

Network of Cooperating Smart Sensors for Global-View Generation in Surveillance Applications

Dietmar Bruckner
Member, IEEE

Vienna University of Technology
bruckner@ict.tuwien.ac.at

Rosemarie Velik
Member, IEEE

Vienna University of Technology
velik@ict.tuwien.ac.at

Gerhard Zucker
Member, IEEE

Vienna University of Technology
zucker@ict.tuwien.ac.at

Abstract – This paper presents the framework of a novel approach to combine multi-modal sensor information from a network of distributed smart nodes possessing audio and video modalities for gaining valuable supplementary information compared to traditional video-based observation systems or even just CCTV systems. The nodes are equipped with wireless communication modules allowing high-range, medium-bandwidth, and secure communication in a “cable-replacement” fashion. Some of the nodes also possess wired Ethernet connections to allow the transport of video streams to the supervisor. In a stepwise procedure the nodes first discover neighbor nodes by means of wireless connection strength and then try to find overlaps on the semantic level by utilization of the loopy belief propagation algorithm. This procedure ensures scalability over the network and allows establishing a global view of behavior in some region shared by several nodes. The processing architecture including the physical sensor nodes is called SENSE (smart embedded network of sensing entities) [1, 2].

I. INTRODUCTION

Due to increased requirements regarding human safety, observation systems become more and more powerful in terms of number of observing nodes on the system level as well as object and person detection algorithms on the node level. This leads to large amounts of data presented to the security staff on a number of screens. The security operators have to monitor some ten up to a hundred of screens at the same time – trying to find unusual or dangerous situations that need to be further investigated by personnel on site.

What is still missing in current observation systems is learning in terms of auto-calibration and situation awareness. A basic situation aware system could for example also provide the security staff with a rating of how “normal” the current activities in the field of the nodes are and therefore limit the number of nodes to be monitored at the same time. The auto-calibration capability is required to significantly lower installation and maintenance costs.

The SENSE system¹ has the aim to detect scenarios and report unusual or explicitly dangerous situations to the operator. This scenario detection is done in a hierarchical processing architecture consisting of 9 layers, including

amongst others, low level feature extraction, tracking, estimating probability of several parameters, checking against rules for dangerous situations, and reporting to the user [3]. One important feature of SENSE is the sharing of information with neighbors that have a common observation field. How this is recognized and the implications for bandwidth-efficient routing are the main focus of this work and discussed in later chapters.

II. SYSTEM DESIGN

SENSE intends to overcome problems with current centralized networks. SENSE uses a completely decentralized approach (see Fig. 1). The system consists of a number of identical, autonomous acting nodes mounted at fixed locations. Each node possesses sensors with static sensing parameters (in our test application these are cameras without zoom or tilt and microphones with stable characteristics), gathers information from its surroundings, and interprets it. Consequently, each node can be seen as a stand alone system. However, the nodes are able to share that knowledge with their neighbouring nodes. Each node acquires knowledge not just from its own sensors, but from its neighbours, and integrates this knowledge into its own world view. In this way, a global world view is created autonomously.

III. VIDEO SURVEILLANCE SYSTEMS MARKET

From the architectural point of view, contemporary systems are mainly centralized. In older systems video signals are gathered through an analog video network and coded and processed in control rooms or server racks. More actual systems are deployed on IP networks (see Fig. 2). Embedded coders/decoders are attached to general purpose cameras to code images locally. These video streams are available in the network to be store and processed on-line or off-line. This kind of architectures constitutes a step forward to the use of distributed computing. Examples of these systems are:

- **IBM digital video surveillance** provides an integrated video surveillance and security solution that helps organizing view, monitor and digitally recording activity

¹ This work was supported in part by the European Commission under Grant 033279 FP6-2005-IST. The website address is www.sense-ist.org.

while archiving video over an IP network, allowing digital streaming and enabling real-time access.

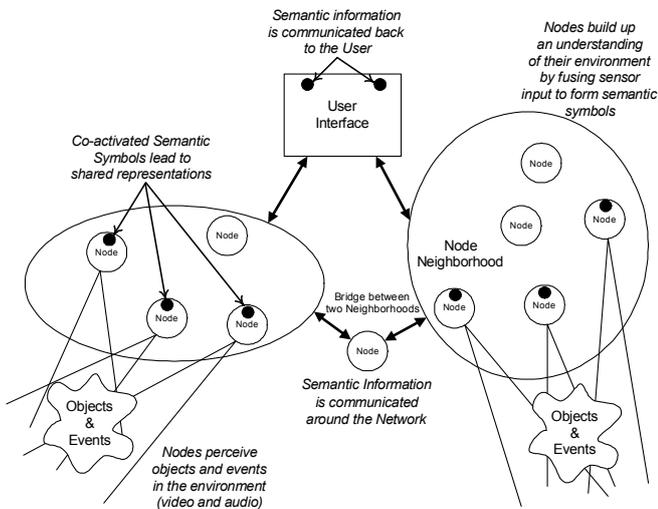


Fig. 1. SENSE approach

- **GE's VisioWave IVP** platform delivers an architecture that leverages existing infrastructure systems and equipment such as CCTV keyboards and domes. It monitors system health and analyzes video streams to report camera performance and alerting users to camera problems. It also allows adding intelligence with third party plug-ins.
- **iOmniScient Intelligent Surveillance** provides a suite of intelligent video analytics software which will assist in improving safety and security. They offer a wide variety of applications; their IQ Product Series is open in that it can cope with both IQ cameras as well as with analogue cameras. Further it can be easily connected into existing CCTV infrastructure. The system is highly scalable and can be distributed or centralized which means they offer centralized and “distributed” servers for processing.
- **IndigoVision** provide users with powerful tools to assist in the complex task of security monitoring and control. Using the plug-in technology, algorithms analyze live video locally and identify predetermined events of interest as they happen.

The processing performed in local embedded encoders is not limited to simply encoding/decoding video signal. It can be extended by plug-in installation. Manufacturers provide plug-ins for sophisticated tasks like tracking and motion detection and sometimes a framework for plug-in development.

However, even distributed video processing systems are considered centralized from the SENSE system point of view. The SENSE system is focused on high-level processing, not only processing of raw sensor signals. Alarms are generated by local processing in a node taking into account multimodal local sensor fusions and high level information

correlations among neighborhood nodes. A commercial approximation of this kind of systems is:

- **Blue Eye Video** offers multi-sensory systems that give information to understand the customers or visitors behavior. For instance B-TRACK uses data from a network of sensors to analyze pedestrian paths.

Finally, it should be noted, that besides the mentioned drawbacks of the market solutions, there are serious reservations on to what extent the often ambitious claims of the manufacturers can be kept. Without actually buying or test-installing one of those systems one cannot seriously compare the actual performance.



Fig. 2: General structure of a video surveillance system. Image taken from [4].

Another important point is that the stand alone video systems are optimized to suppress some kind of messages or on the other hand to generate particular kinds of messages also in case of quite low probabilities for a particular event. The latter can be produced by smoothing algorithms. By using such a system within the SENSE architecture we would have no figures about this and it would be a hard problem to design for example a tracking algorithm or a parameter inference machine without having reliability measures of the underlying processing of events.

IV. SYSTEM SET-UP

This chapter gives an overview on the SENSE system on system level view by describing initialization, node removal and neighbour establishing.

The side conditions for designing the system are:

- The goal is the provision of a conceptual any-to-any node communication
- In a stable system, a node on the semantic, the application level, communicates only with neighbours with overlapping sensing areas, e. g. overlapping camera views
- A maximum of four semantic neighbours per node is assumed due to bandwidth restrictions and the assumption

that there is no benefit in having more than four nodes observing the same area

- Configuration of neighbourhoods is performed solely through detection of correlation (see description of Loopy Belief Propagation algorithm below)
- Hence, at start-up, each node communicates with “every” other node (Fig. 3); wireless coverage (reception range) depends on environmental conditions and is therefore not ideal circle; therefore, stars are used in the pictures below whenever non-ideal situations are represented.
- Principally, wireless communication is assumed. Therefore, directly reachable nodes are a superset of nodes with overlapping sensor views (because of the large coverage of the wireless module), but only a subset of all nodes in the network.
- In current implementation, two kinds of nodes exist with respect to communication: “Simple” nodes are able to

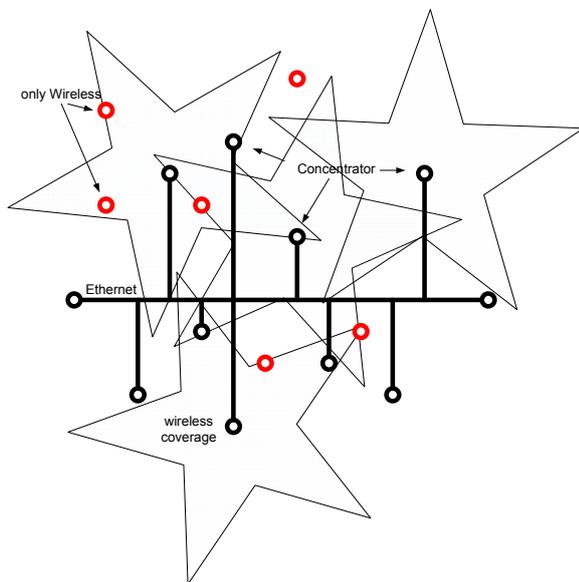


Fig. 3: SENSE system unconfigured

communicate directly with one “concentrator” node.

- “Concentrators” perform local communication between their wirelessly connected nodes
- They also communicate with user interface nodes via ethernet cable, in order to enable image/video transmission from sensor nodes to the user interface.
- Ethernet is also used to communicate with other concentrators
- Each concentrator with its wirelessly associated nodes establishes a local cluster.
- Each concentrator has up to four neighbour concentrators, with which it exchanges at startup information from its local cluster nodes. These will be learned via wireless communication as in simple nodes.

- During learning phase, nodes detect their real neighbours. This information is used to restrict communication among those nodes by the concentrators.
- Establishing of the neighbourhoods will take some time due to the necessity of finding correlations between locations in the local views of nodes on the semantic level. The steps are:
 - A DHCP server is assumed to provide new nodes with IP addresses in order to avoid having this capability on each concentrator node and not knowing which one has to perform the service. This server can be the same PC which runs the user interface.
 - So, each node tests, if it is connected to the wired network. If it is, it treats itself as concentrator and requests an IP address from the DHCP server.
 - If not, it sends a broadcast wireless message to all reachable nodes asking for a concentrator (Fig. 4).
 - If there is none, it builds a micronet with its reachable neighbours and let the loopy belief propagation algorithm find a maximum of 4 neighbours with the highest area of overlap and good connectivity.
 - One node will be negotiated to control the message flow as the concentrator would.
 - Periodic broadcasts from the nodes in the micronet will be used to find a concentrator node when such node will be powered on.
 - If there is just one concentrator available, the new node connects to that one and starts finding a semantic overlap with it.

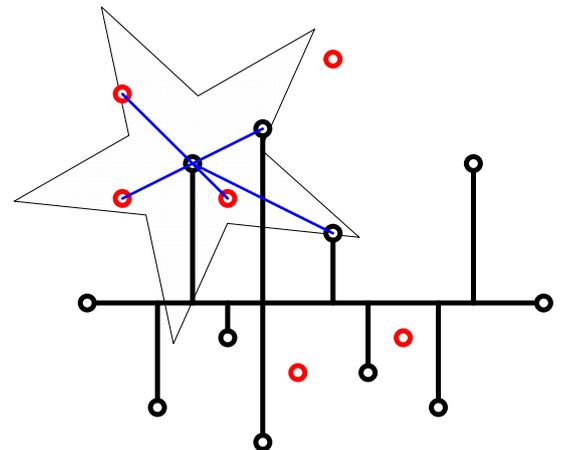


Fig. 4. Wireless coverage of one concentrator node

- If there are more than one, it has to find the concentrator with best connectivity.
- After having found its wireless neighbours the nodes reduce their transmission power in order to avoid disturbance of also covered, but “unimportant” neighbour nodes (Fig. 5).

- Finally, neighbours don't just have to be found dependent on wireless connectivity, but also upon wired connected nodes.

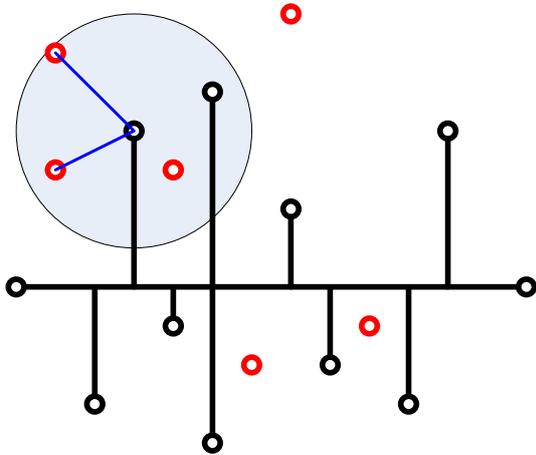


Fig. 5. Neighbouring nodes of one concentrator with strongest connection

- Therefore concentrators also ask for wirelessly reachable neighbour concentrators and then try to find physical overlaps of sensor areas.
- This policy – with the additional boundary condition that the transmission starts with very low power at the beginning and increases step-by-step – will ensure scalability within the SENSE system.

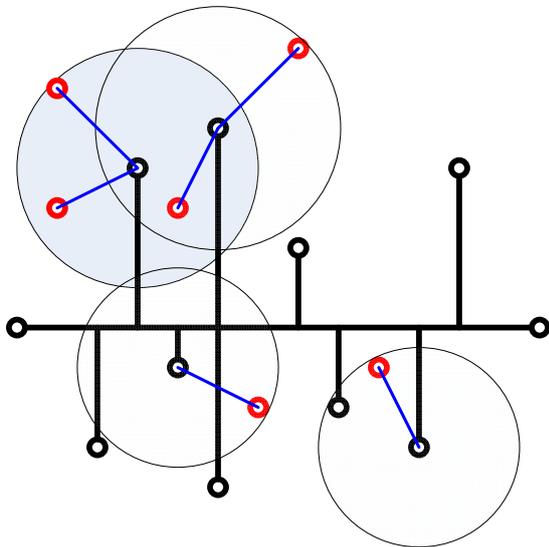


Fig. 6. Connections on the physical layer

- After finding correlations between wired connected nodes and between concentrators and purely wireless nodes, correlations between purely wireless nodes connected with different concentrators have to be found.

- Therefore concentrator nodes will forward LBP messages from their wireless nodes via another concentrator to another wireless node.
- Due to wireless bandwidth restrictions this task is of lower priority than normal LBP messages between concentrators and wireless nodes.
- Having this considerations in mind while designing the installation, most of this cases can be avoided by clever arrangement and sufficient amount of concentrators (Fig. 7).
- Hence, at startup a slightly different negotiation procedure takes place as with a purely wireless network, but the higher bandwidth is provided by the Ethernet which discharges the wireless communication.
- Once, neighbourhoods are recognized, communication is restricted among "overlapping" nodes on a semantic level, the network topology is hidden.
- When a simple node is added, the abovementioned procedures take place, meaning its concentrator distributes its messages to the local nodes and the neighbour concentrators, which causes to re-configuration using the loopy belief propagation algorithm.

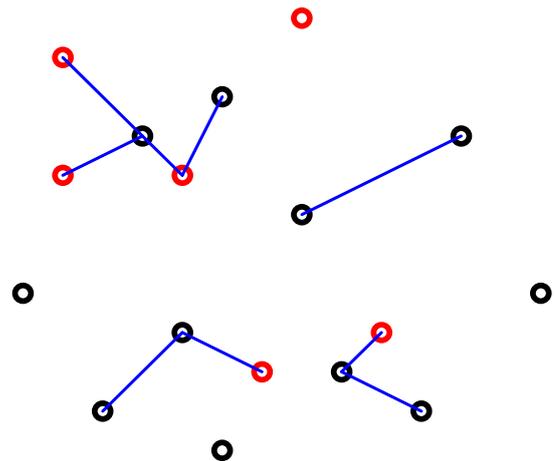


Fig. 7. Connections on the semantic layer – ignoring the user messages

- When a simple node is removed, the neighbours communicating with it detect it, report it, and then re-adapt their correlation information.
- When a concentrator is removed, this corresponds to the loss of a bridge node in Fig. 1, together with the associated cluster. This leads to reconfiguration of the wireless network topology. Either all simple nodes find another concentrator – which will be the normal case because of the high range of the chosen transceivers – or the remaining simple nodes will form a micronet and wait for a concentrator.
- The other nodes in the SENSE system which had overlap with one of the lost nodes will find out about the network break and inform the user.

- The last possible case – that the lost network has no overlap with the remaining – will be covered by the periodic broadcast messages for updating the network.
- Lost connections will be reported to the user.

V. LOOPY BELIEF PROPAGATION

In general belief propagation is an iterative algorithm for computing marginal distributions of functions on a graphical model. Belief propagation is, for example, applied in pair wise Markov random fields or Bayesian networks. Loopy Belief Propagation (LBP) is an extension of the belief propagation framework from polytrees to general networks and was developed by Pearl [10]. Generalization implies that graphs can have loops, thus the name. Convergence of LBP may be an issue, since it does not always converge; however, non-convergence cases are rare.

The goal is to create a global view of normality, i.e. the system learns not locally in a node, but globally which behavior is common. The nodes have overlapping sensing areas (not all of them, but some) and have a symbolic (statistical) representation of normality in their local views. For establishing global views about normality in a "stochastic" environment we need a method with the following properties:

- it has to be based on statistical methods,
- it has to be decentralized, and
- it is computationally inexpensive

The belief propagation algorithm for general (loopy) graphs is based on statistical methods and is decentralized. Since its creation 20 years ago, it has been widely adopted in application areas like error correcting codes (e.g. the "Turbo code" [5]) or machine learning [6]. Recently it has been a candidate for application for sensor networks communication [8] as well as image understanding [7].

the Loopy Belief Propagation iteratively updates stochastic variables in nodes; this approach appears well suited for the system concept described in this paper, since it employs decentralized processing. It has furthermore been shown that The LBP algorithm has proven to be effective in learning contexts, where the Expectation-Maximization (EM) algorithm is used to learn model parameters, and LBP is used to perform an approximate calculation of the required expectations. This form of approximate EM algorithm is theoretically well-founded, and can be seen as a minimization of a Bethe free energy [9]. Loopy Belief Propagation also performs well empirically. Again, this fits well within the system concept, where parameters will be learned using an Expectation-Maximization algorithm with LBP as the approximate inference algorithm. Additionally, LBP is very effective in sparse network topologies consisting of long chains of neighbors [6], which can be translated to a sensor network topology as described above.

At run-time the system undergoes two different states: in the first state the system starts up and infers correlations between different nodes, that is, between different semantic symbols in

the nodes. It does so by sending broadcast messages to all other nodes in range. After this state the system goes to the second state where it has learned or inferred all the correlations and sends messages only to those nodes that are really neighbours (in the sense of having semantic symbols that are correlated with the local semantic symbols).

VI. CONCLUSION AND OUTLOOK

In this paper a novel approach for setting up the network architecture of a multi-modal multi node surveillance system is presented. The semantic processing architecture finds overlaps between nodes on a semantic level and restricts the communication accordingly. Due to the wireless signal strength measuring-based prioritized neighborhood detection procedure scalability of large-scale SENSE systems is ensured.

REFERENCES

- [1] www.sense-ist.org (SENSE project website)
- [2] G. Zucker (ne Pratl), and L. Frangu: "Smart Nodes for Semantic Analysis of Visual and Aural Data", Proceedings of the IEEE INDIN, p. 1001-1006, 2007
- [3] D. Bruckner, J. Kasbi, R. Velik, and W. Herzner: "High-level Hierarchical Semantic Processing Framework for Smart Sensor Networks", to be published on the IEEE HSI, 2008
- [4] <http://www.indigovision.com/learnabout-ipvideo.php> (Indigo Vision IP Video Management Introduction)
- [5] R. McEliece, E. Rodemich, and J. Cheng. "The Turbo decision algorithm." In Proceedings of 33rd Allerton Conference on Communications, Control and Computing, Monticello, IL., 1995
- [6] K. Murphy, Y Weiss, and M. Jordan, "Loopy belief propagation for approximate inference: An empirical study." In Proceedings of Uncertainty in Artificial Intelligence, 1999, pp. 467-475
- [7] James M. Coughlan and Sabino J. Ferreira, "Finding Deformable Shapes Using Loopy Belief Propagation." Computer Vision -- ECCV 2002, Lecture Notes in Computer Science, Springer Berlin/Heidelberg, Volume 2352/2002, pp 53-73, 2002.
- [8] C. Crick and A. Pfeffer, "Loopy belief propagation as a basis for communication in sensor networks", In Proceedings of Uncertainty in Artificial Intelligence (UAI), 2003.
- [9] T.M. Heskes, W.A.J.J. Wiegierinck, O.R. Zoeter. "Approximate Expectation Maximization." In Proceedings NIPS 16, volume 16 (pp. cd-rom), 2004.
- [10] Pearl, J., "Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference", Revised Second Printing, Morgan Kaufmann, San Francisco, 1988