

# Real-time Multimedia Monitoring in Large-Scale Wireless Multimedia Sensor Networks: Research Challenges

Matteo Cesana, Alessandro Redondi  
 Politecnico di Milano,  
 20129 Milano, Italy  
 Email: {cesana, redondi}@elet.polimi.it

Nestor Tiglao\*<sup>†</sup>, António Grilo\*  
 \*INESC-ID/INOV/IST, 1000-029 Lisboa, Portugal  
<sup>†</sup>EEEI, University of the Philippines  
 Email: {nestor.tiglao, antonio.grilo}@inesc-id.pt

Jose M. Barcelo-Ordinas, Mohammad Alaei  
 Universitat Politècnica de Catalunya,  
 08034 Barcelona, Spain  
 Email: {joseb,malaei}@ac.upc.edu

Petia Todorova  
 Fraunhofer FOKUS,  
 10589 Berlin, Germany  
 Email: petia.todorova@fokus.fraunhofer.de

**Abstract**—Wireless Sensor Networks (WSNs) have enjoyed dramatic developments over the last decade. The availability of CMOS cameras and microphones enlarged the scope of WSNs paving the way to the development of Wireless Multimedia Sensor Networks (WMSN). Among the envisaged WMSN applications, Real-time Multimedia Monitoring constitutes one of the most promising. However, the resource requirements of these applications place difficult challenges in terms of network lifetime and scalability. This paper starts by identifying the main characteristics and requirements of Real-time Multimedia Monitoring applications and then highlights key research directions that may help to overcome those challenges.

## I. INTRODUCTION

Over the last decade, Wireless Sensor Networks (WSNs) have attracted increasing interest within research community and manufacturers. WSNs are composed of small sized battery operated network devices geared with processing capabilities, wireless communication interfaces, and sensing functionalities. With the diminishing cost of communication devices, WSNs have emerged as ideal solutions to a large number of applications in both civilian and military contexts. With the increased availability of CMOS cameras and microphones, as well as more powerful and energy-efficient WSN nodes, the latter can now be equipped with multimedia sensor suites to collaboratively monitor a given area, leading to the concept of Wireless Multimedia Sensor Networks (WMSN). A general survey on WMSNs can be found in [1]. In this paper, we focus on Real-time Multimedia Monitoring applications. Being particularly demanding in terms of computational and energy resources, this class of applications raises interesting and difficult research challenges, demanding innovative solutions for the combination of the optimization techniques at the different layers of the protocol stack. This paper presents the main research challenges on promising technologies, algorithms and protocols that can be used to support Real-time Multimedia Monitoring. The rest of this document is structured as follows.

Section II presents the main characteristics and requirements of Real-time Multimedia Monitoring applications. Section III addresses the suitable communication protocols and techniques. Finally, Section V presents some conclusions, pointing the way to future research on the topic.

## II. REAL-TIME MULTIMEDIA MONITORING APPLICATIONS

Real-time Monitoring in visual camera networks refers to an incredibly broad area of tasks. Face and object recognition or classification, vehicle tracking and crowd monitoring are just few examples of typical application scenarios. Monitoring can be done in a supervised or unsupervised manner. In the former case, still images or video are sent from the visual sensors to a central monitoring place, where a human operator is constantly required to watch the multimedia streams and reacts properly if an anomalous event is detected. Conversely, in unsupervised monitoring, the multimedia content is automatically analyzed by means of specialized computer vision algorithms to infer the semantic content. This latter option is clearly very attractive and promising, and system based on that have emerged in the past few years under the name of "smart camera systems". Depending on the specific requirements of the reference monitoring scenario, two broad classes of real-time monitoring application can be defined: image/video-based monitoring and feature-based monitoring. Applications belonging to the former class require the remote delivery of image/video flows which are then analyzed (either by human operators or by automatic systems) in search for anomalous behaviors and/or hazardous events. In the latter class of applications, only a succinct representation of an image is delivered remotely to implement specific detection functionalities (face detection, object recognition, etc.). In the following, the distinctive characteristics of both application scenarios are commented, clearly highlighting the main reference application fields and the consequent system's level

requirements.

### A. Images and Video

Supervised monitoring requires still images or video to be transmitted from the visual sensor to a central operation place, where a human operator controls the visual stream. Clearly, such multimedia stream must be compressed to reduce bandwidth, energy, delay and memory requirements.

Digital images and video coding has been subject to scientific research and serious standardization in the last 20 years. Currently, the worldwide accepted standard for video compression is H.264/AVC ([2]), which provides very good rate-distortion performance due to the use of advanced techniques, such as spatial prediction for intra coding, multiple reference frames motion compensation, variable block-size intra/inter block coding and advanced entropy coding schemes. The gain in performance on the coding side is balanced by a tremendous complexity in the encoder design, which is still too heavy to be implemented and used on current sensor network platforms.

In order to find more suitable solutions for WMSN, an object-based approach can be used ([3], [4]). The rationale behind these encoders is to constraint motion estimation and encoding only to those part of the frame which involves salient motion, in order to reduce the computational complexity. Usually, background subtraction techniques are employed in order to distinguish the background scene from the foreground objects. Recently, a different encoding paradigm known as distributed video coding (DVC [5]) has been proposed as a candidate technique for video compression in WMSN. In DVC, the computational complexity is shifted from the encoder to the decoder, hence it seems promising for applications built on top of visual sensors. Unfortunately, successful implementations of such technique are still not present. The integration of all the aforementioned video compression algorithms within a visual wireless sensor network has been subject of study in the past few years. The surveys of Misra et al [6], Soro and Heinzelman [7] and Akyldiz et al. [1] summarize the work done towards this integration.

### B. Visual Features

Automatic monitoring is accomplished with the aid of some synthetic description of the underlying pixel content. Instead of processing the image or the video directly on their spatial representation, a set of descriptors or features is extracted by means of specialized algorithms.

Even if there is no standard or general strategy for monitoring based on visual features, the use of local photometric descriptors such as SIFT [8] and SURF [9] seems very promising. Such local image features have in fact demonstrated excellent performance in a broad range of monitoring tasks: they are invariant and show robustness with respect to scale, rotation, affine transformations and illumination, hence they form the basis for recognition based applications. Examples include, but are not limited to, people/object identification [10][11], environmental and habitat monitoring [12][13], traf-

fic and parking monitoring [14] [15] [16] and face recognition [17][18].

Those analysis tasks are usually based on a widely accepted processing paradigm, that we refer to as compress-then-analyze. In a nutshell, still images or video are acquired by cameras deployed in the monitored area, compressed with some encoding technique (see II-A) and then sent to a central controller that analyze the visual content (i.e. compute and process SIFT features) leveraging superior computational resources. However, applying such compress-then-analyze paradigm to visual sensor networks is particularly challenging due to the resources available on camera nodes. Limited memory, CPU, energy and bandwidth impose hard constraints, resulting in low spatial and temporal resolutions of the delivered content, that may impair further feature extraction and analysis.

A different approach to visual analysis can be used by reversing the traditional compress-then-analyze paradigm. In this new approach, camera sensors use their computational resource to extract and compress local image features directly, instead of compressing the visual sequence. These features may then be transmitted to a central controller for following steps of the analysis process. Clearly, this approach is advantageous if computing, compressing and transmitting local features costs less than doing it for images and video, given that the accuracy of the final task remains the same. The final cost can be expressed as the total energy required by a camera sensors to compute, encode and transmit features instead of video. For what concerns the computation of local features, some work has been done recently to reduce SIFT complexity without affecting performance (see [9][19]). However, practical implementations of feature extraction schemes on visual sensors has not been studied in detail yet: a working system has been presented in [20], where the features detection / extraction algorithm is based on SURF and takes about 10/20 seconds to be completed.

1) *Coding techniques:* In the last few years an increasing number of works is facing the problem of compressing local features. We can spot two different approaches to feature compression: (i) improving the performance of existing descriptor algorithms and (ii) designing descriptors algorithms that produces low-bitrate visual features. Several examples of the first approach can be found in the recent literature. In [21] the authors applied Principal Component Analysis to the normalized SIFT gradient patch, yielding to a more compact representation that is also more distinctive and robust to image deformations. In [22] Chandrasekhar et al. propose a general framework for transform coding image features using a Karhunen-Loeve Transform learned from a large training set. Recently, a new class of low-bitrate descriptors was developed in [23]: here in contrast to prior work, the gradient statistics around interest points are explicitly exploited to build an histogram that can be efficiently compressed. Such CHoG (Compressed Histogram of Gradients) descriptors are shown to be particularly attractive, since they provide a compact, yet comprehensive, representation of the local image patches. As

an example, the performance of a 53 bits CHoG descriptor matches the one of a 1024 bit SIFT descriptor. In [24] the same authors also proposed a low-latency image retrieval system with progressive transmission of CHoG descriptors: a feedback is used when a match is found on the server so that the querying client can stop the transmission process. In that case, descriptors are sorted by their Hessian blob response, with the assumption that descriptors with low Hessian response are poorly localized and less discriminative. Another approach to descriptors sorting is given in [25]: descriptors are ordered in such a way to minimize their inter-distance, then only the difference between a descriptor and its following is coded and transmitted, in order to provide additional rate savings. Finally, an hybrid approach is presented in [26]: here, patches of pixels around the detected keypoints are JPEG compressed and sent to a central controller where they are further processed to compute the relative descriptor. This approach allows to obtain very low bit rate, while unchanging the accuracy of the following analysis task. Table I summarize different features compression techniques, together with their target bitrate.

TABLE I  
FEATURE COMPRESSION TECHNIQUES.

Method	Bitrate (bits / descriptor)
SIFT [8][26]	128 - 1024
SURF [9][22]	64 - 512
CHOG [23][24]	53 - 100
PCA-SIFT [21]	160
KLT-SIFT [22]	256
SORTING SIFT [25]	10 - 40
COMPRESSED PATCHES [26]	38 - 57

2) *How many features?:* So far, we have reviewed different techniques to encode local features. However, a natural question that can be raised is the following: how many feature descriptors are needed to perform visual analysis with a specific target accuracy? In other words, what is the total amount of data that is generated when e.g. object recognition must be performed on an input image? Clearly, there's no general answer to this question. First, the number of descriptors that can be extracted from an image is dependent on both image-specific and algorithm-specific parameters. Image-specific parameters are e.g. the image resolution (the higher the resolution, the higher the number of features extracted) and the visual content (images rich of details produce more features). Algorithm-specific parameters depends on the particular detector algorithm used (i.e. corner/blob detector) as well as on particular parameters of the detector algorithm. As an example, in SIFT and SURF it is possible to tune a threshold on the Hessian response value of each detected keypoint: keypoints with low hessian values are discarded and their correspondent feature is not computed. Hence, varying this threshold has impact on the number of features that are finally extracted. Another factor that has impact on the number of features to be extracted to obtain a target accuracy is the complexity of the visual task, that can be ultimately seen as the dimension of the database against which the analysis is performed. In a nutshell, recognizing an object within a small set of other objects

obviously requires less features than recognizing the same object in a big set of objects with similar visual characteristics. Examples of works that study how recognition accuracy varies when varying the number of feature descriptors or the database size can be found in [8][16] and [27]. In [28] the authors go a step further, evaluating the impact of both the number of feature descriptors and their quantization rate on the final accuracy. A rate-accuracy model is defined and used to optimally allocate resources in a WMSN, with the aim of maximizing its lifetime.

### III. RELIABLE AND EFFICIENT DELIVERY OF VISUAL DATA IN WMSNs

This section presents relevant techniques for the support of Real-time Multimedia Monitoring applications at the protocol stack level at the MAC layer and above.

#### A. MAC Layer

The MAC Layer is responsible for providing medium access control to competing nodes guaranteeing QoS metrics such as delay, losses, distortion, fairness, etc. In order to achieve these QoS metrics, a wireless MAC protocol has to be able to, [29], minimize medium access delay, minimize collisions in case of having contention-based access, maximize reliability, minimize energy consumption, minimize interference, and maximize adaptivity to changes. Video streaming applications require a steady flow of information and delay-bounded delivery of packets. Contention-based MAC has difficulties in providing these requirements due to the collision nature of the medium, the interference from other nodes and the multi-path fading and shadowing effects over the link. Providing video over wireless sensor networks requires the need of joint reliable transport and resource allocation while saving as much energy as possible, which demands a cross-layer architecture that integrates all these features. As a conclusion, a WMSN MAC layer, besides providing medium sharing, has to be able to support reliable communications, save energy and be QoS-aware. Service differentiation is a technique that includes a wide range of mechanisms supporting QoS in networks. The available differentiation mechanisms at MAC layer that allow tuning QoS metrics in WMSN are:

- 1) Power control [30]: consists of adjusting the transmission power in order to assure correct transmission of packets, minimizing the interference with other nodes.
- 2) Traffic Class differentiation: many QoS MAC proposals differentiate traffic based on local decisions. For example, Q-MAC, [31], classifies urgent traffic based on a mixture of parameters such as transmission hops, amount of residual energy and queue proportional load. Other proposals, [32] [33] pre-assign a set of priorities to different kind of services. In general, multimedia traffic has to be identified with respect to other applications according to their QoS requirements, i.e. service classification. Moreover, finer tuning based on local decisions such as residual energy or queueing load can be applied in order to define forwarding or dropping strategies

Then, we define traffic classification as how the MAC treats each traffic: pre-assigned traffic classes based on QoS requirements or local classification based on local decisions (e.g. number of hops towards destination).

- 3) Control the access to the medium [29]: Contention-free protocols control the time-slot assignment. In this way, high priority traffic can reserve higher number of slots with respect other lower priority sensor applications. However, contention-free networks need tight synchronization and complexity increases if the sensor network scales, i.e., when new nodes are added to the network. On the other hand, in contention-based protocols, there are several ways to control the access and frequency of transmissions. The contention window (CW) size control allows senders to differentiate traffic by selecting different timers. The back-off exponent control allows alleviating contention by reducing the probability of collisions. Inter-frame space control also allows traffic differentiation since traffic classes will access the medium with different precedence.
- 4) Duty-cycling control, [34]: topology control allows power management by controlling active/sleeping periods or designing low-duty MAC protocols. Sensor nodes processing video streaming can have higher active periods than those sensors handling less priority traffic. The effect of this technique will be lower delays and higher throughputs on high priority traffic in comparison with best effort traffic.
- 5) Queuing and scheduling mechanisms: several queues can be set up for the different traffic classes. Then, a scheduler can give more service to those high priority traffic classes. Furthermore, mechanisms like Weighted Fair Queuing (WFQ) can guarantee minimum throughput - bandwidth control - in per queuing classes and inside a priority class.
- 6) Error control mechanisms: MAC protocols provide several error control mechanisms to provide resilience against bit errors and losses. ARQ mechanisms provide hard QoS at the cost of increasing latency and reducing bandwidth. FEC mechanisms include higher redundancy allowing recovering of data against errors at the cost of memory and processing and higher transmission latency due to the longer packets.

Among these mechanisms there are three of the utmost importance when designing a MAC protocol for video streaming in WMSN using contention-based MAC protocols. A first requisite is that nodes have to implement intra-node traffic class differentiation. That means a queuing management and scheduling mechanism that separate traffic according to their classes and receive service according to their priority. Static approaches assign a priority to each traffic class. However, this priority does not change over time or depending on the load conditions. A dynamic priority assignment reacts to the load conditions of the network. Typical decision parameters are the number of hops towards the sink, the number of

hops already traversed, the amount of the remaining delay budget, the remaining energy, etc. Furthermore, it is possible to separate traffic classes in different queues that are served according to their priority. Separate queues can be scheduled with a mechanism like WFQ that allows guaranteed bandwidth services between the different traffic classes. Each queue can also be controlled with a scheduler to guarantee fairness between the same traffic class flows. The objective of these techniques is to achieve bounded delays and guarantee certain minimum bandwidth per traffic class and per traffic flow.

A second requisite for video streaming is achieving inter-node traffic class differentiation. A packet scheduled by a node has to compete for the medium with other nodes, without knowing whether its priority is higher or lower than the packets scheduled by its neighbors. Contention Window (CW) size control allows senders to differentiate traffic assigning shorter CW to high priority traffic and larger CW to low priority traffic. In this way, high priority traffic has more probability to access the medium than low priority traffic. However, careful tuning is necessary since lower traffic classes can considerably increase the latency in case of highly loaded networks. Back-off exponent control and IFS control also play a role in the system and is challenging finding a trade-off on the parameters for each traffic class in order to bound delays. Furthermore, these parameters also impact bandwidth since they control retransmissions due to packet collision. Finally, a dynamic Contention Window size control mechanism that adapts to load conditions can significantly improve delays among traffic classes at the cost of adding complexity.

A third requisite for video streaming is parameterizing when a node has to wake up. Node synchronization is challenging since video streaming is bursty by nature and static duty-cycle protocols, e.g. STEM [35], do not behave well with highly variable traffic conditions. Predicting the amount of video traffic as well as duty-cycle adjustment are necessary conditions in order to achieve low delay budgets and to avoid idle listening or early sleeping. Video traffic characterization depends on the used video coding techniques. Predicting the amount of traffic that a sensor node is going to receive is a challenging task. Furthermore, in case of using multipath routing, the task can be ever more challenging. Finally, dynamic active periods imply that nodes have to buffer a minimum number of packets before retransmission. This implies a trade-off on the length of the active period with respect to buffer size and energy consumption.

Few works in Wireless Sensor Networks at MAC layer are aimed for video streaming. Some MAC protocols are designed with traffic classes in mind, but without being specific to video streaming. Most of these works control the access to the medium, differentiating between traffic classes. However, since they do not have video streaming as their main application, it is not clear whether these MAC protocols adapt well to high volumes of bursty data traffic, i.e. large video frames that have to be fragmented in multiple back-to-back lower layer packets. Table II compares several MAC protocols in terms of the implemented features.

TABLE II  
SUMMARY OF DESIRABLE PROTOCOL FEATURES FOR SOME CONTENTION-BASED MACs.

Protocol	Multimedia traffic	Traffic Class differentiation	Controlling the access to the medium	Dynamic Duty-cycle control	Queue management (Priority queues)	Bandwidth Control (WFQ)
PSIFT [36]	No	Yes (local-based)	CW size/IFS	No	No	No
PR-MAC [32]	No	Yes (pre-assigned)	CW size/IFS	No	No	No
RL-MAC [33]	No	Yes (pre-assigned)	CW size	Yes	Yes	No
Q-MAC [31]	No	Yes (local-based)	No	No	Yes	Yes
PQ-MAC [37]	No	Yes (pre-assigned)	IFS	Yes	Yes	No
Sexena MAC [38]	Yes	Yes (pre-assigned)	CW size	Yes	Yes	No
Diff-MAC [39]	Yes	Yes (pre-assigned)	CW size	Yes	Yes	Yes
CoSenS [40]	No	Yes (pre-assigned)	Back-off	Yes	Yes	No

### B. Network Layer

In traditional WSNs, the Network Layer is mainly responsible for routing traffic from the source (usually sensor nodes) to the destination (usually sink node or actuator nodes) in an energy-efficient way. QoS metrics such as delay, jitter and throughput are relegated to a background role in comparison with energy consumption and network lifetime. In contrast, WMSN applications in general and Real-time Multimedia Monitoring in particular, require real-time performance as well as high data rates. In this case, traditional WSN routing protocols such as Directed Diffusion [41] become insufficient. Energy-efficiency is still of the utmost importance in WMSNs, and especially so because multimedia applications are typically more demanding in terms of energy. However, new routing protocols are needed which are able to find the best trade-off between the often contradictory requirements of energy-efficient and real-time transmission. The following are desirable features of routing protocols for monitoring and tracking WMSNs:

- 1) Traffic differentiation and joint optimization of multiple QoS goals to deliver multimedia traffic: Among the QoS parameters are delay bound, loss rate, and throughput. For example, delay-sensitive traffic should be granted a shorter and faster path, while more delay-tolerant bulky traffic might be sent through an energy-rich but longer path. A special case of traffic differentiation, concerns differentiation between packets belonging to the same stream. For example in MPEG video transmission, I, P and B frames should be differentiated regarding reliability. This requires a deeper awareness about the structure or contents of multimedia data.
- 2) Resource balancing (mainly energy, load, memory and processing power): The objective is to avoid routing bottlenecks and thus to maximize network lifetime and goodput, while minimizing delay. It includes approaches such as multipath routing.
- 3) Fast adaptation to changes in the location of monitored objects: Routes must be computed and established quickly in response to changing networks conditions and real-time delay bounds must be satisfied.
- 4) Support to in-network processing and cross-layer optimization: The goal is to provide closer interaction and

cooperation between the layers of the network protocol stack in order to improve overall performance. When in-network processing is enabled, selected routes should cross more powerful multimedia processing nodes (e.g. more powerful CPU, energy, memory, etc.).

- 5) Scalability: Large-scale monitoring with thousands of nodes is only possible if the routing overhead is kept to a minimum. This can be achieved by means of approaches such as hierarchical routing, opportunistic routing, clustering and geographic forwarding.
- 6) Energy harvesting awareness: Since WMSNs typically consume a significantly larger amount of energy in comparison with scalar WSNs, the network lifetime is too short if energy harvesting mechanisms are not implemented. Routing optimization is different depending on availability of energy harvesting, since the main goal moves away from maximizing the network lifetime. In case the WMSN nodes are able to gather energy from external sources, the task of the routing protocol is to find sustainable flows, maximizing the throughput and minimizing the delay.

Table III lists several state-of-the-art network layer solutions together with the respective features.

### C. Transport Layer

Traditional transport protocols like TCP are not suitable for WMSNs since they were mainly designed to operate in wired networks. In addition to reliability, fairness, and congestion control, WMSNs also impose additional requirements such as energy-efficiency and tight delay bounds. Here we discuss the state-of-the-art WSN transport protocols. The experimental study in [53] shows that existing protocols provide very poor performance for real-time multimedia delivery. However, we are slowly seeing the development of transport protocols that are able to improve the delivery of multimedia traffic. Based on state-of-the-art approaches, the following features are useful in the design and development of transport protocols for Real-time Multimedia Monitoring WMSNs:

- 1) Differentiated reliability: Since some multimedia applications can be loss-tolerant, differentiated reliability should be supported.
- 2) Reliability-delay trade-off: Reliability and delay should be jointly addressed and adapted based on the unique re-

TABLE III  
SUMMARY OF DESIRABLE ROUTING PROTOCOL FEATURES FOR WMSNs.

Protocol	Multimedia QoS	Tracking & Monitoring	Energy-aware	Multi-path	Scalability	Cross-layer interaction	In-network processing
Multimedia-aware MMSPEED [42]	Yes	No	No	Yes	Geo. Forwarding	Network, MAC	No
3R [43]	Yes	No	No	Yes	No	Network, MAC	No
Mahapatra, et. al [44]	Yes	No	Yes	No	Geo. Forwarding	Network, MAC	No
Akkaya & Younis [45]	Yes	No	Yes	No	No	No	No
Mohajerzadeh & Yaghmaee [46]	Yes	No	Yes	No	No	No	No
RAP [47]	Yes	No	No	No	Geo. Forwarding	Network, MAC	No
DCAR [48], MIXIT [49]	No	No	No	Yes	No	Network, PHY	Yes
MORE [50]	No	No	No	Yes	Opp. Routing	Network, PHY	Yes
APOLLO/PISA [51]	Yes	No	Yes (e.-harvesting)	No	Geo. Forwarding	Network, MAC	No
D-APOLLO [52]	Yes	No	Yes (e.-harvesting)	No	Geo. Forwarding	Network, MAC	No

quirements of the multimedia application, since usually one is achieved at the expense of the other. Network coding is a promising approach to avoid the retransmission delays. It can be combined with ARQ. Some approaches based on TCP combine these mechanisms.

- 3) Media-centric collective reliability: When there is redundancy in multimedia data, reliability must not be regarded from a packet-based perspective, treating individual media flows independently. Instead, the objective of reliability is to provide an usable fused media stream from the media streams sent by individual sources. This feature relies heavily on cross-layer optimization.
- 4) Congestion control: Congestion avoidance and rate control mechanism must be designed considering the stringent and unique demands of multimedia traffic.
- 5) Cross-layer optimization: Transport protocols must be adapted dynamically in close coordination with other layers, e.g. application, MAC and network layers. Cross-layer interaction with the in-network processing mechanisms at application-layer as well as with the supporting routing mechanisms is the only way to maximize energy-efficiency.

A summary of the main features of those transport protocols is listed in Table IV. A more comprehensive survey of WMSN transport layer protocols can be found in [6][7][1][54].

#### D. Collaborative sensing in WMSN

Sensing management and scheduling policies define how nodes sense the environment. Multimedia nodes are characterized by a directional sensing model defined by the Field of View (FoV). Objects covered by a camera can be distant and the captured images will depend on the relative positions and orientations of the cameras towards the observed object. In planned deployments, nodes usually cover the whole area with low redundancy. The objective is that each camera covers a unique set of points in such a way that the amount of area covered is maximized with the minimum number of cameras, e.g., art gallery problem. In random deployments, however, nodes are deployed in such a way that multiple cameras cover the same area, implying redundant sensing areas. Collaborative sensing consists of group nodes according to their common sensing coverage and schedules them in order to wake them up to fulfill the application tasks in such

a manner that the amount of data to be sent to the sink is minimized. The FoV is one of the main candidates for grouping nodes. If several nodes have FoV areas that intersect, they can be grouped and collaborate in sensing the area. The main trade-off is that at each wake up period part of the FoV is common with respect to clustered nodes, but part of the FoV is missed. However, in general, if the duty-cycle of a clustered node is less than the time a mobile object spends in crossing a FoV, the object will be detected by one of the nodes of the cluster. Alaei et al in study [61][62][63] several clustering mechanism for object detection and object monitoring and propose scheduling mechanisms showing how collaborative sensing reduces the amount of energy a node spends in comparison with scenarios without collaboration.

#### IV. DISCUSSION AND RESEARCH CHALLENGES

Visual analysis based on local features can be integrated with wireless multimedia sensor networks to enable automatic monitoring in a broad range of tasks. However, due to the limitations introduced by current visual sensor architectures, it is imperative to optimize every stage in the analysis process. First of all, practical implementations of state-of-the-art local features extraction algorithms on top of currently available visual sensor nodes must be investigated and optimized in order to enable energy-efficient real-time feature extraction. Secondly, advanced techniques to encode visual features are needed, in order to minimize the amount of data that requires to be transmitted to ensure high accuracy.

Real-time Multimedia Monitoring applications requires the integration of appropriate mechanisms at the different layers of the protocol stack. The protocol architectures proposed in the literature are usually focused on the optimization of specific layers, overlooking cross-layer optimization to a significant extent. Supporting video streaming applications at MAC layer imply the integration of service differentiation techniques. The MAC has to design traffic class differentiation and queuing and scheduling mechanisms. However, the most difficult and challenging task is achieving dynamic duty-cycling control due to the high variability of the video traffic. Furthermore, the amount of high priority traffic that arrives to a node also depends on the neighbourhood and is tightly coupled with the higher layers. At the network layer, no single existing proposal integrates all the desirable

TABLE IV  
SUMMARY OF DESIRABLE TRANSPORT PROTOCOL FEATURES FOR WMSNs.

Protocol	Reliability vs. delay trade-off	Differentiated reliability	Media-centric reliability	Congestion control	Cross-layer optimization
ESRT [55]	No	No	No	Yes	No
STCP [56]	Yes	Yes	No	Yes	No
RCRT [57]	No	Yes	No	Yes	Yes
TRCCIT [58]	Yes	Yes	No	Yes	Yes
MDTSN [59]	Yes	Yes	No	No	No
DMRC [60]	Yes	Yes	No	No	Yes

characteristics required by Real-time Multimedia Monitoring applications. In particular awareness of in-network processing hubs and energy harvesting are still not fully exploited in WMSN routing protocols. Media-centric collective reliability has been mostly neglected in transport layer proposals, requiring a better exploitation of transport-application and transport-routing cross-layer optimization. When an ARQ-based transport protocol relies on caching, transport-routing optimization becomes a must, especially in heterogeneous WSNs where nodes have different storage capabilities. Two radio neighbor sensors may cover different FoV while two sensors that cover the same FoV are not necessarily radio neighbours. How nodes that sense the same data can collaborate to reduce the amount of traffic generated is a challenge. Thus, scheduling the best node to detect, monitor or track an object will increase network lifetime and will also improve related QoS parameters such as delay or distortion.

#### V. CONCLUSIONS

Real-time Multimedia Monitoring constitutes a particularly resource-hungry class of WMSN applications, since it combines contradictory goals in terms of QoS, network lifetime and scalability. This paper has presented the key characteristics of Real-time Multimedia Monitoring in WMSNs, identifying promising directions to overcome the associated challenges. Like in typical WSN applications, the feasibility of Real-time Multimedia Monitoring depends on suitable network architecture design, as well as the coordination of optimization processes at different levels. From the analysis of the related work on WMSNs, it can be concluded that there still exist areas that are not fully exploited, which are probably essential to enable Real-time Multimedia Monitoring. Collaborative sensing is essential to minimize the traffic redundancy and thus to increase the energy efficiency and network lifetime of WMSNs. Finally, application-centric cross-layer optimization is essential to achieve the required QoS in the most energy-efficient way, namely the integration of energy-harvesting techniques and the provisioning of media-centric collective reliability.

#### ACKNOWLEDGMENT

This work was supported by the Euro-NF Network of Excellence (Grant Agreement No. 216366). Nestor Tigliao is supported by a UP ERDT PhD Scholarship Grant. INESC-ID is also supported by FCT (PIDDAC funds). UPC is supported by TIN2010-21378-C02-01 and SGR2009-1167.

#### REFERENCES

- [1] I. F. Akyildiz, T. Melodia, and K. R. Chowdhury, "A survey on wireless multimedia sensor networks," *Computer Networks*, vol. 51, no. 4, pp. 921–960, 2007.
- [2] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the h.264/avc video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, July 2003.
- [3] T. Tsai and C.-Y. Lin, "Exploring contextual redundancy in improving object-based video coding for video sensor networks surveillance," *IEEE Transactions on Multimedia*, vol. PP, no. 99, p. 1, 2011.
- [4] Y. Yu and D. Doermann, "Model of object-based coding for surveillance video," in *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on*, vol. 2, pp. 18–23, 2005.
- [5] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proceedings of the IEEE*, vol. 93, no. 1, Jan. 2005.
- [6] S. Misra, M. Reisslein, and G. Xue, "A survey of multimedia streaming in wireless sensor networks," *IEEE Communications Surveys and Tutorials*, vol. 10, no. 1-4, pp. 18–39, 2008.
- [7] S. Soro and W. R. Heinzelman, "A survey of visual sensor networks," *Advances in Multimedia*, vol. 2009, 2009.
- [8] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, 2004.
- [9] H. Bay, T. Tuytelaars, and L. J. V. Gool, "Surf: Speeded up robust features," in *ECCV (1)*, 2006, pp. 404–417.
- [10] I. O. de Oliveira and J. L. S. Pio, "People reidentification in a camera network," in *Computer Science and its Applications, 2009. CSA '09. 2nd International Conference on*, Dec. 2009, pp. 1–8.
- [11] D. Figueira and A. Bernardino, "Re-identification of visual targets in camera networks: A comparison of techniques," in *ICIAR (1)*, ser. Lecture Notes in Computer Science, M. Kamel and A. C. Campilho, Eds., vol. 6753. Springer, 2011, pp. 294–303.
- [12] T. Ko, S. Ahmadian, J. Hicks, M. H. Rahimi, D. Estrin, S. Soatto, S. Coe, and M. P. Hamilton, "Heartbeat of a nest: Using imagers as biological sensors," *TOSN*, vol. 6, no. 3, 2010.
- [13] O. Pizarro, P. Rigby, M. Johnson-Roberson, S. Williams, and J. Colquhoun, "Towards image-based marine habitat classification," in *OCEANS 2008*, Sept. 2008, pp. 1–7.
- [14] H. Bhaskar, N. Werghi, and S. Al-Mansoori, "Rectangular empty parking space detection using sift based classification," in *VISAPP*, L. Mestetskiy and J. Braz, Eds. SciTePress, pp. 214–220.
- [15] J.-Y. Choi, K.-S. Sung, and Y.-K. Yang, "Multiple vehicles detection and tracking based on scale-invariant feature transform," in *Intelligent Transportation Systems Conference, 2007. ITSC 2007. IEEE*, 30 Oct. 2007, pp. 528–533.
- [16] Z. Zhang, M. Li, K. Huang, and T. Tan, "Boosting local feature descriptors for automatic objects classification in traffic scene surveillance," in *ICPR*. IEEE, pp. 1–4.
- [17] J. Luo, Y. Ma, E. Takikawa, S. Lao, M. Kawade, and B.-L. Lu, "Person-specific sift features for face recognition," in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, vol. 2, April 2007, pp. II–593–II–596.
- [18] M. Bicego, A. Lagorio, E. Grosso, and M. Tistarelli, "On the use of sift features for face authentication," in *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW '06. Conference on*, June 2006.
- [19] M. Grabner, H. Grabner, and H. Bischof, "Fast approximated sift," in *ACCV (1)*, 2006, pp. 918–927.
- [20] A. Yang, S. Maji, C. Christodias, T. Darrell, J. Malik, and S. Sastry, "Multiple-view object recognition in band-limited distributed camera networks," in *Distributed Smart Cameras, 2009. ICDSC 2009. Third ACM/IEEE International Conference on*, 30 Oct. 2009–Sept. 2 2009, pp. 1–8.

- [21] Y. Ke and R. Sukthankar, "Pca-sift: A more distinctive representation for local image descriptors," in *CVPR (2)*, 2004, pp. 506–513.
- [22] V. Chandrasekhar, G. Takacs, D. Chen, S. S. Tsai, J. Singh, and B. Girod, "Transform coding of image feature descriptors," M. Rabbani and R. L. Stevenson, Eds., no. 1.
- [23] V. Chandrasekhar, G. Takacs, D. M. Chen, S. S. Tsai, R. Grzeszczuk, and B. Girod, "Chog: Compressed histogram of gradients a low bit-rate feature descriptor," in *CVPR*, 2009, pp. 2504–2511.
- [24] V. R. Chandrasekhar, S. S. Tsai, G. Takacs, D. M. Chen, N.-M. Cheung, Y. Reznik, R. Vedantham, R. Grzeszczuk, and B. Girod, "Low latency image retrieval with progressive transmission of chog descriptors," in *Proceedings of the 2010 ACM multimedia workshop on Mobile cloud media computing*, ser. MCMC '10, 2010, pp. 41–46.
- [25] R. J. H. Y. Chen, L. Y. Duan and W. Gao, "Sorting local descriptors for low bit rate mobile visual search," *IEEE International Conference on Acoustic, Speech, and Signal Processing*, 2011.
- [26] M. Makar, C.-L. Chang, D. M. Chen, S. S. Tsai, and B. Girod, "Compression of image patches for local feature extraction," in *ICASSP*, 2009, pp. 821–824.
- [27] J. J. Foo and R. Sinha, "Pruning sift for scalable near-duplicate image matching," in *ADC*, 2007, pp. 63–71.
- [28] A. Redondi, M. Cesana, and M. Tagliasacchi, "Rate-accuracy optimization in visual wireless sensor networks," in *Image Processing, 2012. Proceedings. (ICIP2012). Submitted to IEEE International Conference on*, 2012.
- [29] M. Yigitel, O. Incel, and C. Ersoy, "Qos-aware mac protocols for wireless sensor networks: A survey," *Computer Networks*, 2011.
- [30] N. Pantazis and D. Vergados, "A survey on power control issues in wireless sensor networks," *IEEE Communications Surveys & Tutorials*, vol. 9, no. 4, pp. 86–107, 2007.
- [31] Y. Liu, I. Elhanany, and H. Qi, "An energy-efficient qos-aware media access control protocol for wireless sensor networks," in *Mobile Adhoc and Sensor Systems Conference, 2005. IEEE International Conference on*. IEEE, 2005, pp. 3–pp.
- [32] A. Firoze, L. Ju, and L. Kwong, "Pr-mac a priority reservation mac protocol for wireless sensor networks," in *Electrical Engineering, 2007. ICEE'07. International Conference on*. IEEE, 2007, pp. 1–6.
- [33] Z. Liu and I. Elhanany, "Rl-mac: a qos-aware reinforcement learning based mac protocol for wireless sensor networks," in *Networking, Sensing and Control, 2006. ICNSC'06. Proceedings of the 2006 IEEE International Conference on*. IEEE, 2006, pp. 768–773.
- [34] G. Anastasi, M. Conti, M. Di Francesco, and A. Passarella, "Energy conservation in wireless sensor networks: A survey," *Ad Hoc Networks*, vol. 7, no. 3, pp. 537–568, 2009.
- [35] C. Schurgers, V. Tsiatsis, and M. Srivastava, "Stem: Topology management for energy efficient sensor networks," in *Aerospace Conference Proceedings, 2002. IEEE*, vol. 3. IEEE, 2002, pp. 3–1099.
- [36] K. Nguyen, T. Nguyen, C. Chaing, and M. Motani, "A prioritized mac protocol for multihop, event-driven wireless sensor networks," in *Communications and Electronics, 2006. ICCE'06. First International Conference on*. IEEE, 2006, pp. 47–52.
- [37] H. Kim and S. Min, "Priority-based qos mac protocol for wireless sensor networks," in *Parallel & Distributed Processing, 2009. IPDPS 2009. IEEE International Symposium on*. IEEE, 2009, pp. 1–8.
- [38] N. Saxena, A. Roy, and J. Shin, "Dynamic duty cycle and adaptive contention window based qos-mac protocol for wireless multimedia sensor networks," *Computer Networks*, vol. 52, no. 13.
- [39] M. Yigitel, O. Incel, and C. Ersoy, "Design and implementation of a qos-aware mac protocol for wireless multimedia sensor networks," *Computer Communications*, 2011.
- [40] B. Nefzi and Y. Song, "Qos for wireless sensor networks: Enabling service differentiation at the mac sub-layer using cosens," *Ad Hoc Networks*, 2011.
- [41] C. Intanagonwiwat, R. Govindan, and D. Estrin, "Directed diffusion: a scalable and robust communication paradigm for sensor networks," in *Proceedings of the 6th annual international conference on Mobile computing and networking*, ser. MobiCom '00, 2000, pp. 56–67.
- [42] S. Darabi, N. Yazdani, and O. Fatemi, "Multimedia-aware mmspeed: A routing solution for video transmission in wmsn," in *Advanced Networks and Telecommunication Systems, 2008. ANTS '08. 2nd International Symposium on*, dec. 2008, pp. 1–3.
- [43] M. Krogmann, M. Heidrich, D. Bichler, D. Barisic, and G. Stromberg, "Reliable, real-time routing in wireless sensor and actuator networks," *ISRN Communications and Networking*, 2011.
- [44] A. Mahapatra, K. Anand, and D. P. Agrawal, "Qos and energy aware routing for real-time traffic in wireless sensor networks," *Computer Communications*, vol. 29, no. 4, pp. 437–445, 2006.
- [45] K. Akkaya and M. Younis, "An energy-aware qos routing protocol for wireless sensor networks," in *Distributed Computing Systems Workshops, 2003. Proceedings. 23rd International Conference on*, may 2003, pp. 710–715.
- [46] A. Mohajerzadeh and M. Yaghmaee, "An efficient energy aware routing protocol for real time traffic in wireless sensor networks," in *Ultra Modern Telecommunications Workshops, 2009. ICUMT '09. International Conference on*, oct. 2009, pp. 1–9.
- [47] C. Lu, B. Blum, T. Abdelzaher, J. Stankovic, and T. He, "Rap: a real-time communication architecture for large-scale wireless sensor networks," in *Real-Time and Embedded Technology and Applications Symposium, 2002. Proceedings. Eighth IEEE*, 2002, pp. 55–66.
- [48] J. Le, J. Lui, and D. M. Chiu, "Dcar: Distributed coding-aware routing in wireless networks," in *Distributed Computing Systems, 2008. ICDCS '08. The 28th International Conference on*, june 2008, pp. 462–469.
- [49] S. Katti, D. Katabi, H. Balakrishnan, and M. Medard, "Symbol-level network coding for wireless mesh networks," in *Proceedings of the ACM SIGCOMM 2008 conference on Data communication*, ser. SIGCOMM '08, 2008, pp. 401–412.
- [50] S. Chachulski, M. Jennings, S. Katti, and D. Katabi, "Trading structure for randomness in wireless opportunistic routing," in *Proceedings of the 2007 conference on Applications, technologies, architectures, and protocols for computer communications*, ser. SIGCOMM '07, 2007.
- [51] D. Noh, J. Kim, J. Lee, D. Lee, H. Kwon, and H. Shin, "Priority-based routing for solar-powered wireless sensor networks," in *Wireless Pervasive Computing, 2007. ISWPC '07. 2nd International Symposium on*, feb. 2007.
- [52] D. Noh, I. Yoon, and H. Shin, "Low-latency geographic routing for asynchronous energy-harvesting wsns," *Journal of Networks (JNW)*, vol. 3, no. 1, pp. 78–85, 2008.
- [53] O. Akan, "Performance of transport protocols for multimedia communications in wireless sensor networks," *IEEE Communications Letters*, vol. 11, no. 10, pp. 826–828, october 2007.
- [54] I. Almalkawi, M. Guerrero Zapata, J. Al-Karaki, and J. Morillo-Pozo, "Wireless multimedia sensor networks: current trends and future directions," *Sensors*, vol. 10, no. 7, pp. 6662–6717, 2010.
- [55] O. Akan and I. Akyildiz, "Event-to-sink reliable transport in wireless sensor networks," *IEEE/ACM Transactions on Networking*, vol. 13, no. 5, pp. 1003–1016, oct. 2005.
- [56] Y. Iyer, S. Gandham, and S. Venkatesan, "Step: a generic transport layer protocol for wireless sensor networks," in *Computer Communications and Networks, 2005. ICCCN 2005. Proceedings. 14th International Conference on*, oct. 2005, pp. 449–454.
- [57] J. Paek and R. Govindan, "Rrcrt: Rate-controlled reliable transport protocol for wireless sensor networks," *ACM Transactions on Sensor Networks*, vol. 7, pp. 20:1–20:45, October 2010.
- [58] F. Shaikh, A. Khelil, A. Ali, and N. Suri, "Trccit: Tunable reliability with congestion control for information transport in wireless sensor networks," in *Wireless Internet Conference (WICON), 2010 The 5th Annual ICST*, march 2010, pp. 1–9.
- [59] J. Mingorance-Puga, G. Macia-Fernandez, A. Grilo, and N. Tiglao, "Efficient multimedia transmission in wireless sensor networks," in *Next Generation Internet (NGI), 2010 6th EURO-NF Conference on*, june 2010, pp. 1–8.
- [60] S. Pudlewski and T. Melodia, "Dmrc: Distortion-minimizing rate control for wireless multimedia sensor networks," in *Mobile Adhoc and Sensor Systems, 2009. MASS '09. IEEE 6th International Conference on*, oct. 2009, pp. 563–572.
- [61] M. Alaei and J. Barcelo-Ordinas, "Node clustering based on overlapping fovs for wireless multimedia sensor networks," in *Wireless Communications and Networking Conference (WCNC), 2010 IEEE*. IEEE, 2010, pp. 1–6.
- [62] —, "A method for clustering and cooperation in wireless multimedia sensor networks," *Sensors*, vol. 10, no. 4, pp. 3145–3169, 2010.
- [63] —, "Mcm: multi-cluster-membership approach for fov-based cluster formation in wireless multimedia sensor networks," in *Proceedings of the 6th International Wireless Communications and Mobile Computing Conference*. ACM, 2010, pp. 1161–1165.