Cross-layer Optimization in Video Sensor Networks

Tommaso Melodia, SUNY Buffalo, USA
Wendi Heinzelman, University of Rochester, USA
tmelodia@eng.buffalo.edu, wheinzel@ece.rochester.edu

1. Introduction

Video Sensor Networks (VSN) (also referred to as *multimedia sensor networks* [1], or *visual sensor networks* [2]) are being made possible by the integration of low-power wireless networking technologies with inexpensive CMOS cameras and microphones. Video sensor networks are self-organizing, intelligent wireless systems of embedded and resource-constrained devices deployed to retrieve, distributively process, store, correlate, and fuse multimedia streams. VSNs can potentially constitute a viable alternative to ubiquitously deployed wired surveillance systems. Once properly regulated, the availability of affordable VSN systems will enhance the ability of private citizens and law enforcement officers to observe and monitor locations and events in an unprecedented way, enabling sophisticated real-time scene analysis. We envision that users will be able to gather information about the physical environment by issuing simple textual queries, accessing remote VSNs connected to the Internet through application-level gateways.

The characteristics of VSNs diverge considerably from wired network paradigms such as the Internet, and even from traditional sensor networks. VSN applications require the sensor network paradigm to be re-thought in view of the need for *adaptation* and *cross-layer optimization* to deliver video/multimedia content with predefined levels of quality of service (QoS). While minimizing the energy consumption has been the main objective in sensor network research, mechanisms to efficiently deliver application-level QoS (e.g., target video distortion), and to map these requirements to network-layer metrics have not been primary concerns.

In this position paper, we discuss some key research challenges in video sensor networks from a network design perspective. In particular, we discuss the need for adaptation and cross-layer optimization in protocol design, and the need for a tighter integration and co-design between networking functionalities and image/video sensing and processing.

2. Cross-Layer Networking

In multi-hop wireless networks the attainable capacity of each wireless link depends on the interference level at the receiver. This, in turn, depends on the interaction of functionalities that are distributively handled, such as power control, routing, and rate policies. Hence, capacity and delay attainable at each link are location dependent, vary continuously, and may be bursty in nature, thus making QoS provisioning a challenging task. Therefore, there is a strict interdependence among functions handled at all layers of the communication stack, which are inherently and strictly coupled due to the shared nature of the wireless communication channel. In addition to this, performance metrics in VSNs may be directly tied to the perceived video quality at the receiver rather than to traditional network metrics such as throughput. Furthermore, the required video quality may vary over time, and the network must adapt to meet these changes, providing the required video quality while minimizing resource utilization. Cross-layer protocol architectures, including information-sharing and layer fusion designs, enable such coupling of the protocols and allow adaptation for optimal resource utilization in VSNs.

Cross-layer Information-sharing. An information-sharing architecture that provides shared data repositories that all protocols can access is described in [3]. In this architecture, called X-Lisa, there is a common interface to the data repositories that allow protocols to update data and to read data when needed. This architecture includes middleware support such that application-level information can be shared with network protocols. This is vital for VSNs, where the QoS information (e.g., distortion) must be considered by the protocols for optimal resource utilization. Additionally, X-Lisa supports proactive event notification, such that protocols can subscribe to be notified of a change in any data stored in the repositories (for example, changes in link quality) or of any changes in application QoS. This enables the protocols to react immediately to important changes in the network or the application goals, ensuring efficient operation of the VSN while continuously meeting QoS goals.

Cross-layer architectures such as X-Lisa provide protocols with access to network-level and application-level information. One important challenge is how to use this information to adapt the protocols. For example, as link conditions or QoS requirements change, at the PHY layer, the transmit power or the packet length may be adjusted, while at the MAC layer, duty cycle or back-off windows may be adjusted. How best to adapt and coordinate the adaptation of the different layers and across different VSN nodes is an important area of research.

Video Quality-driven Cross-layer Optimization. An alternate approach that considers an integrated, cross-layer architecture for video streaming in VSNs is pursued in [4]. A multi-hop wireless network of video sensors deployed for surveillance applications is considered, and the focus is on reliable and real-time transport of video traffic. The objective is to design algorithms to *efficiently* and *fairly* share the common network infrastructure among the video streams generated by different video sensors, to deliver high-quality video on resource-constrained devices. To achieve this objective, the Distortion-Minimizing Rate Control (DMRC) algorithm is proposed, a decentralized cross-layer control algorithm that jointly regulates the end-to-end data rate, the video encoding rate, and the channel coding rate at the physical layer to minimize the distortion of the received video. The end-to-end data rate is chosen to avoid congestion while maintaining fairness in the domain of *video quality* (rather than data rate as in traditional rate control algorithms). Once the end-to-end data rate has been determined, the sender calculates the optimal proportion of video encoder rate and channel encoder rate based on the overall rate available and on the current quality of the wireless channel on the source-destination path, with the objective of minimizing the video distortion at the receiver. Cross-layer rate control algorithms designed to minimize the energy consumption while preserving video quality will be another important area of research.

3. Integration Of Video Sensing, Processing And Networking

Sensing and processing of multimedia content has mostly been approached as a problem isolated from the network design problem. Hence, research that addressed the content delivery aspects has typically not considered the characteristics of the source content. However, the sensing, processing and delivery of multimedia content are not independent, and their interaction has a major impact on performance. VSNs will support in-network processing algorithms operating on the sensed data. Hence, the QoS required at the application level will be delivered by means of a combination of cross-layer optimization and in-network processing of sensed data streams that describe the phenomenon of interest from multiple views, with different media, and using multiple resolutions. Thus, it is necessary to develop application-independent and self-organizing architectures to efficiently perform in-network processing of multimedia content. Examples of such interactions between sensing, processing and networking include compressed sensing paradigms as well as power-rate-distortion frameworks for resource allocation.

Video Encoding Based on Compressed Sensing. As an example of integration between sensing, processing and networking, video encoders based on the recently proposed compressive sensing (CS) paradigm [5] may offer a viable solution to the problems of encoder complexity and limited resiliency to channel errors that characterize predictive encoders. Compressed sensing is a new paradigm that allows the recovery of signals from far fewer measurements than methods based on Nyquist sampling. In particular, the main result of CS is that an N-dimensional signal can be reconstructed from M noise-like incoherent measurements as if one had observed the $M/log(N)$ most important coefficients in a suitable base [6]. Hence, CS can offer an alternative to traditional video encoders by enabling imaging systems that sense and compress data simultaneously *with low-complexity encoders*.

In [7], the performance and potential of CS-based video transmission in video sensor networks was evaluated. In CS, the transmitted samples constitute a random, incoherent combination of the original image pixels. This means that, unlike traditional wireless imaging systems, in CS no individual sample is more important for image reconstruction than any other sample. Instead, *the number of correctly received samples* is the only main factor in determining the quality of the received image. Hence, a peculiar characteristic of CS video is its *inherent and fine-grained spatial scalability*. The video quality can be regulated at a much finer granularity than traditional video encoders, by simply varying the number of samples per frame. Also, as shown in [7] a small amount of random channel errors does not affect the perceptual quality of the received image *at all*, since, for moderate BERs, the greater sparsity of the "correct" image will offset the error caused by the incorrect bit. CS image representation is completely *unstructured*: this fact *makes CS video* *more resilient than existing video coding schemes to random channel errors*. This simple fact has deep consequences on protocol design for end-to-end wireless transport of CS video (especially at the link and transport layers of the protocol stack), which are to be addressed in future research.

Power-Rate-Distortion Models. Another example of the need for tight integration of the networking with the sensing and processing is shown in the resource optimization frameworks that have recently been developed. Recognizing that VSNs are limited both in bandwidth and in energy, these frameworks extend the traditional rate-distortion (R-D) models for image processing to include power consumption. These power-rate-distortion (P-R-D) frameworks can be used to determine the optimal allocation of the constrained resources to the different components of a VSN node [8], [9]. Initial work on P-R-D frameworks for VSNs looked at the integrated optimization of power and rate for the compression and transmission modules [8]. More recent work includes the sensing in the optimization framework to ensure that all aspects of the VSN node are considered in resource allocation for a particular QoS (distortion) [9]. These works show the potential benefit of integrated optimization considering a simplified networking scenario. How to extend these frameworks considering multiple cameras' resources and medium access control issues is a challenging problem.

4. Conclusions

In conclusion, VSNs have unique challenges due to their high data rates and severe energy constraints coupled with the bandwidth-limited and time-varying nature of wireless networks. Cross-layer designs that integrate decisions not only of the network layers but also considering the sensing and processing are crucial to the efficient operation of future VSNs.

References

[1] I. F. Akyildiz, T. Melodia, and K. R. Chowdhury, "A Survey on Wireless Multimedia Sensor Networks," *Computer Networks (Elsevier)*, vol. 51, no. 4, pp. 921–960, Mar. 2007.

[2] S. Soro and W. Heinzelman, "A survey of visual sensor networks," *Advances in Multimedia*, 2009.

[3] C. Merlin, C.-H. Feng and W. Heinzelman, "Information-sharing Architectures for Sensor Networks: the State of the Art," in *ACM Mobile Computing and Communications Review (MC2R)*, Vol. 13, No. 4, Oct. 2009, pp. 26-38.

[4] S. Pudlewski and T. Melodia, "DMRC: Distortion-minimizing Rate Control for Wireless Multimedia Sensor Networks," in *Proc. of IEEE International Conference on Mobile Ad-hoc and Sensor Systems (MASS)*, Hong Kong S.A.R., P.R. China, October 2009.

[5] D. Donoho, "Compressed Sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[6] J. Romberg, "Imaging via Compressive Sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 14–20, 2008.

[7] S. Pudlewski and T. Melodia, "On the Performance of Compressive Video Streaming for Wireless Multimedia Sensor Networks," in *Proc. of IEEE Int Conf on Communications (ICC)*, Cape Town, South Africa, May 2010.

[8] Z. He and D. Wu, "Resource allocation and performance analysis of wireless video sensors," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, No. 5, pp. 590-599, May 2006.

[9] M. Marijan, W. Heinzelman, G. Sharma and Z. Ignjatovic, "Optimal Resource Allocation for Wireless Video Sensors with Power-Rate-Distortion Model of Imager," in *IEEE Midwest Symposium on Circuits and Systems (MWSCAS)*, 2009.

Tommaso Melodia [M2007] (tmelodia@buffalo.edu) is an Assistant Professor with the Department of Electrical Engineering at the University at Buffalo, The State University of New York (SUNY). He received his Ph.D. in Electrical and Computer Engineering from the Georgia Institute of Technology in 2007. He had previously received his Laurea (integrated B.S. and M.S.) and Doctorate degrees in Telecommunications Engineering from the University of Rome La Sapienza, Rome, Italy, in 2001 and 2005, respectively. He coauthored a paper that was recognized as the Fast Breaking Paper in the field of Computer Science for February 2009 by Thomson ISI Essential Science Indicators. He is an Associate Editor for the Computer Networks (Elsevier) Journal. He serves in the technical program committees of several leading conferences in wireless communications and networking, including IEEE Infocom, ACM Mobicom, and ACM Mobihoc. He was the technical co-chair of the Ad Hoc and Sensor Networks Symposium for IEEE ICC 2009. His current research interests are in modeling and optimization of multi-hop wireless networks, cross-layer design and optimization, cognitive radio networks, multimedia sensor networks, and underwater acoustic networks.

Wendi Heinzelman is an associate professor in the Department of Electrical and Computer Engineering at the University of Rochester. She hold a secondary appointment in the Computer Science Department at Rochester. Dr. Heinzelman also currently serves as Dean of Graduate Studies for Arts, Sciences and Engineering at the University of Rochester. Dr. Heinzelman received a B.S. degree in Electrical Engineering from Cornell University in 1995 and M.S. and Ph.D. degrees in Electrical Engineering and Computer Science from MIT in 1997 and 2000, respectively. Her current research interests lie in the areas of wireless communications and networking, mobile computing, and multimedia communication. Dr. Heinzelman received the NSF CAREER award in 2005 for her research on cross-layer architectures for wireless sensor networks, and she received the ONR Young Investigator Award in 2005 for her work on balancing resource utilization in wireless sensor networks. She is an Associate Editor for the IEEE Transactions on Mobile Computing, an Associate Editor for the ACM Transactions on Sensor Networks, and an Associate Editor for Elsevier Ad Hoc Networks Journal. Dr. Heinzelman is a senior member of the IEEE and the ACM, and she is co-founder of the N^2 Women (Networking Networking Women) group.