Managing the Information Flow in Visual Sensor Networks

K. Obraczka
CE Department, UCSC
Santa Cruz, CA 95064
katia@soe.ucsc.edu

R. Manduchi
CE Department, UCSC
Santa Cruz, CA 95064
manduchi@soe.ucsc.edu

J.J. Garcia–Luna–Aveces
CE Department, UCSC
Santa Cruz, CA 95064
jj@soe.ucsc.edu

Abstract

Sensor networks, or sensor webs, which consist of a large number of interconnected sensing devices, have been the subject of extensive research in the past couple of years. Typical applications of sensor networks include monitoring of possibly very large, remote and/or inaccessible areas, surveillance, smart environments like meeting rooms, buildings, homes, and highways.

Our focus is on visual sensor networks, which are networks of cameras equipped with enough processing power to support local image analysis. This paper describes ongoing research at UCSC in visual sensor networks and highlights the research challenges to be addressed. It motivates the need for the tight coupling between vision techniques and communication protocols for more effective monitoring/tracking capabilities (by having sensors operate in a coordinated manner), as well as energy- and bandwidth-efficient protocols which will prolong the operational life of the sensor network.

Keywords

Sensor networks, distributed sensing, network–level QoS, data aggregation, power aware clustering.

INTRODUCTION

In recent years, considerable attention has been given by the research community to the design of large ensembles of interconnected sensors (usually called sensor networks or sensor webs). Some of the key enabling technologies for the development of such systems include the increasing availability of cheap and miniaturized sensing devices, embedded computers, wireless connection, and power supplies. They provide an ideal solution to applications that require monitoring possibly very large, remote and/or inaccessible areas over extended periods of time. To meet these applications' needs, sensor networks must be scalable, fault–tolerant and self–managing.

Possibly the biggest single challenge faced by researchers in this field is the management of the information flow within the network. These systems are typically highly asymmetric: all sensors are "sources" of new data, but the recipients (the "sinks") are much fewer than sensors - perhaps just one central unit. Not all the information produced by all sensors, however, is of interest to the final user at all times. Furthermore, various forms of correlation are usually presents across the network, especially if some redundancy is allowed in the sensor placement to guarantee higher robustness and fault tolerance. The key to managing the possibly huge amount of information produced by the sensors is to make the best use of available local computational power. Local processing (at the sensor level or in intermediate nodes) allows one to control the data acquisition process, remove cross–sensor correlation, aggregate data streams, in such a way that only what is really needed is routed to the user.

This paper describes ongoing research at UCSC in the field of "Visual Sensor Networks", that is, networks of cameras that are equipped with enough processing power to support image analysis functionalities. Networks of visual sensors are the solution of choice for a number of societal, research, and educational applications, including:

SURVEILLANCE: Protection of large facilities (airports, plants, stadiums) requires that mechanisms for detecting and tracking intruders over large areas be put in place. If a large number of miniaturized cameras is disseminated throughout the facility, events may be detected and analyzed by visual processing, and video streams of interest may be transmitted to the operator control unit.

ENVIRONMENTAL MONITORING: There are many situations in which vast and/or inaccessible areas should be visually monitored to detect unusual events or to acquire environmental data over long periods of time. Examples include toxic locations, disaster sites, traffic control in freeways, as well as natural environments such as forests, deserts, and even planetary exploration.

SMART MEETING ROOMS: Meetings or lectures involving remote participants need a higher level of interactivity than currently available. In order to actively participate to a discussion or to a lecture, remote users should enjoy a rich, dynamic visual experience. The current speaker should be tracked as he or she moves in the room, and visual attention should be switched to interveners in the discussion when appropriate. "Autonomous videography" requires that a sufficiently high number of cameras (possibly omnidirectional, or with pan/tilt/zoom capabilities) are placed in the room, and that the cameras coordinate among them or by means of a central controller to select and transmit the "best" view at each time.

One main problem in sensor networks design is the management of the information flow in the network. In the case of webs of cameras, this problem is multifaceted. The amount of data produced by a camera depends on the abstraction level at which this information is represented, as computed by onboard processing. Visual data streams from cameras detecting interesting events should be routed with high priority to the final user. Cameras may cooperate, exchanging information with specific quality of service requirements. Thus, in order to design a visual sensor network to accomplish a specific task, it is first necessary to model and quantify all communications needs peculiar to cooperative visual processing, and then to translate such needs into specific constraints on the underlying communication network. Be-

sides “traditional” power conservation issues, visual sensor network protocols (at the network and media access control layers) are faced with a number of new challenges such as supporting multiple, potentially high bit rate flows with different priorities, handling communication among heterogeneous devices (cameras, control units, gateways to the wired infrastructure, etc.), and ensuring different levels of service guarantees (e.g., delay and delay jitter).

The remainder of this paper is organized in two main parts. Section deals with the visual sensor processing part of the system, and outlines the proposed research activity related to the communication of different scene representation levels and to sensor collaboration. Section describes our work in the communication network part of the system, hilighting a number of different aspects of the problem: interconnection protocols, channel access with QoS and power constraints, network–level QoS, data aggregation, and power–aware clustering.

VISUAL SENSOR PROCESSING

Over the past two decades, computer vision research has progressed remarkably, both in terms of algorithm development and of system integration. Yet, very little attention has been given to the job of communicating the results of the vision action. Also, cooperative vision, whereby two or more cameras with possibly overlapping fields of view extract visual information over extended areas in dynamical situations, is still a largely unexplored task, especially when realistic inter–sensor communication scenarios are considered. Shaping the information flow between cameras (in a decentralized scheme) or between each camera and a central unit (in a hierarchical scheme) so as to maximize the outcome of the vision task, while adapting to the current network constraints, is still an open problem. In the following, we analyze such problems in detail and discuss possible solutions.

Visual Representation and Communication

Consider an outdoor surveillance application, whereby a camera observes a region of space with the goal of detecting and reporting any intrusions. Current mainstream commercial technology would simply transmit the visual stream as generated by the camera for monitoring by a human in the operator control unit (OCU), flooding the network with mostly useless information (i.e., the background in the absence of intruders). To reduce the bit-rate, motion-compensated video encoding (MPEG-style) may be employed. The data rate using traditional techniques, however, may still be too high for bandwidth–constrained networks. If the peak bit–rate the network can support is fixed, motion–compensated encoding will typically cause high lossy distortion in moving areas [20], thus making the recognition task by a human operator harder. Reducing the frame rate is another naive technique for data reduction, but it increases the risk of missing short transient events. If there is enough onboard computational power, motion segmentation can be performed at the camera level, by means of motion vector clustering [19] or parametric layered image representation [15]. More visual features (color, 2-D shape), together with available a–priori knowledge and modeling, may be used to classify image segments and assign different priorities to them. In this context, the priority of a segment corresponds to its probability of containing an “intrusion" event. The prioritized data may be encoded using Region-Of-Interest (ROI) progressive compression (such as in JPEG-2000 [16]), which allocate more bits to areas with higher priority.

More visual processing may be implemented to track image segments from frame to frame, allowing one to 1) gather more information about the moving areas (thereby enabling better event classification), 2) predict the future trajectory of a moving object (thus coordinating with other cameras, and 3) further reduce the output data rate, by only transmitting “innovative" information that was not predicted during the tracking process (this can be seen as a higher form of motion–compensated encoding). If the network cannot support the transmission of image segments, then the segmentation map, together of a set of parameters for each segment (such as the distribution of color and texture and/or a parametric motion description) may be transmitted instead, reducing the bit-rate considerably. If onboard computational power permits, higher recognition tasks can be performed at the camera level, including: classification of a moving object as human–not human, gait analysis, face recognition. Information over multiple frames can be condensed into a small set of parameters and, ultimately, represented by a binary decision on whether an intrusion happened.

This example has outlined a possible hierarchy of task–oriented scene representation levels. Depending on the available bandwidth and on user requirements, the sensor may transmit: 1) the full video stream, 2) ROI–encoded segmented images, possibly at different frame rates, 3) segmentation maps with related parameters, 4) scene classification with confidence measures, all the way down to a single bit describing the event.

Sensor Collaboration

The main reason for deploying visual sensor networks is to cover a wider area than possible with a single camera. However, if two sensors are not too far apart, one may expect some degree of correlation between what is viewed by them. The most obvious case is when the field of view (FOV) of the two cameras overlap. Viewing the same surface from two different positions has two main advantages. Firstly, event detection can be made more robust by cross-validating information. For example, if two cameras in a surveillance system see at the same time a moving object approximately located in the same position in space, then the confidence of this event detection is increased. Standard sensor fusion techniques may be implemented for this purpose. Secondly, range information can be estimated by triangulation, allowing for a very powerful geometric scene description. We will call such forms of cooperation “spatial–based collab-

oration". Spatial–based collaboration requires that: (i) It must be known which couples of cameras have overlapping FOVs; (ii) All camera pairs with overlapping FOVs should be jointly calibrated, by recovering their epipolar geometry [8]. Pairwise relative positions may then be propagated to global relationship in the sensor web. (iii) The cameras must be temporally synchronized for the image registration to be meaningful. Even if two cameras are too far apart for their FOV to overlap in the operative range, they still may cooperate effectively. Assume that one camera is tracking a moving object (e.g., a potential intruder in an area under surveillance, a wild animal in a monitored region, a walking person in a smart meeting room or classroom). This camera may signal a nearby camera that the object is about to enter its field of view, effectively "handing off" the tracked entity [2],[4]. This is an instance of "predictive collaboration". An interesting form of predictive collaboration can be implemented when a camera can operate at different power/performance modes for energy conservation. For example, in a low–performance mode the camera's processor would execute image analysis at low resolution and low frame rate. Once an unusual event has been detected, the camera would switch to high–performance mode, enabling higher resolution image analysis (for better localization) and higher frame rate (for better tracking.) Note that the available computational resources may also be efficiently managed by means of a "window of attention" strategy, that is, by performing higher resolution analysis only an the image region most likely to contain interesting information. As a camera detects a moving event, it may send an "alert" packet to the nearby sensors, advising them to switch to high–performance mode and be prepared to a likely event detection. If other cameras with overlapping FOVs detect the same event, the moving surface may be triangulated to infer its position and its velocity in 3-D space. Motion information may be used for predicting which cameras are most likely to see the moving body next, and therefore should be alerted, possibly by multi–hop transmission.

Given the two types of collaborations among cameras discussed above, one should address the issue of what kind of information should be exchanged among cameras (in a distributed system) or between each camera and the control unit (in a centralized system) for efficient collaboration, and how this translates into QoS requirements for the underlying network. The kind of information that needs to be exchanged locally within the network depends on the representation domain of interest for a particular task. For example, stereo triangulation requires in principle that one camera can access the full image produced by the other camera in the pair. Since this may require too high a data rate, one may considered a reduced representation, for example as formed by suitable sets of features (such as edges or contours, or perhaps richer local descriptors as mentioned earlier), which could be encoded much more compactly. Another image–level instance of information exchange is related to target hand–off between cameras when range information is not available. However, target hand–off, as well as a number of higher–level prediction and recognition tasks, may be implemented much more efficiently if 3–D information (as produced by stereo matching) is available. This kind of data can be represented in symbolic form, for example by means of a parametric model of the moving object. Another kind of 3–D representation proposed in the literature is based on occupancy grids [5],[11],[17]. With such a mechanism, local reasoning may be used to infer changes in the occupancy map (which is assumed to be shared by the different cameras in the cluster), and only such changes are transmitted to the other cameras.

SENSOR COMMUNICATION

One of the main contributions of this research is to promote synergy between the fields of computer vision and (wireless) computer networks, addressing the requirements of scalable, distributed visual sensor webs. As discussed in the previous section, the ability of sensors to communicate and collaborate in accomplishing tasks is critical for the types of applications we target. In this section, we describe the research challenges associated with developing communication protocols for visual sensor interconnection. Knowledge of the current state of the network is critical to evaluate the optimal representation level (with associated data rate) to be used at any given time. Power efficiency is another constraint sensors and control units need to consider when deciding what data representation level to generate and transmit. Visual sensing imposes different service requirements on the underlying network. One typical requirement is delay and jitter guarantees for sensor collaboration. Another form of quality of service that needs to be supported is prioritized data streams which is critical to enable reactive behavior.

Interconnection Protocol

One of the innovative aspects of our research is to design new network- and MAC-layer protocols that explicitly address the requirements of visual sensing applications. This tight coupling between vision techniques and communication protocols will result not only in more effective monitoring/tracking capabilities (by having sensors operate in a coordinated manner), but also in energy- and bandwidth-efficient protocols which will prolong the operational life of the sensor network.

In the network we are developing, sensor nodes use the Flexible Interconnection Protocol, or FLIP [12], as the underlying communication protocol. FLIP is a network-layer protocol designed to interconnect devices with varying power, communication, and processing capabilities. Through the use of customizable headers, FLIP can offer close to optimal overhead for limited-capability devices in one extreme, and yet can still provide full functionality for more powerful devices in the other extreme. We will use FLIP to handle communication: (1) among sensors, (2) among control units, and (3) between sensors and control units. Visual sensor networks may also be connected to the wired IP infrastructure.

Gateways will be responsible for performing the translation between FLIP and IP and vice-versa. Flip must account for the needs of visual sensor networks including deciding which functionality the protocol should (or not) provide (e.g., types of routing algorithms, reliability, security). FLIP's flexible header will allow us to customize the protocol for the specific needs of vision–based applications. For example, in order to implement prioritized flows, we will incorporate priority information in the FLIP header. FLIP's flexible headers allows functionality to be tailored not only to specific applications but also to specific devices. In a hierarchical sensor structure, where group of sensors report to a local control unit, priority information may be carried just by streams emanating from the control units as a result of their exchange with local sensors. Priority information may then be used by intermediate nodes when routing/forwarding data. A simple binary priority scheme may be used, where flows are either prioritized or not. We will also study a n-ary scheme than can work in concert with hierarchical data representation.

FLIP will be for the visual sensor network what IP is for the Internet: it will "glue" all components together, provide required functionality (as appropriate for each type of component), and enable coexistence of various core network mechanisms. For instance, through FLIP, we can incorporate different sensor network communication paradigms such as directed diffusion [10] and SPIN [9], as well as new mechanisms needed to address the specific requirements of visual sensing applications (e.g., point-to-point communication to handle data flows between sensors and sinks, prioritized flows, efficient data collection, and topology management).

Channel Access with QoS and Power Constraints

Medium-access control (MAC) protocols based on collision avoidance have been widely used in wireless LANs and ad hoc networks mainly due to their simplicity and good performance compared to carrier sensing multiple access (CSMA). With a collision-avoidance MAC protocol, a node that needs to transmit data to a receiver first sends a request-to-send (RTS) packet to the receiver, who responds with a clear-to-send (CTS) if it receives the RTS correctly. A sender transmits a data packet only after receiving a CTS successfully. Several variations of this scheme have been developed, including IEEE 802.11. We have shown [6] that, in order to avoid data packets from colliding with any other packets at the intended receivers in networks with a single channel, the senders have to sense the channel before sending their RTSs. This is not the case in 802.11; furthermore, recent simulation studies of 802.11, including our own, show that up to 40% of packets that are sent after a successful RTS/CTS exchange have to be retransmitted after the corresponding ACKs are not received properly.

Key limitations in applying existing contention-based MAC protocols to visual sensor networks are: (i) inability to ensure any channel-access delay guarantees, (ii) unnecessary use of transmission power during unsuccessful handshakes, and (iii) inability to support collision-free multicast and broadcast transmissions. Achieving these goals requires a conflict-free channel access method.

Conflict-free channel access protocols today are based on fixed slot assignments (e.g., TDMA) that do not scale well, topology-independent assignments (e.g., [3]) that cannot support efficient reuse of multiple channels, or topology-dependent assignments (e.g., [21, 14, 13]) that, because of their reliance on mini-slotting, can become impractical for high data rates. Furthermore, the impact of node mobility and link errors and failures has received very little attention in prior work on transmission-scheduling algorithms.

In sensor networks, it is critical to reduce as much as possible the effort needed in achieving conflict-free transmissions. Furthermore, visual sensor network applications require one-to-many communication, but, for the purpose of sensor collaboration, also need one-to-one communication. Simply applying scheduling solutions based on collision-free broadcasts to sensor networks would waste precious bandwidth when unicast transmissions are needed.

We have developed a new family of protocols for conflict-free unicasting, multicasting, and broadcasting based on topology-dependent scheduling algorithms that work on the basis of the identifiers of nodes one and two hops away [1]. In a nutshell, our distributed channel-access scheduling mechanism implements anticipatory collision resolution at each node using knowledge of the nodes that reside in the two-hop neighborhood of a node. Hence, nodes need only exchange the identifiers of their neighbors to be able to carry out conflict-free transmissions scheduling. We are investigating augmenting our activation multiple access schemes to account for differences in the bandwidth and latency requirements of different flows (e.g., to handle flows carrying data at different representation levels), the need to preserve battery life at sensor nodes, and the use of directional antennas as an additional component of distributed scheduling. Central issues in this research will be: (i) developing new hybrid activation heuristics that assign transmission priorities to the flows that depart from nodes, rather than just the nodes, in a way that flows with different QoS requirements gain conflict-free access to the channel over multiple hops from source to destination(s); (ii) incorporating the remaining battery life at nodes and the power required to listen and transmit as part of the scheduling; (iii) developing an algorithm for the exchange of neighborhood information that is efficient in its consumption of power. A key element of this is reducing the amount of time a node needs to listen to the channel in order to obtain neighborhood information needed for conflict-free scheduling; (iv) extending the notion of two-hop neighborhood to take into account the fact that nodes more than two hops away from a node may still interfere with the node, depending on transmit and receive powers.

Network-Level QoS

Traditional network control is reactive in nature and assumes a fairly homogeneous transmission medium. Another key difference is that in wireless (especially ad hoc) networks,

links can be established and terminated much more dynamically. In sensor networks, directional antennas, power control, and waveform control can be used in combination with the scheduling of transmissions to improve the utilization of the bandwidth available, as well as to preserve the battery life of untethered nodes. In essence, this means that, to deliver information more effectively, network nodes must control both the routing decisions made and the topology over which routing takes place.

In this project we propose a major departure from the state of the art by: (i) proactively using knowledge of the environment as an integral part of the network control protocols, (ii) taking advantage of multiple transmission modes and media to provide the QoS required by network users, and (iii) controlling the topology of the network in order to improve the routing of information over the sensor network. Awareness of "collateral network information" (e.g., location of routers and destinations, time, and channel characteristics) can be used in many ways to improve the behavior of traditional reactive protocols, to name a few: reducing "guard bands" of channels, using long-range or short-range links according to the location of the destination and the characteristics of the links, modifying the frequency of control information exchange to preserve power, repositioning routers to reconstitute a network, routing information to where a destination is expected to arrive, and predicting the presence of new neighbors at certain transmission powers.

We are investigating the design of a new class of network-level protocols that support QoS and are location aware, time aware, multichannel aware, platform aware, service aware, and topology aware. In particular, we are developing protocols to support QoS proactively over visual sensor networks by: (i) aggregating flows based on their classes and destinations, thus eliminating a key scaling problem of the Intserv architecture; (ii) using multiple loop-free paths (called multipaths and computed distributedly using routing algorithms we have recently designed and verified for wired and wireless nets [7, 18]) to distribute aggregated flows, which eliminates the failure-prone nature of virtual circuits; (iii) establishing signaling to reserve resources for aggregated flows only between neighbors, which is much more robust and adaptive than end-to-end signaling; (iv) integrating routing and reservation control so that packets are forwarded over multipaths, which reduces congestion and tolerates link and node failures; (v) forwarding time-critical or priority flows over multiple segments of a multipath to reduce latency or increase the likelihood of delivery; (vi) integrating routing and resource reservation with link management to control proactively and dynamically the bandwidth allocated to unicast and multicast links from a node to its neighbor(s).

REFERENCES

[1] Bao, L., and Garcia-Luna-Aceves, J. A new approach to channel access scheduling for ad hoc networks. In Proc. ACM MobiCom 2001 (2001).

[2] Cai, Q., and Aggarwal, J. Tracking human motion in structured environments using a distributed-camera system. IEEE Trans. Pattern Anal. Machine Intell. 21 (November 1999), 1241–1247.

[3] Chlamtac, I., and Farago, A. Making transmission schedules immune to topology changes in multi-hop packet radio networks. IEEE/ACM Transactions on Networking 2 (February 1994), 23–29.

[4] Collins, R., and et al. A system for video surveillance and monitoring. Tech. Rep. CMU-RI-TR-00-12, Carnegie Mellon University, 2000.

[5] Elfes, A. Using occupancy grids for mobile robot perception and navigation. IEEE Computer (June 1989), 46–57.

[6] Fullmer, C. L., and Garcia-Luna-Aceves, J. J. Solutions to hidden terminal problems in wireless networks. In Proceedings ACM SIGCOMM (1997).

[7] Garcia-Luna-Aceves, J., and Spohn, M. Transmission-efficient routing in wireless networks using link-state information. Special Issue on Energy Conserving Protocols in Wireless Networks (2000).

[8] Hartley, R., and Zisserman, A. Multiple View Geometry in Computer Vision. Cambridge University Press, Cambridge, 2000.

[9] Heinzelman, W., Kulik, J., and Balakrishnan, H. Adaptive protocols for information dissemination in wireless sensor networks. In Proceedings of the International Conference on Mobile Computing and Networking (MobiCom) (August 1999).

[10] Intanagonwiwat, C., Govindan, R., and Estrin, D. Directed diffusion: A scalable and robust communication paradigm for sensor networks. In 6th International Conference on Mobile Computing and Networking (MobiCom) (August 2000), ACM.

[11] Nakazawa, A., Kato, H., and Inokuchi, S. Human tracking using distributed vision systems. In Proc. Int. Conf. on Pattern Recog. (1998), vol. 1, pp. 593–596.

[12] Solis, I., Obraczka, K., and Marcos, J. FLIP: a flexible protocol for efficient communication between heterogeneous devices. IEEE ISCC 2001 (July 2001).

[13] Tang, Z., and Garcia-Luna-Aceves, J. A protocol for topology-dependent transmission scheduling. In Proc. IEEE Wireless Communications and Networking Conference 1999 (WCNC 99) (1999).

[14] Tang, Z., and Garcia-Luna-Aceves, J. Collision-avoidance transmission scheduling for ad-hoc networks. In Proc. IEEE ICC 2000 (2000).

[15] Tao, H., Sawhney, H., and Kumar, R. Dynamic layer representation with applications to tracking. In Proc. IEEE CVPR (2000), vol. 2, pp. 134–141.

[16] Taubman, D., and Marcellin, M. JPEG2000: Image Compression Fundamentals, Standards, and Practice. Kluwer Academic Publishers, 2001.

[17] Ukita, N., and Matsuyama, T. Incremental observable-area modeling for cooperative tracking. In Proc. IEEE ICPR (2000), vol. 4, pp. 192–196.

[18] Vutukury, S., and Garcia-Luna-Aceves, J. A simple approximation to minimum-delay routing. In Proc. ACM SIGCOMM 99 (1999).

[19] Wang, J., and Adelson, E. Representing moving images with layer. IEEE Trans. Image Proc. 3 (September 1994), 625–638.

[20] Wang, Y., Ostermann, J., and Zhang, Y.-Q. Video Processing and Communications. Prentice Hall, Upper Saddle River, NJ, 2002.

[21] Zhu, C., and Corson, M. S. A five phase reservation protocol (FPRP) for mobile ad hoc networks. In Proceedings IEEE INFOCOM (1998).