Towards Smart Infrastructures for Modern Surveillance Networks

René Golembewski and Michael Rossberg and Guenter Schaefer
Telematics and Computer Networks Research Group, Ilmenau University of Technology, Germany

Abstract

Nowadays, large-scale surveillance systems can be found in wide distributed security infrastructures or, more regional, in big cities with highly frequented public transportation networks. Already the latter ones consist of thousands of cameras with continuously enhancing quality, producing enormous amounts of multidimensional and high-resolution multimedia data. While new hardware technologies lead to the evolution of so called *Smart Cameras* with embedded processing and communication capabilities, distributed algorithms become more and more important concerning efficiency and robustness against attacks or malfunctions. Ongoing improvements in algorithms for robust object recognition and growing distributed processing power, are expected to enable surveillance systems to work more and more autonomously, similar to wireless sensor networks. Thus, one important step in this direction is to support distributed applications and algorithms by developing intelligent communication infrastructures and help spawning an efficient and robust network of smart devices. Within this article the problem of distributed topology estimation is discussed in detail and a first solution approach is given. Furthermore, we provide an overview of our simulation framework aimed to support computer vision research by evaluating distributed algorithms in large, heterogeneous networks.

1 Introduction

Security in urban areas is one important field of research and development due to the expectation that in the future more than half of the 7 billion humans are expected to live in cities. One major goal is to keep people, buildings, and infrastructure free from vandalism and crime. But also the planning and optimization of urban traffic or public transportation needs intelligent solutions to support the administrative personnel.

To achieve these goals future surveillance infrastructures will primarily make use of distributed smart cameras (DSCs) that are characterized by the combination of integrated systems, computer vision (CV), and sensor networking technology [1,2], and communicating via arbitrary, heterogeneous transport networks.

A distributed setup in a metropolitan area will hence consist of hundreds or thousands of DSCs, spanning all over the city and analyzing the events. The network connectivity for each system can be provided by any kind of Internet uplink.

But using public networks for communication implicates strong security requirements and an efficient use of resources in order to keep operational costs at affordable levels. Figure 1 depicts an exemplary network of seven high-resolution cameras, attached to computing devices for further processing and distributed communication without any centralized instance.

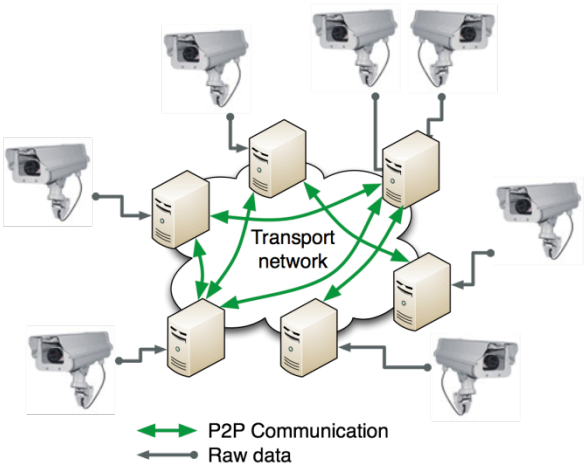

Figure 1 Distributed Smart Camera Network

While this outlook goes far beyond current operating principles, where only small amounts of cameras are used for live monitoring and most of the material is solely analyzed reactively in case of an incident, there is no alternative to online analysis, if preventive steps shall be taken. Current urban surveillance infrastructures rely on centralized components for these analytical steps, leading to potential bottlenecks and generating single points of failure. Exposed targets for sab-

otage attacks are a direct consequence of the centralization, and also privacy should be considered, if we think about a solitary huge database with potentially intimate information.

Thus, the forecasted change towards a distributed communication paradigm requires the development of new algorithms in many different fields of application. CV approaches should be validated in distributed systems, and networking performance has to be made measurable to prove functionality. Since refining and evaluating those algorithms is not trivial in the absence of suitable test-setups, benchmarking frameworks can help to deal with these problems.

The rest of this article describes scenarios with derived objectives and states the need for proper benchmarking solutions. We are presenting first results and give a case study for the problem of topology estimation in distributed systems. Finally, a conclusion is made and further research topics are outlined.

2 Scenarios and Objectives

In the following section, after a short definition of our understanding of smart camera networks, exemplary scenarios are presented and analyzed in order to identify the requirements. Afterwards, most recent issues for the development of smart communication infrastructures are extracted.

2.1 Smart Camera Networks

Smart Camera Networks (SCNs) form loose alliances of DSC nodes, connected by any kind of heterogeneous networking infrastructure, e.g., the Internet. The main task of a camera device is not longer the delivering of video material to a central server farm or a dedicated processing cluster, as with growing advance in signal processing and embedded devices [3] DSCs can transmit meta-information and aggregated results of their local analysis, only.

To define a more abstract model, we attribute all the features of standard PCs to the interconnected smart devices, because of the continuously increasing processing power of embedded devices (cf. Fig. 1). Thus, a DSC is able to make use of long-term memory, general-purpose processors, special image processing hardware, and an external power supply. Especially, the latter delimits DSCs from the concept of wireless sensor networks, in which the nodes are usually battery-powered and may thus not perform powerful calculations. From an operator's point of view, the whole network shall provide collaboratively processed information, and it should also work fully autonomously even in the case of partial failures (graceful degradation).

2.2 Topology Estimation in Smart Camera Networks

In order to support location-aware applications and enable efficient communication algorithms, an efficient virtual communication topology is required, e.g. cameras observing similar events should form a cluster in order to align own observations with a macroscopic situation.

An important issue in this context is the self-calibration of the networks, e.g., to automatically deduct the logical topology, estimate camera positions, and detect structural problems. A definite arbitration, which cameras interact, may solely happen based on common events, as the physical topology and also the geographical circumstances are not necessarily decisive to determine the logical neighborhood. Thus, one main objective is using exclusively visual information to fulfill this task. Furthermore, it cannot be expected that different cameras have shared fields of view (FOVs) to support calibration or object tracking, because it is not economical in many large-scale scenarios if multiple DSCs cover approximately the same area. Therefore, we suggest to perform such a mapping by using object recognition, e.g., a person or car that appears in one camera and was shortly seen before in another one indicates that both cameras are linked.

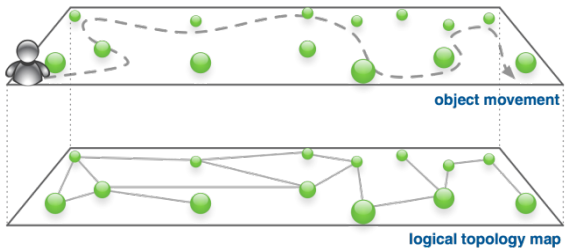

Figure 2 Topology Estimation depending on the movement of objects

Other difficulties arise with steadily changing environmental conditions and mobility. Approaches trying a calibration with special well-known training objects would not be applicable in that case, due to the constantly demand for recalibration.

At the present time, most recent topology estimation solutions known to the authors [4] do not fully comply with the already mentioned objectives:

- High number of distributed cameras
- Large area of application
- No overlapping fields of view

- Highly dynamic adaptivity
- Difficult environmental conditions

We are working on a solution for the given estimation problem and try to identify the main difficulties when relying on noisy detection results and unreliable transport networks. Decoupled from individual CV algorithms, we want to give an approximation of the required communication effort to achieve self-calibration, and identify ways to optimize, support, and extend networks of DSCs.

2.3 Intelligent Video Surveillance of Safety-Critical Areas

Due to the mostly limited security personnel capacities, it is not possible to monitor all available imaging devices at a time. Furthermore, humans are often not able to fully recognize large and complex scenes, spanning over many different camera views, because of the flood of information. Therefore, intelligent surveillance systems should analyse the whole scenario in real-time and inform security service personnel only in case of an alert. This can be the case if the system identifies suspicious luggage, or strange behaviour of people.

However, there are also situations where manual intervention is needed, like the search for specific persons or the pursuit of potential criminals. Intelligent surveillance systems should be able to give an advice of the most probable position and prior movement patterns.

Contributions of ongoing research in computer vision technologies show steady improvements on many topics like face recognition [5,6], age estimation [7], gait recognition [8], pose classification [9], and motion detection [10]. The resulting challenge is now to communicate the outcome of those CV approaches to the network in an efficient and yet secure manner, enabling algorithms to generate hypotheses of observed situations.

2.4 Urban Traffic Planning

One more interesting chance of camera surveillance is to support the establishment of *Smart Cities* [11] and *Smart Traffic Systems*. By adding more functionality to distributed traffic observing cameras, it can be possible to deduct more efficient rules for traffic management or navigation software. Taking into account the detected (and anonymized) official registration numbers of vehicles, an estimation of complex flows can be made to extend the traditional plain traffic counting facilities.

But the more decisive part of transforming existing systems to a distributed solution is to achieve robustness, high availability, and privacy, independently from a centralized server infrastructure. Again, in this scenario we can find high dynamics and fast changing circumstances, potential fields of application for self-calibrating algorithms.

3 Solution approaches

In this section we want to state the first solution approaches of our research. Because of the large variety of targeted distributed systems, we introduce an evaluation framework, built upon a discrete-event simulator with replaceable transport networks, which is used to develop novel algorithms and validate results.

3.1 Evaluation Environment

Figure 3 gives a high level view of our systematic strategy to cope with many kinds of different problem domains: The core simulation of our environment is coupled to the extensible modules for Communication, Traffic Generation, Computer Vision, and Visualization. Because of the strict separation of communication and platform specific programming code, all implemented features can easily be migrated on native Linux/Unix systems.

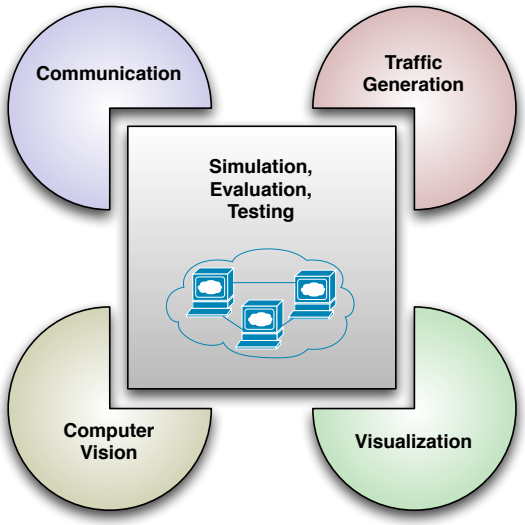

Figure 3 System Overview

The principal challenge in developing a smart infrastructure for modern surveillance networks is to start with a reasonable and well-founded model of possible scenarios and conditions. While, the size of DSCs is well known to reach hundreds or thousands of cameras, there physical and geographical topology cannot be assumed to be randomly structured. Thus, within our framework a *Traffic Generation* part is in charge

of providing realistic camera topologies and of generating human traffic that is presented to the *Computer Vision* part subsequently as input and enabling users of the framework to evaluate any desired CV algorithm by just implementing it as an extension of the simulated DSC node. Hereby the output of the CV is not limited to a named person. Every kind of detectable thing can be transformed into a complex feature vector with properties like speed, structure, size, direction, or in the case of humans, e.g., face, gait, posture, and age. This vector is an extensible data structure, e.g., an XML file.

The actual communication mechanism between the DSCs is detached from the simulated transport network where arbitrary topologies and protocols can be used, like UDP or TCP, optional over wired links or radio. All acts of communication are organized in a layered fashion, and the DSC nodes are forming a logical overlay network, depending on the implemented behaviour. Due to the fact that DSCs are handling sensitive data, also security and privacy issues have to be considered and evaluated. The simulation therefore will support the implementation of attacker models as well as according countermeasures.

3.2 Distributed Topology Estimation Algorithm: A Case Study

To underpin the strengths of the presented approach a first implemented and evaluated use case is presented. The previously concerned distributed topology estimation algorithm that uses only the movement patterns of visually identified objects, and without overlapping FOVs shall reveal the potential of our modular benchmarking platform.

The used algorithm can be seen as a form of distributed consensus algorithm with spatio-temporal-analysis of distributed events that are communicated between DSC nodes and used for finding correspondences afterwards. While trivial broadcasting of events is working in simple scenarios, it has scalability issues, and will only be considered for comparison here.

Instead, within the proposed system nodes will build local clusters, based on an occasional message exchange with other nodes. If multiple events are detected at several nodes within a configurable time period, they assume to have a logical relationship resulting as an edge in the vision graph. Subsequent events will only be communicated within clusters and the emerging vision graph describes the logical topology of the camera network.

To evaluate the algorithm a scenario with an area of $1000m \times 1000m$ and 100 DSCs was simulated, whereby the density of nodes in the center is higher than in the outer zone in order to model an urban downtown area. Figure 4 shows the directed input topology graph with edges for simulated traffic flows. Edges and also nodes have an additional capacity parameter, modelling the maximum number of objects on lanes and places. This parameter allows the simulation of congestion and also to evaluate ways for detecting it.

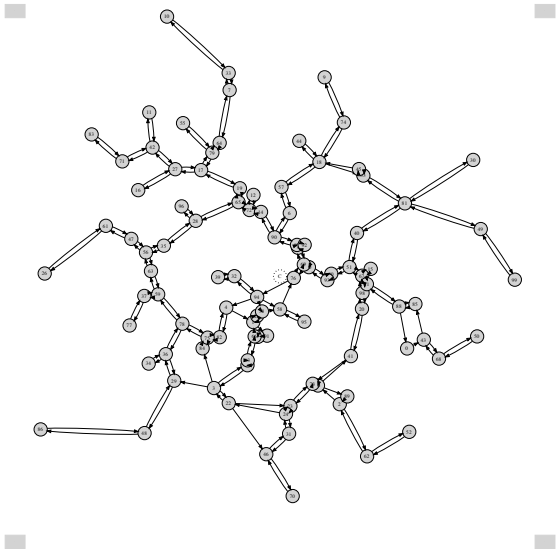

Figure 4 Input Topology Graph with Traffic Flows

Traffic data is generated simultaneously by 200 persistent objects, which start from arbitrary entry points and move on the previously defined input topology. Each object's speed is modelled by a normal distribution $N\ (\mu = 1.5, \sigma = 0.5)\ m/s$. After reaching the destination node, a new visual entity is generated from a set of $2 \cdot 10^9$ feature vectors and a new destination is chosen.

The detection of objects is simulated via an out-of-band communication between traffic generator and DSCs, to prevent an influence on the traffic measurements in the transport network. Further parameters for detection rate and false-positive rate describing the quality of object recognition. For our simulation we have chosen a detection rate $d = 0.9$ and a false positive rate $r = 0.1$ for event correspondence.

The transport network for nodes is modelled by a single routing instance with adjustable delay and datarate on all links, simulating Internet transmission properties. The applications within the cameras use UDP for communication and all nodes are assumed to reach others directly via the communication infrastructure.

After a simulated time of 3600s we get the result that is depicted in Figure 5. In this case we used broadcasting of events for all nodes to reach global

knowledge and demonstrate the accuracy of our implemented estimation algorithm, when the bound detection rate and false positives are deactivated.

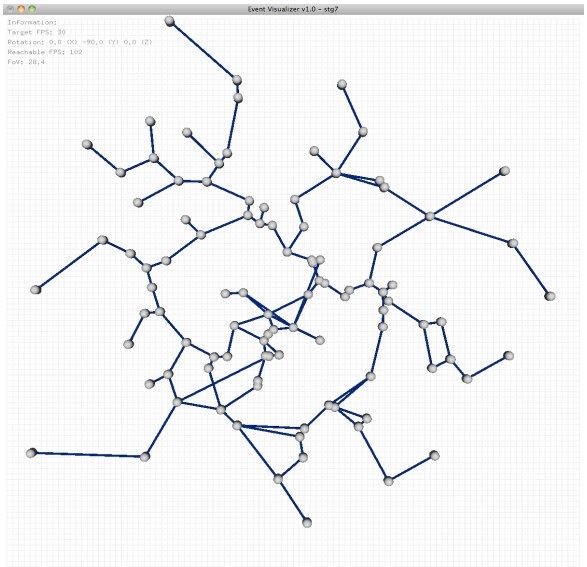

Figure 5 Estimated Vision Graph (Broadcasting)

Based on the first simulation run, we can now have a look at the evaluation of the proposed clustering algorithm that we expected to be more sophisticated, concerning communication expenditure. Again, Figure 6 shows the estimated neighborship graph, but now incorrect and/or missing edges can be seen due to the limited communication.

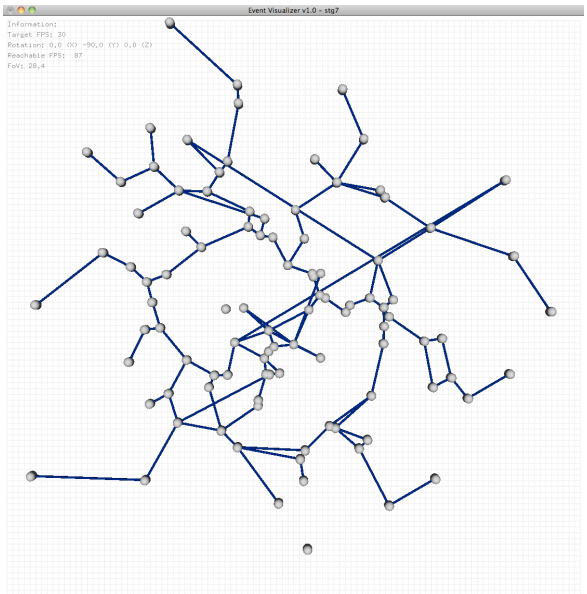

Figure 6 Estimated Vision Graph (Clustering)

In order to underpin the advantage of the saved communication resources that were traded in for a slightly worse detection rate, another experiment was conducted. Here the accumulated message frequency for node communication was recorded in 32 independent simulation runs for each scenario to reach statistical significance. Figure 7 shows the distribution of message frequencies over different node counts, as well as the 99% confidence interval and minimum/maximum. According to expectations, broadcasting needs much more messages to be transmitted than the optimized clustering algorithm. Preferring the communication with detected neighbors and reducing the notification of potential non-involved nodes can save resources, while only slightly delimiting the quality of topology estimation.

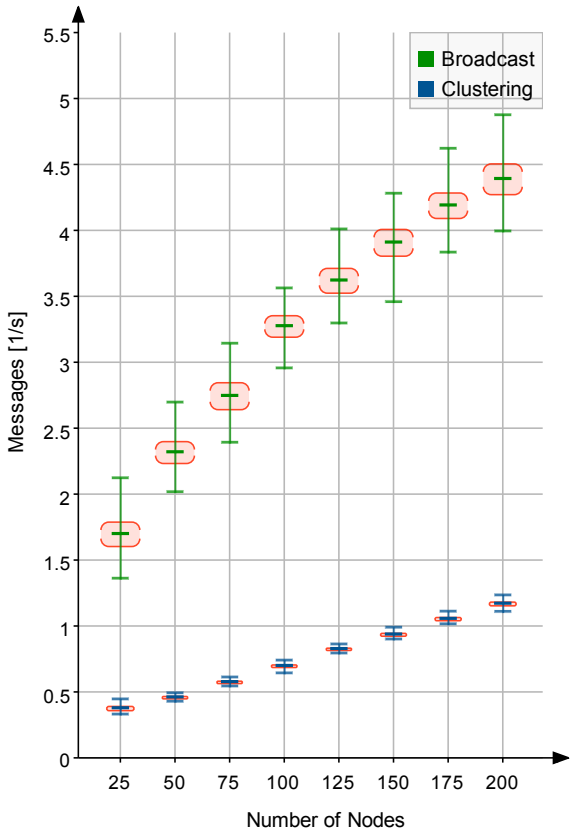

Figure 7 Comparison of communication efforts in reference system and presented approach

In order to allow for a detailed evaluation of parameter changes and influences of different disturbance variables, metrics have to be developed and tested. With our further research we want to provide not only new approaches to DSC self-configuration, but also an extensible framework to evaluate these distributed algorithms and visualize processes in real-time. The latter should help developers monitoring the system behaviour and identify problems.

5 Conclusion

Within this article, a first approach for modeling and evaluating a distributed topology estimation algorithm based on smart camera networks was presented. We introduced a high-level description of our simulation framework and named objectives for future camera networks.

Ongoing research will focus on the development of more sophisticated distributed algorithms for object and situation detection, as well as aspects of efficiency, security, and models for realistic camera environments.

References

[1] B. Rinner and W. Wolf, "An Introduction to Distributed Smart Cameras," *Proceedings of the IEEE*, vol. 96, Oct. 2008, pp. 1565-1575.

[2] M. Valera and S. Velastin, "Intelligent distributed surveillance systems: a review," *Vision, Image and Signal Processing, IEE Proceedings-*, IET, 2005, p. 192–204.

[3] Y.M. Mustafah, A. Bigdeli, A.W. Azman, and B.C. Lovell, "Face detection system design for real time high resolution smart camera," *Third ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC)*, IEEE, 2009, pp. 1-6.

[4] R.J. Radke, "A survey of distributed computer vision algorithms," *Aghajan (Eds.), Handbook of Ambient Intelligence and Smart Environments*, Citeseer, 2008, pp. 1-21.

[5] V. Atienza-Vanacloig, J. Rosell-Ortega, G. Andreu-Garcia, and J. Valiente-González, "People and luggage recognition in airport surveillance under real-time constraints," *19th International Conference on Pattern Recognition ICPR*, IEEE, 2008, p. 1–4.

[6] G.-F. Lu, Z. Lin, and Z. Jin, "Face recognition using regularised generalised discriminant locality preserving projections," *IET Computer Vision*, vol. 5, 2011, p. 107.

[7] Y. Fu, G. Guo, and T.S. Huang, "Age synthesis and estimation via faces: a survey.," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, Nov. 2010, pp. 1955-76.

[8] D. Kim and J. Paik, "Gait recognition using active shape model and motion prediction," *IET Computer Vision*, vol. 4, 2010, p. 25.

[9] V. Maik, D.T. Paik, J. Lim, K. Park, and J. Paik, "Hierarchical pose classification based on human physiology for behaviour analysis," *IET Computer Vision*, vol. 4, 2010, p. 12.

[10] M. Murakami, J.K. Tan, H. Kim, and S. Ishikawa, "Human motion recognition using directional motion history images," *Control Automation and Systems (ICCAS), 2010 International Conference on*, IEEE, 2010, p. 1445–1449.

[11] A. Caragliu, C. del Bo, and P. Nijkamp, "Smart cities in Europe," *3rd Central European Conference in Regional Science – CERS*, 2009, pp. 45-59.