The Future of Camera Networks: Staying Smart in a Chaotic World

Lukas Esterle  
Aston Lab for Intelligent Collectives Engineering (ALICE)  
Aston University, Birmingham, UK  
l.esterle@aston.ac.uk

Richie McBride  
Edesix Ltd., Edinburgh, UK  
r.mcbride@edesix.com

Peter R. Lewis  
Aston Lab for Intelligent Collectives Engineering (ALICE)  
Aston University, Birmingham, UK  
p.lewis@aston.ac.uk

Xin Yao  
University of Birmingham, UK  
x.yao@cs.bham.ac.uk

ABSTRACT
Camera networks become smart when they can interpret video data on board, in order to carry out tasks as a collective, such as target tracking and (re-)identification of objects of interest. Unlike today’s deployments, which are mainly restricted to lab settings and highly controlled high-value applications, future smart camera networks will be messy and unpredictable. They will operate on a vast scale, drawing on mobile resources connected in networks structured in complex and changing ways. They will comprise heterogeneous and decentralised aggregations of visual sensors, which will come together in temporary alliances, in unforeseen and rapidly unfolding scenarios. The potential to include and harness citizen-contributed mobile streaming, body-worn video, and robot-mounted cameras, alongside more traditional fixed or PTZ cameras, and supported by other non-visual sensors, leads to a number of difficult and important challenges. In this position paper, we discuss a variety of potential uses for such complex smart camera networks, and some of the challenges that arise when staying smart in the presence of such complexity. We present a general discussion on the challenges of heterogeneity, coordination, self-reconfigurability, mobility, and collaboration in camera networks.

KEYWORDS
future camera networks, surveillance, mobile smart cameras, body worn cameras, humans in the loop, distributed control

1 INTRODUCTION
More powerful embedded processors and improvements in memory in the last fifteen years have led to a revolution in camera technology: camera networks evolved towards smart camera networks. These smart cameras can observe the physical world, process the acquired images on board, and communicate aggregated information and extracted knowledge rather than the video feed alone. This enables them to detach from central components, analysing imagery locally, making decisions and acting on them autonomously. This in turn allows for faster reactions to emerging situations, as well as increased flexibility, scalability, and robustness in surveillance networks [41]. However, this also introduces significant challenges in terms of target tracking and re-identification.

In this paper we will analyse the state-of-the-art and especially the state-of-practice of camera networks, and discuss what is possible and what is not yet possible from a smart camera network perspective, and with a particular focus on continuous target tracking and required re-identification. We discuss the challenges arising from mobility and uncertainty in highly heterogeneous surveillance networks, and the problems that come along with them. We dive into detail on coordination problems and how we can overcome them in centralised and decentralised settings. Finally, we describe a potential future outlook for smart camera networks, that support people through carrying out highly responsive tasks autonomously.

The rest of this paper is structured as follows. The next section discusses the current state of practice in surveillance networks as operated by industry. Section 3 sheds light on the issues with mobile cameras and outlines potential solutions. Section 4 discusses potential benefits and obstacles in using heterogeneous cameras in a network. Section 5 explores the requirement of coordination in large camera networks arising due to rapidly unfolding situations. Section 6 discusses the need of future camera networks to autonomously form teams and cooperate in order to achieve their goals. Section 7 provides an outlook and concludes the paper.

2 CURRENT PRACTICE
The development of smart cameras enables them to operate autonomously. However, when networked, they can cooperate in order to carry out more complex tasks and achieve collective goals [13, 14, 27, 36, 46]. As with other types of cyber-physical system, more recently, self-awareness [24, 29] has been introduced in smart cameras, enabling them to reason about their own behaviour, and adapt accordingly in changing environments [15, 42].
These developments of camera networks also changed their original purpose. While manual surveillance is still an important aspect of camera networks, other application areas have also emerged. Those that have received most attention by researchers include:

- **Autonomous distributed target tracking** coordinates tracking of objects with a single camera at a time. This ensures efficiency of network-wide resources, but requires either perfect re-identification or knowledge of the environment to ensure sufficient re-identification of targets among different cameras [10, 13, 31, 34, 48].

- **Autonomous multi-camera tracking**, by contrast, tries to cover each target with multiple cameras at the same time. Several issues may arise, such as: how to observe the object from different angles, while keeping a minimum of overlap; or how to place cameras to ensure continuous coverage of objects by multiple cameras [4, 7, 21]. In addition, cameras need to ensure they track the same object. This requires efficient data association across multiple cameras [6, 19].

- **(k-)Coverage and barrier optimization**, the *art-gallery problem*, aims to ensure coverage of a specific area. Barrier optimization refines this to ensure objects cannot pass between areas undetected. $k$ refers to the number of cameras observing the same area at any given time [8, 11, 14].

- **Search and rescue/follow** often mounts cameras on mobile robots and aims to find specific targets autonomously. These targets can be stationary or mobile as well as collaborative or evasive [12, 18, 22, 35, 40, 47]. Cameras have to be able to re-identify a target based on a given model. In case of mobile cameras (e.g. PTZ or drones), cameras need to be controlled in such a fashion that they do not deplete their resources before the target has been found.

- **Personal data collection and performance review** is a novel area of application of camera networks for personal data collection and performance review of individuals. The idea is, to have video material of interactions within the personal range of employees. This information can be used to improve performance and also when incidents occur.

- **Guidance and control** utilises captured video material to map the environment and hence control and guide robotic systems and autonomously driving cars safely through it [49, 51]. However, we can also use camera networks to guide people through their environment, allowing them to avoid hazardous situations or simply find the best and fastest way through an unknown environment.

When considering camera networks, one typically thinks first of static or PTZ (Pan-Tilt-Zoom) cameras, where fields of view (FOV) can only to a limited extend be changed. Alternatively, networks of drones equipped with visual sensors introduce mobility and allow the network to observe situations from different viewpoints if necessary. Mobile cameras, by contrast, can relocate within their environment, being worn by humans or mounted on mobile robots.

Indeed, the efforts highlighted above have so far been either focused on static and pan-tilt-zoom cameras, or limited to fully controllable robots acting as mobile cameras. The sector of body worn video, which has rapidly developed in terms of both technology and growth of the market, has been widely neglected by researchers so far. Similarly, hybrid networks, where different types of sensors are brought together, are often used in search and rescue operations, where stationary sensors support mobile robots [20, 25, 50]. Hybrid camera networks have also received little attention to date [9, 52].

In particular, there is little academic consideration of the impact of important recent developments in mobile camera technology. These include new market sectors focussed on body worn video (BWV) and incident based video (IBV), which employ small video devices often worn on lanyards or special harnesses, e.g., as shown in Figure 2. These cameras can be switched on by the wearer when necessary, and can stream, record and/or process video feeds on board, or in conjunction with a local mobile device. Figure 1 shows these two developing market sectors, and their overlap with traditional CCTV and more recently developed smart CCTV applications.

Figure 1: Market sectors (light shaded circles) and their overlap, in camera systems. Camera networks of different types interplay with each other. Each market sector also considers smart devices (darker shaded areas) where information can be processed on the device and communicated to others.

Figure 2: Edesix’s VideoBadge. Source: http://edesix.com/

While CCTV is usually used to observe a specific area, BWV and IBV are employed by individual users. Users are categorised into two groups either where mobile cameras are carried as part of their job (BWV) or for their own safety and when video should be recorded for legal purposes (IBV). BWV is worn by personnel that often and repeatedly come into situations where video evidence has to be taken (e.g., a police officer, or a bailiff), IBV, on the other hand, is only turned on in rare situations (e.g., a shopkeeper at the airport is faced with an incident). In both cases, however, the worn cameras are used to capture evidence of a situation. Current surveillance networks, deployed in the field, combine standard equipment, mounted at fixed locations, with novel, mobile BWV/IBV (cp. Figure 1).

There are several benefits of introducing BWV/IBV into smart camera networks, but also a number of challenges. On one hand, additional mobile cameras are added to the fixed set of cameras mounted on walls and ceilings. This increases the amount of information captured from the environment and allows for more...
When introducing mobility to camera networks, the question arises of how to control the mobility behaviour. This question becomes even more pressing when cameras are worn by humans (as opposed to robots) who may not comply with instructions. In such cases, we can only talk about suggested movement rather than controlled movement. Here, we discuss the benefits and challenges associated with both fully controlled and suggested movement.

The main benefit of fully-controlled movement is being able to select the position and orientation of each camera. This is usually available in robotic systems, but may be constrained in some cases, such as with fixed-wing UAVs. This allows a more stable video feed, even when in motion, in comparison to BWV. However, robotic systems are often not able to move as fast as humans and, depending on the robot, may have trouble with obstacles in the environment.

A more realistic scenario is a hybrid network consisting of cameras mounted on humans, where human decision-making behaviour forms part of the network’s decentralised control process. Having humans in the loop allows for more reactive camera networks, as humans are expert at acting on instinct and making good, rapid decisions. The wearer of a camera might make a decision before a controller could decide what each camera might need to do next. This can be a benefit as well as a disadvantage. On the one hand, a person with BWV might follow a target more thoroughly and hence provide continuous and reliable target tracking. On the other hand, the person might decide not to follow a person of interest and hence not provide any video feed at all. A smart controller able to account for the needs and uncertainties of the human, however, might suggest a BWV user to move in a direction that removes the person from potential harm. Alternatively, a suggestion to a user might involve moving only if safe to do so, along with alternative safer options, that still provide value to the network.

4 HETEROGENEITY

Heterogeneity is an increasingly prevalent property of complex computational and cyber-physical systems [3, 26, 28], and smart camera networks are no exception. Heterogeneity in camera networks may take various forms (e.g., as depicted in Figure 4), including variation in hardware between nodes, physical properties of the camera, processing power and storage, and connectivity and bandwidth. A camera network with platform heterogeneity is composed of different types of cameras, possibly including static and fixed mounted ones, PTZ cameras able to change their orientation and zoom, BWV able to relocate, and robot-mounted cameras. This can often lead to other sources of platform heterogeneity, for example, fixed mounted cameras are most likely to have a wired network connection, and can usually rely on a constant power supply. In contrast, mobile cameras have to rely on batteries as a power supply and typically only have wireless communication interfaces. Another manifestation of heterogeneity arising from platform heterogeneity is in terms of image resolution. Mobile cameras are fully embedded devices and rely on very small sensors. Due to their lower processing capabilities and available storage, acquired images may have a lower resolution to begin with. Considering multi-camera continuous target tracking and re-identification, cameras in the network with lower processing capabilities may slow down the network. Furthermore, lower image quality may skew the acquired model of a target and accelerate the problem of concept drift in target tracking and online target model learning [2].
Cameras’ behaviours can also vary, leading to behavioural heterogeneity. The classic approach is for cameras to use a common algorithm or behavioural strategy. However, cameras are often located in different areas, and hence are subject to different experiences. This challenges the common assumption when deploying smart camera networks, that a single behaviour, captured in a single software solution, should be used. Indeed, results from Lewis et al.’s study of heterogeneity in smart camera networks [27] support the idea that to achieve efficient network-wide behaviour, each device should have a specialised behaviour, based on its own local perspective and context. This result mirrors those from studies of heterogeneity elsewhere in multi-agent and self-organising systems, where specific benefits include adaptation to unknown situations or environments. For example, Anders et al. [1] showed that heterogeneity among nodes in a network can lead to better achievement of system wide objectives, and Salazar et al. [44] highlight the importance of understanding and harnessing heterogeneity when nodes can adapt independently online, in response to uncertainty and changes in the environment. In dynamic task allocation, Campbell et al. [5] show that variation in agent behaviour increases the system’s ability to adapt to varying stimuli. Focussing on sensor networks, Yarvis et al. [53] explore the impact of heterogeneity at the network level, and Römer et al. [43] observe that, while wireless sensor networks were initially conceived as homogeneous networks of near-identical sensors, in practice many applications contain platform heterogeneity. Further, Prasath et al. [37] recommend heterogeneity in wireless sensor networks to generate near-optimal configurations. This raises a further challenge: while heterogeneity can improve global efficiency, how should “good” behaviour for each individual camera be determined? In large networks, for even reasonable sets of possible behaviours at the individual camera level, this leads to a combinatorial explosion of possible network behaviour configurations, and identifying by hand the most appropriate configuration at a particular point in time is not feasible. Thus, some automation, through a-priori optimisation, or by online machine learning (e.g., as in [27]) will be needed.

Figure 4 illustrates some challenges: 5 cameras, relying purely on their visual information, is tasked to observe a single target. Cameras $C_1$ and $C_2$ observe it from the front. Since their view angles are similar, identification by both cameras is not a problem. However, little additional information is gathered. The object is occluded from $C_3$, so in order for the camera to contribute knowledge, it would need to move. $C_4$ can see the object from behind. Assuming the model holds sufficient information, $C_4$ also identifies the target, and adds more new knowledge than $C_3$. $C_5$ should be able to re-identify the object based on $C_1$ or $C_2$’s information, while also providing new information about the object. If cameras are calibrated, their location and orientation information can be used to aid re-identification. Otherwise, the network relies on the generalisation of the learnt model to other points of view.

5.1 Centralised Coordination

A global controller knows where each camera is located, its relation to other cameras in the network, and the content of their FOVs. Hence, the coordinator knows what targets are covered by how many cameras, and which targets are not covered at any given time. Achieving goals can be considered an optimization problem in a game-theoretic setting [31]. This, however, requires the cameras to constantly communicate their information to the central coordinator. With fixed cameras on wired connections, this might not be a big issue. Furthermore, the central controller being aware of the location and orientation of each individual camera and their FOVs makes re-identification of targets a much easier task. However, in mobile BWV, information exchange has to be performed via wireless communication channels. In rapidly unfolding situations, required infrastructure might not be available for cameras to receive commands from the central entity. In addition, having constantly changing environments, a camera might not be able to wait for control input from the coordinator.

5.2 Decentralised Coordination

In decentralised coordination, coordination of the network is split up among multiple clusters of cameras. For each cluster, a local coordinator is selected, perhaps through voting. Additionally it is required that the communication range is at least the range at which a camera can detect and potentially re-identify a target with sufficient confidence. In comparison to centralised coordination, in decentralised coordination, the communication effort is reduced among neighbours. This does not necessarily reduce network-wide communication, however, clusters currently not engaged in tracking a target might not have any communication at all.

One of the main challenges is how to select clusters efficiently at runtime due to (i) the directed sensing of cameras, (ii) the possibility to change their pose with PTZ cameras, and (iii) potentially even change their location when employing mobile cameras. Having a highly dynamic network of mobile cameras, selection of such a coordinator with a high frequency at runtime introduces an additional problem. Finally, when targets move across multiple clusters, where each one is handled by a different coordinator, this cluster hand-off needs to be handled efficiently. However, on a local level, a cluster coordinator still has global knowledge, and would be used to ensure positive re-identification of targets among different cameras.
worked, and involve no interaction between cameras. Conversely, An important challenge, especially in heterogeneous networks, is different communication constraints, using shared situation awareness to maximise the performance of each team member. Raubal and Winter [39] define a framework for agents to negotiate how to perform a task. They consider, at runtime, if a task requires collaboration and analyse the trade-off between urgency, risk, and distance involved. In camera networks, Qureshi and Terzopoulos [38] use contract net to assign tasks to cameras at runtime. They use auctions to form coalitions of cameras to achieve specific tasks. Similarly, Esterle et al. [10, 13] use auctions to find neighbouring cameras for continuous tracking. They introduce artificial pheromones to adapt to changes in the network. Li and Zhang [30] use a utility based on the number of cameras observing a target and on the overlapping area of the target to determine which cameras should collaborate to achieve the best results. This is done exhaustively for all cameras in a centralised way. SanMiguel and Cavallaro [45] propose a coalition-based collaborative tracking framework. In negotiation, cameras join teams to collaboratively track targets, based on their available resources and expected contribution.

However, these approaches do not account for the dynamics of the network introduced by mobility, nor the unpredictable behaviour of citizens contributing video data. This additional degree of complexity will require further investigation to improve continuous multi-camera target tracking and re-identification.

7 OUTLOOK

One recent trend in industry is the proliferation of mobile cameras, including body worn video. However, research often still focusses on homogeneous camera networks. In this paper we shed some light on the upcoming challenges that will arise by incorporating BWV into surveillance networks. Future camera networks will be required to respond to any given situation autonomously or with minimal human interaction. When required, these networks can be extended by BWV on local personnel and/or privately owned equipment. This sudden heterogeneity in the network can bring benefits, but needs to be handled accordingly. Since this heterogeneity cannot be foreseen, this needs to be done autonomously and during runtime. In addition, BWV requires new techniques to compensate for motion introduced by rapid movements. The additional cameras, however, allow for immediate extension of the network and hence supplementary video data can be made available when required. Nevertheless, the security team may not be able to completely rely on BWV added by non-security personnel as they may not comply with requests, e.g. relocating to a position if it is potentially harmful. BWV also offers benefits to lay personnel, as security services can use their video footage to guide them safely out of harm’s way. In a given emergency, police special forces may deploy additional mobile camera nodes in the form of more BWV or drones. Recent developments allow those drones to land on and depart from walls, enabling them to conserve energy. These drones would also operate autonomously, extending the camera infrastructure, and communicating with the rest of the network when required. Special forces wearing BWV could again be guided towards the reported incident.

Heterogeneity and self-organisation enable the network to attend to multiple tasks at the same time, through parallelisation and specialisation according to local situations. Autonomous coordination of mobile smart camera networks will allow the assessment

5.3 Self-organised Coordination

Existing large camera deployments, especially BWV, are not networked, and involve no interaction between cameras. Conversely, networked deployments are smaller and centrally controlled, introducing a bottleneck and a single point of failure. To address the issue of scalability, self-organising and distributed coordination can be used. In self-organising cameras, the network also benefits from higher flexibility and more robustness to changes in the network.

However, self-organisation in a camera network requires a high degree of interaction among the individual cameras, which may result in higher communication and be error prone. Additionally, not having a central controller results in cameras acting based on local rules and defined behaviour. This might lead to cameras not behaving as expected resulting in targets not being tracked or to many cameras covering the same targets at the same time.

Introducing mobility, whether in the form of robots or BWV, requires the network to account for changing goals, resources, and rapidly changing environments. This requires the control of the system to be highly adaptive, robust, and flexible. In order to achieve this, individual and collective online learning at the camera and network level is required. Furthermore, the system and individual cameras need to evaluate their own performance as well as the performance of others at runtime. This will allow them to self-adapt to such demanding change. Further, current approaches do not consider the relationship and interplay of different goals and available resources. To increase the efficiency of camera systems, trade-offs need to be modelled during runtime and multiple objectives in our learning approaches need to be considered. Guarantees for emergent behaviour in self-organising camera systems are required to ensure given goals in the network are achieved. However, this would require the network to have a single common goal. Furthermore, techniques on how to handle malicious/non-cooperative/counter-productive nodes within a network are required.

6 COLLABORATION AND TEAM WORK

An important challenge, especially in heterogeneous networks, is cooperation between individual cameras to achieve common goals. Foeken and Kester [16] analyse adaptive team formation under different communication constraints, using shared situation awareness to maximise the performance of each team member. Raubal and Winter [39] define a framework for agents to negotiate how to perform a task. They consider, at runtime, if a task requires collaboration and analyse the trade-off between urgency, risk, and distance involved. In camera networks, Qureshi and Terzopoulos [38] use contract net to assign tasks to cameras at runtime. They use auctions to form coalitions of cameras to achieve specific tasks. Similarly, Esterle et al. [10, 13] use auctions to find neighbouring cameras for continuous tracking. They introduce artificial pheromones to adapt to changes in the network. Li and Zhang [30] use a utility based on the number of cameras observing a target and on the overlapping area of the target to determine which cameras should collaborate to achieve the best results. This is done exhaustively for all cameras in a centralised way. SanMiguel and Cavallaro [45] propose a coalition-based collaborative tracking framework. In negotiation, cameras join teams to collaboratively track targets, based on their available resources and expected contribution.

However, these approaches do not account for the dynamics of the network introduced by mobility, nor the unpredictable behaviour of citizens contributing video data. This additional degree of complexity will require further investigation to improve continuous multi-camera target tracking and re-identification.
of and response to situations faster, even in a chaotic and fast-changing environment, while generating additional valuable and up-to-date information in real time.

ACKNOWLEDGMENT

This work is supported by the SOLOMON project (Self-Organisation and Learning Online in Mobile Observation Networks) funded by the European Union H2020 Programme, grant n°705026.

REFERENCES


Lukas Esterle, Peter R. Lewis, Richie McBride, and Xin Yao