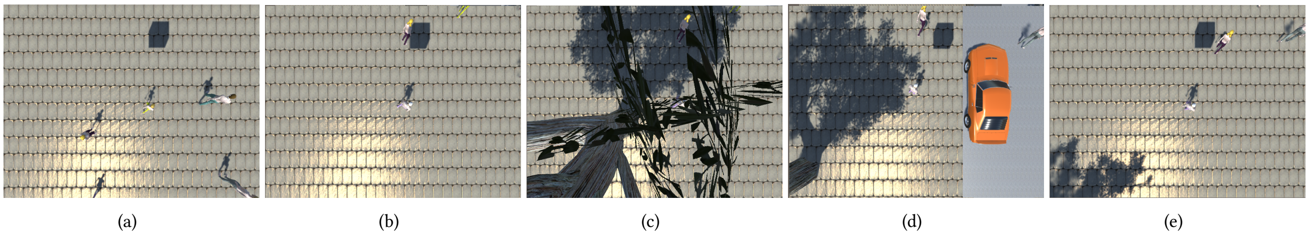

Figure 6: Image sequence of a group of pedestrian moving from the bottom left of the environment (a) to the top right (e) captured by a UAV surveying the scene.

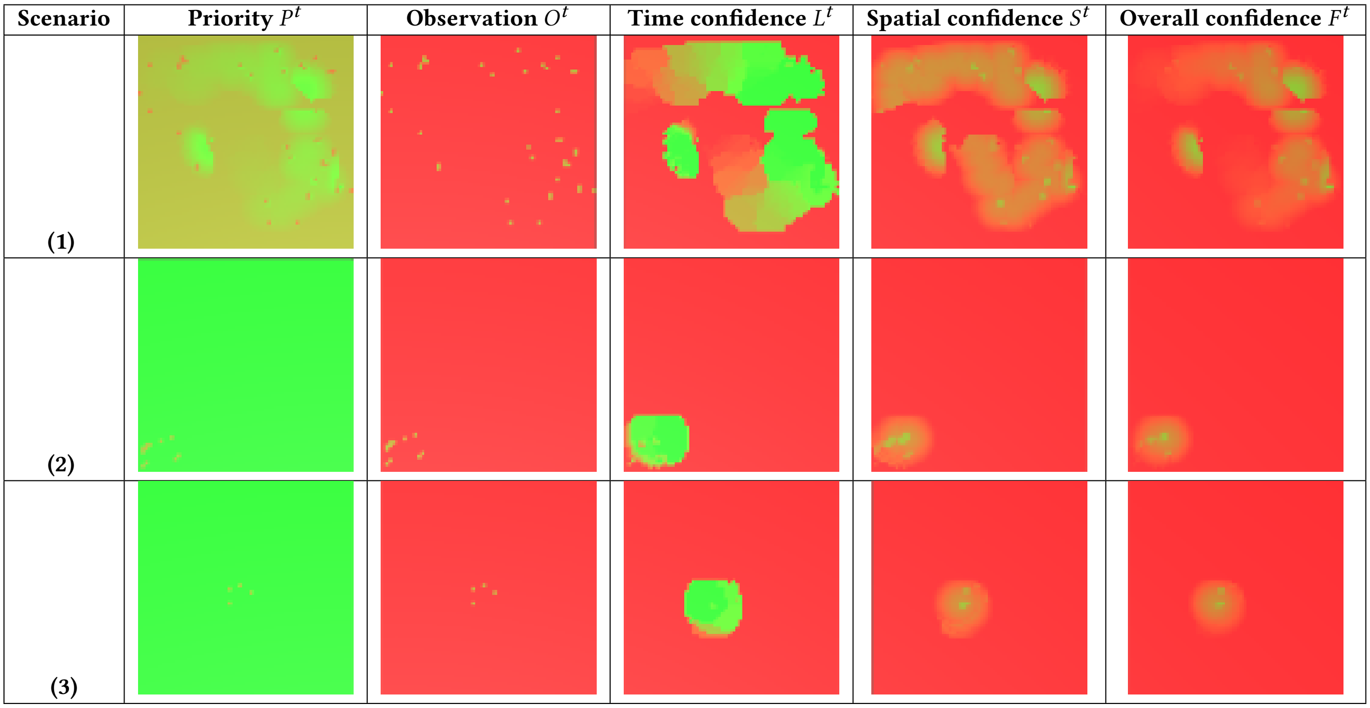

Figure 7: Graphical representation of priority P^t, observation O^t, time confidence L^t, spatial confidence S^t and overall confidence F^t for 3 different scenarios: (1) Camera Network Sample, (2) Tracking sample at time $t = 0$, (3) Tracking sample at time $t = 10$. In (2) and (3) the UAV focuses on the observation matrix, such that the next priority map depends only on previous observations. Red represent the value 0, and green represents value 1.