Multiview Image Compression and Transmission Techniques in Wireless Multimedia Sensor Networks: A Survey

Xiaoqin Wang, Y. Ahmet Şekercioğlu and Tom Drummond
Department of Electrical and Computer System Engineering, Monash University, Melbourne, Australia
Email: {Xiaoqin.Wang|Ahmet.Sekercioglu|Tom.Drummond}@monash.edu

***Abstract*—We present a survey of recent research works on multiview image compression and transmission techniques developed for Wireless Multimedia Sensor Networks (WMSNs). We classify them into two categories with respect to the coding methods adopted: (i) in-network processing with joint coding schemes, and (ii) distributed source coding schemes. The survey also includes a comprehensive evaluation of the limitations of each approach. Based on the review results, we discuss future research directions, and identify the ways of more efficient transmission of the spatially correlated and redundant data in WMSNs.**

I. Introduction

The recent advances in video technology, inexpensive camera sensors, and development of new distributed processing algorithms have enabled a new kind of wireless sensor networks (WSNs): Wireless Multimedia Sensor Networks (WMSNs). These networks primarily work by capturing multimedia content such as video/audio streams, and use wireless links [1]. They differ from the conventional wireless sensor networks that deal with scalar data. The potentially high-volume visual information captured in these networks creates unique challenges. The energy, bandwidth and processing capability constraints of the sensor nodes, as well as the nature of the wireless links that interconnect them, are more severe in WMSNs as compared to WSNs.

It is critically important to efficiently utilize the limited resources of WMSN nodes. For example, correlations in the data captured through multiple sensors have great potential to alleviate the restrictions imposed by the resource constraints. When the density of sensors increases or multiple camera sensors are deployed to provide multiple views of a field of interest, the correlation of the visual information observed by different sensors increases. If each sensor has to send the observed correlated visual content to a common receiver independently, the network traffic increases significantly with potentially highly redundant information. Obviously, processing and transmitting such redundant data over multi-hop links wastes channel capacity and energy resources of the nodes. Maximum coverage, through optimal placement of camera sensors, is an efficient method to cut down the redundant data transmission [2], [3]. Additionally, in a densely deployed network without a deterministic sensor placement, redundant sensors can be turned off to save channel capacity and energy in a self-actuated manner [4], [5] allowing the remaining sensors fulfilling the monitoring and routing functions effectively. However, multiple camera sensors with overlapped field of views (FoVs) provide multiple views, multiple resolutions and through that way, enhance observations of the environment, and become necessary in many applications such as object tracking, or 3-D reconstruction. In order to utilize the limited bandwidth efficiently in multiview environments, many research works are proposed to exploit correlation and minimize the amount of redundant visual data transmitted. In-network processing is a key design in correlation-based communication strategies. It indicates cooperative processing of visual content among intermediate sensors which reduces the amount of data transmitted throughout the network [1]. In this scenario, correlation among the views are exploited at encoders, and multimedia data is encoded jointly. The advantage of this scenario is the amount of overall multimedia data transmitted in the sensor network is minimized by eliminating redundancy at encoders. The distributed source coding (DSC) algorithms are another promising approaches to efficiently compress the redundant data in WMSNs[6]. Here, each encoder operates independently but relies on joint decoding at the sink. The advantage of this approach is an encoder does not need to share its information directly which can lead to significant energy saving of the nodes running the encoders.

The rest of the paper is organized as follows. In Section II, algorithms related to in-network processing with joint coding are discussed. Section III reviews the papers concerning distributed source coding for multiview images. Future research directions are discussed in Section IV, followed by our concluding remarks.

II. In-Network Processing with Joint Coding

Since the multiview images are usually highly correlated, joint coding schemes with encoders accessing images of multiple views achieve higher compression performance than traditional mechanisms with independent coding schemes. In these in-network processing schemes, the spatial correlation is estimated and removed at the encoders before transmission by using the joint coding schemes. Then the uncorrelated visual content is delivered in the network collaboratively.

A. Spatial Correlation Estimation

In this category, the correlation among images observed by camera sensors with overlapped FoVs should be determined first. The correlation characteristics for images observed by different cameras can be acquired through using geometric approach [7], [8], [9] and image processing approach [10], [11].

1) Geometric Approaches: Ma et al. [8] proposed an algorithm to determine correlation degree between two sensors by studying each camera sensor's location and FoV. The idea of this algorithm is to derive the correlation degree as the portion of overlapping sensing area to the entire area of the FoV. Camera sensors' locations and orientations can be estimated from the works in [12]. The view of the sensing field are treated as a set of discrete points. A criterion is introduced to determine whether each discrete point in one sensor's FoV falls in the other sensor's sensing regions or not. If a point is also in another sensor's FoV, the overlapping sensing area and the correlation degree between the images collected by these two camera sensors increase.

Han et al. [7] proposed a grid-based approach and a relative position-based approach to determine the correlation degree by using 3D directional sensing model. In the grid-based approach, the overall FoV is divided into small grids. Then each small grid is checked to determine if it is in other sensor's FoV. This approach is valid when the size of grid is much smaller than the size of FoV. In the relative position-based approach, the overlapped area is based on the relationship of camera sensors' pose. For details of the correlation degree calculation mechanism, [7] can be referred.

The works presented in [9], [13] use pinhole imaging model to acquire the correlation characteristics of visual information in WMSNs. The idea of the algorithm is to adopt the characteristics of projection model and distance disparity between corresponding positions of 7 fixed feature points on different image planes to determine the correlation coefficient.

2) Image Processing Approaches: In many cases, information of each camera sensor's location and FoV is not obtained by its neighbors and the central receiver. In such situations, the spatial correlation between multiview images can be exploited by using image processing approaches.

In the scheme proposed by Wu et al. [11], the spatial correlation between neighboring sensors is determined by image shape matching [14], and it is computationally lightweight. Regarding the constraints of bandwidth, energy, and computational capability of the image sensors, the observed image contents are represented by a small set of image feature points. In this algorithm, neighboring sensors with overlapping FoV should communicate with each other to determine the correlation degree between the images collected. This algorithm operates in three phases. In the first phase, dominant edges on images collected by sensors with overlapped FoV are extracted. Any edge detection algorithm can be adopted in this phase, such as Canny algorithm [15] and Robert operator [16]. And feature points are then extracted along the edges. In the second phase, a mechanism including a shape context descriptor and bipartite graph matching algorithm[14] is employed to find the best matching points between two images. After two groups of feature points are extracted from two correlated images by shape context descriptor, bipartite graph matching algorithm is adopted to exploit the best one-to-one matching between two sets of the feature points. In the third phase, after the finite set of correspondences between feature points on two shapes is obtained, modeling transformation [14] is adopted to estimate a plane transformation which can map arbitrary pixel on one image to another. Eventually, a displacement field that maps any position in one image to its interpolated location in the correlated image and the overlapping FoV can be determined by employing two separate thin plate spline (TPS) function [17], [18].

Stauffer et al. [10] developed an algorithm to determine the correlation between video without the knowledge of sensors' pose. The proposed algorithm works in a frame-to-frame tracking manner on a set of tracking sequences in each camera sensor to model the likelihood of direct correspondence. The sequences are comprised of discrete observations indexed by absolute occurrence time. The homography relating observations between camera sensors is estimated by using a combination approach of works in [19] and [20]. Once the homography relating observation is estimated, the overlapped region is determined by the intersection of the bounding rectangle of one camera sensor with the bounding rectangle of another camera sensor. However, this algorithm is only valid in the environment in which the camera sensors have unobstructed views.

Another multiview encoder has been proposed by San et al. [21], which exploits the inter-viewpoint correlation by a multiview geometry based disparity vector predictor. A disparity vector (DV) is defined as a vector connecting the corresponding points between the left image and right image [22]. In this scheme, images are decomposed into blocks. It first searches the corresponding object blocks of the same scene but from two different viewing images. It then projects the corresponding block pairs to the images captured by the other sensors. The offset between the corresponding block pairs position in the $f-1$th image and that in the fth image can be obtained. This offset is regarded as a DV candidate. Finally, it adopts a DV Fusion scheme to merge similar candidates into one predicted DV. This disparity vector based algorithm can track the corresponding image blocks in different views accurately and reduce the matching cost.

B. Multiview Image Transmission Strategies

With the knowledge of spatial correlation between multiple views at encoders, the information of the overlapping region can be avoided to be transmitted for multiple times. Many papers have addressed various collaborative image transmission frameworks and strategies to efficiently deliver correlated images in the networks.

In [8], a framework consist of three stages is developed to efficiently transmit correlated visual information and hence

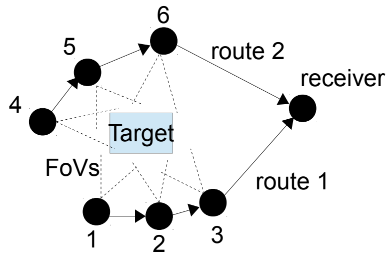

Fig. 1. An example of a multi-hop collaborative image transmission system.

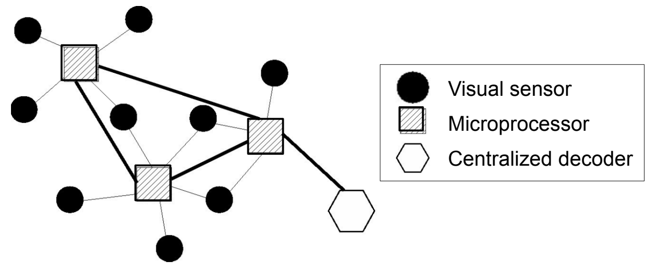

Fig. 2. An example of the network topology for framework in [25]

reduce the network load. In the proposed framework, camera sensors can cooperatively capture the scene and deliver partial information to the sink independently. In the first stage, camera sensors are grouped according to correlation matrix which is determined by using geometry approach. Camera sensors with maximum correlations are allocated in the same group. In order to obtain a fused image as if it was captured by a virtual camera located in the center of the group, various sensing tasks are assigned to different sensors. In the second stage, camera sensors just capture and deliver partial visual information according to the result in previous stage. Data will be sent to the sink independently from various sensors via different routes to balance the network load. In the final stage, once the images are received at sink, they will be fused together to construct a composite image.

Wagner et al. [23] proposed a collaborative in-network compression scenario with super-resolution recovery techniques applied at receiver. In this scenario, in order to determine maximal overlap, images from correlated views are first registered by using image matching method. This image matching method involving image feature points and feature points correspondence is similar to [11] introduced in Section II-A2. Then the low-resolution version of the common image blocks describing the overlapping region is transmitted from each sensor to the receiver. And the super-resolution techniques are applied at the receiver to reconstruct a high-resolution version of the overlapping region. In this work, the super-resolution algorithm require a relatively large number of low-resolution images to reconstruct overlapped region with an acceptable quality. Therefore the camera sensors need to be deployed densely which effectively limits the flexibility and the coverage area of the network.

A collaborative image transmission system is developed in [11] to overcome the limitations in [23]. Fig. 1 shows an example of the network topology. In this system, it is assumed that each sensor could perform feature-based image matching, and sensors on the route to the monitoring center node could access image data collected from previous hops. At each sensor along the path (sensor 2, 3, 5, 6 in Fig. 1), feature extraction and matching are operated on the image received from the previous hop (original image) and the image observed by current hop (reference image). In this phase, the spatial redundancy among images is removed. For example, the redundancy in the images captured by sensor 1 and 2, sensor 2 and 3, sensor 4 and 5, and sensor 5 and 6 is removed in a hop-by-hop manner. Then a transformation process operates to generate the difference image based on the result of redundancy removal. Then, only the original image and difference image are transmitted to the next hop. This will significantly reduce transmission energy comparing with transmitting several individual images independently.

Instead of transmitting low-resolution images of the overlapping region, optimal fractions of the overlapped image are transmitted by various image sensors in the model proposed in [24]. The proposed model works in three phases. In the first phase, the overlapping regions and the non-overlapping regions in the images observed by multiple camera sensors are separated. In order to save the energy at the sensors, each sensor transmits complete visual data corresponding to non-overlapping region and a portion of the visual data corresponding to the overlapping region. Therefore the correlated visual information of the overlapping region can be avoided to be sent repeatedly. In the second phase, a multipath routing protocol, named multi-level rate-oriented routing (MLRR), is developed to find the multiple node-disjoint routes from source sensors to the sink. In this protocol, sensors are classified into multiple levels according to their distance to sink node, and lower data rate are assigned to sensor nodes with less residual energy. And the formula is conducted to compute the total energy consumption of transmitting all image information via multiple paths. In the third phase, the optimized fractions of visual information in overlapping region transmitted by each sensor are determined to maximize network lifetime. Therefore, the total transmitted data is reduced throughout the network, and the end-to-end quality is preserved. In such scenario, extra processing from the sink is expected to reconstruct the viewed image based on received packets of separated image regions.

Chia et al. [25] defined a mechanism for centralized image compression in WMSNs. In this approach, in order to reduce the energy consumption of image processing and transmission at the source sensors, raw or encoded images can be processed by other nodes around the source sensors. The basic idea behind this framework is to construct a topology to support the desired in-network compression in which only some of nodes (microprocessor) have higher processing power for image processing while the other nodes are just pure image sensors without any processing power. Fig. 2 illustrates an example of network topology described in [25]. The camera sensors first transmit the observed images to microprocessor. The overlap redundancy in the neighboring camera sensors is removed at microprocessor by using image stitching technique [26]. The compression technique in [25] is a modified set partitioning in hierarchical trees (SPIHT) algorithm which exploits the inherent similarities across the subbands in a wavelet decomposition of an image. In this framework, image stitching and modified SPIHT algorithm operates on each microprocessor. And if one microprocessor is down, the data can be routed to another nearest microprocessor which grantees the robustness of the proposed framework.

Instead of removing redundancy after camera sensors capture visual information, Dai et al. [9] proposed a mechanism which removes the information redundancy by only activating partial camera sensors. This mechanism adopts an entropy-based approach. Joint entropies between each two camera sensors are acquired first. In order to maximize the joint entropy of multiple camera sensors, the correlation among the active camera sensors should be minimized. A correlation-based algorithm to maximize the joint entropy of multiple cameras is proposed. This proposed algorithm can select camera sensors with least correlation in their captured information and achieve maximum joint entropy.

A summary of the in-network processing schemes with joint coding is shown in Table I.

III. Distributed Source Coding of Multiview Images

Distributed source coding has emerged as a promising alternative technology for achieving efficient compression of multiview images. Distributed source coding refers to the compression of correlated signals captured by various sensors which do not communicate with each other [27]. Surprisingly, based on the foundation works of Slepian-Wolf (SW) [28] and Wyner-Ziv (WZ) [29], it is shown that the distributed encoding with joint decoding can be just as efficient as the wholly joint system. Moreover, regarding the computationally constrained encoders in WMSNs, distributed source coding makes itself more suitable for WMSNs by its ability of shifting the computational complexity from encoder side to decoder side.

A. Background on Distributed Source Coding

From a theoretical perspective, distributed source coding lies its foundation in Wyner-Ziv theorems [29]. Wyner and Ziv considered the problem on coding of two correlated sources X and Y with respect to a fidelity criterion [29]. In their work, they generalized the setup of [28] in which coding of source X is with respect to a fidelity criterion rather than lossless. They have developed a rate distortion (RD) function $R_{*X|Y}(D)$ for the case in which the side information Y is available at the decoder but not at the encoder. For a given target distortion is the required rate to code X if Y is known at both encoder and decoder, and $R_X(D)$ is the minimum rate for coding X without side information. Wyner and Ziv show that there is no rate loss between the joint coding and joint decoding of two correlated Gaussian sources. This result has been extended to the cases in which sources X and Y can have any arbitrary distributions with Gaussian distributed innovation between them.

In order to research Wyner-Ziv limit, both source codes and channel codes are employed in practical coding schemes constructed on the Wyner-Ziv theorem. These coding schemes usually comprise a quantizer followed by a Slepian-Wolf coder. In the coding process, source X is quantized with respect to the fine source code into quantization cells. A codeword is associated with each cell to construct the course codebook. And in order to save rate, only the index of the coset containing the course codeword is coded and transmitted [27]. Based on the coded index, the decoder can recover the source codeword in a given coset by locating the codeword which is the closest to the side information Y.

B. Applications of Distributed Source Coding in Wireless Multimedia Sensor Networks

Zhu et al. [30] proposed a distributed compression framework based on Wyner-Ziv codec with both conventional encoders and Wyner-Ziv encoders in the network. Images acquired from some camera sensors are coded independently using conventional coding techniques such as JPEG. The other camera sensors encode the observed views by using Wyner-Ziv coder. It is assumed no interconnections amongst camera sensors, thus the views are encoded at each sensors independently. The centralized decoder first reconstructs scene geometry from the conventional encoded images, and generate an estimation for the image at each sensor equipped with Wyner-Ziv encoder using the geometry constraint. Then the decoder requests parity bits from each Wyner-Ziv sensor and uses the estimation images as the side information to decode. As the turbo codec consists in Wyner-Ziv codec used in this framework, additional parity bits from the encoders are requested by the centralized decoder until decoding is successful. This feedback mechanism plays an important role which can effect the decoding quality and bandwidth constraint by adaptively requesting parity bits. However, as some camera sensors transmit their full view encoded in conventional method (to provide side information) while the other only transmit parity information, this proposed approach is highly asymmetric which can shorten the network lifetime and can be a problem for some practical applications.

In the framework proposed by Chen et al. [31], the correlation between two multiview images is exploited in the pixel domain. Unsupervised learning of disparity between two multiview images is achieved in proposed Wyner-Ziv multiview image codec by using the Expectation Maximization (EM) algorithm [32]. One image Y is transmitted by conventional coding, and the other image X is encoded independently by a Wyner-Ziv encoder. Based on advanced channel coding technique, the Wyner-Ziv encoder transform quantized transform coefficients of image X to a low-density parity check (LDPC) bitstream. In the loop composed the LDPC decoder, the disparity estimator and the side information generator, each

TABLE I
A COMPARISON OF DIFFERENT IN-NETWORK PROCESSING SCHEMES WITH JOINT ENCODERS

SCHEME	CORRELATION ESTIMATION	ADVANTAGES	DISADVANTAGES
Ma et al. [8]	Geometry approach	Each sensor only need to transmit a fraction of overlapped view.	Accurate information of each sensor's pose need to be obtained.
Dai et al. [9]	Geometry approach	Only M of N ($M \leq N$)camera sensors need to be activated	Accurate information of each sensor's pose need to be obtained; network load is imbalance .
Wagner et al. [23]	Image processing approach	Each sensor only need to transmit a low-resolution version of overlapped view.	Camera sensors have to be deployed densely.
Wang et al. [24]	Image processing approach	Each sensor only need to transmit a fraction of overlapped view.	Missing detailed information such as efficient correlation estimation algorithm to make the scheme realistic
Wu et al. [11]	Image processing approach	Significant reduce transmitted energy along the routing path.	Not robust: network may not work after some sensors die.
Chia et al. [25]	Image processing approach	Robustness: network can still work when some sensors die.	Central nodes require high processing ability and energy consumption.

iteration of LDPC decoding sends a soft estimation of X to the disparity estimator, while the side information for the LDPC decoder is renewed by the disparity estimator in return. Therefore, the decoder learns the disparity between image X and image Y progressively and decodes a lossy reconstruction of image X by employing the received LDPC bitstream and side information image Y.

Different from the distributed compression approaches relying on the use of advanced channel coding techniques, Gehrig et al. [33], [34] proposed a distributed compression approach exploiting the correlation of multiview data by adopting a fully geometrical approach which does not use channel codes. This mechanism represents each scan-line of the images as piecewise polynomials by adopting the prune-join tree decomposition algorithm [35]. Thus, signal is encoded and represented with a quadtree structure which models its segmentation and a set of polynomials. For a scenario with only two camera sensors, only one camera sensor is required to transmit full description of its view to decoder. A threshold is proposed to determine how many information in the subtree of the other sensor's quadtree structure require to be sent to decoder. And each camera needs to send its location of discontinuities following image bits independently. At the decoder, according to each encoder's location information and epipolar constraints, the information acquired from both encoders is used to retrieve all the disparities and match each polynomial piece of the first view with its corresponding region in the second view. In this way, the image observed by second sensor is reconstructed at the decoder. In this distributed compression mechanism, only the necessary information, which allows for a complete reconstruction at decoder, is transmitted. However, the locations of camera sensors are strictly constrained and camera sensors should be deployed in a line. This constraint would highly affect the coverage of the network.

Tošić and Frossard [36] proposed an approach which does not require a special camera sensor arrangement in a WMSN, since more diverse types of geometric correlation among multiview images, such as shifts, rotations and anisotropic scaling, can be exploited by this approach. In this mechanism, the correlation between multiview images is modeled by local geometric transforms of prominent image features called atoms. This model is established on a parametric redundant dictionary of atoms. This parametric redundant dictionary can be constructed by using a dictionary learning algorithm introduced in [37]. Each image can be approximated by a linear combination of a small number of vectors from this dictionary. When the parametric dictionary is adopted on both images, the transform of atoms on the first image to the corresponding atoms on the second image can be identified by two constraints, which are shape similarity constraint and epipolar constraint. A Wyner-Ziv coding scheme is developed on this geometric correlation model. The first image (reference image) is encoded independently, while the second image (Wyner-Ziv image) is encoded by coset coding of atom indexes and the quantization of their respective coefficients [36]. At the decoder, the reference image is used as the side information to decode the Wyner-Ziv image. The corresponding atoms in the reference image and atoms within the coset of the Wyner-Ziv image are matched by decoder. A disparity map which consists of local transforms between the reference image and Wyner-Ziv image is built. Finally, the Wyner-Ziv image can be reconstructed based on the reference image and the disparity map.

In order to enable the ability to deal with occlusion between multiview images, occlusion-resilient coding scheme in proposed in Tošić's later work [38]. In this coding scheme, encoders require to send additional information of the atoms which are not adequately distinguished to appear in sparse approximation in both views. However, as the encoders cannot communicate with each other in distributed compression mechanisms, the encoders have no idea about which of the atoms are occlusions and non-distinguished features in other views. Therefore, in the proposed coding scheme, encoder makes an assumption that at least M out of total N atoms represent occlusions which leads to failures in decoding process. And the shape and position cosets of these atoms are transmitted to decoder additionally.

A framework proposed in [39] comprises of distributed representation of correlated images with quantized linear measurements and the joint decoding algorithm exploiting geometrical correlation among images. The correlation model of this framework is based on the geometric transformations

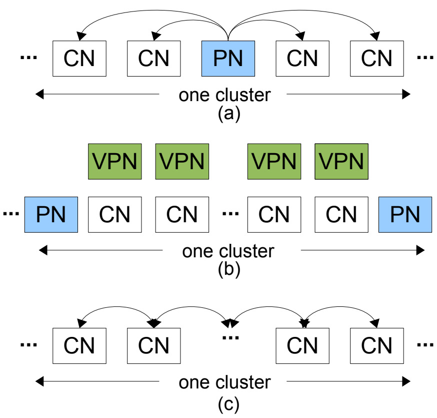

Fig. 3. Network configurations in [41]

acquired by a structured dictionary to provide an effective estimation of the correlation between multiview images. When a scene is observed by two camera sensors, the image captured by the first sensor plays the role of the reference and the side information for the decoding of the second image. The reference image X_1 can be encoded in any conventional compression algorithms. While a uniform quantization algorithm and quantized linear measurements are applied on the second image X_2. Then the measurements of second image are compressed with an entropy encoder. Image features inferring the geometry information are first estimated at the joint decoder. A parametric dictionary of geometric functions is proposed to derive the sparse approximation of the coded first image $\hat{X}_1$ with its geometric features. Instead of finding the corresponding features in the reconstructed second image at decoder, the corresponding features can be estimated from the quantized linear measurements of the second image. The correlation information is generated for further use in decoding the compressed image $\hat{X}_2$ with the side information $\hat{X}_1$. And the experimental results has shown the proposed framework with geometry-based correlation model is more efficient than the distributed compression frameworks with block-based correlation model [40].

Different network configurations for distributed source coding of multiview images were analyzed by Tehrani et al. [41]. This distributed source coder is based on algorithm proposed by Ozonat [42] with parent node (PN) in a network. In [41], a cluster of correlated camera sensors includes PN and child nodes (CNs). The PN transmits full observed image to the receiver, while the CN only transmit partial information. The receiver can reconstruct the view at each CN using the full captured information of PN as the side information. Fig. 3 depicts three kinds of the network configurations proposed in [41]. In the first two kinds of network configurations, each cluster is composed of one PN at least. In a configuration with only one PN, the PN is used to decode all CNs in the same cluster. And in the configuration with two PNs, two view images observed at the corners are chosen as PNs and used to decode all CNs in the middle. And algorithm generating the middle images and using the interpolated image [43] as the virtual parent node (VPN) is proposed to decode partial information of the CNs. In the third type of the network configuration, there is no PN in the network, and the partial information sent by closest neighboring sensors are used to decode partial information at each sensor. In this case, the communications between camera sensors and receiver are symmetrical.

A summary of the distributed coding schemes is shown in Table II.

TABLE II
A COMPARISON OF DISTRIBUTED SOURCE CODING SCHEMES

SCHEME	ADVANTAGES	DISADVANTAGES
Zhu et al. [30]	Use turbo coding to enhance decoding quality in a iterative manner	Serious asymmetric communication
Chen et al. [31]	Gray code is used for bit representation of transform coefficients, which enhances PSNR gain and performance.	Provides limited advantage if disparity is known.
Gehrig et al. [33], [34]	Transmitted data is reduced to the least at encoders.	Decoder needs camera sensors' pose information; deployment of sensors is highly restricted.
Tošić er al. [36]	Sensors can be deployed arbitrarily	System has to be trained first to obtain a disparity dictionary.
Tošić er al. [38]	Occlusion can be identified.	More information has to be sent to decoder.
Thirumalai et al. [39]	More efficient than mechanisms using image block-based correlation model	System has to be trained first to obtain a dictionary of geometric functions.
Tehrani et al. [41]	Symmetrical communication in the network	Decoding quality varies to network structures.

IV. RESEARCH DIRECTIONS

In the past few years, many image compression techniques exploiting correlation among multiview images have been proposed. Those research works provide a wealth of contributions to bandwidth saving, energy preservation and error recovery in WMSNs. However, the problem of achieving efficient image compression and transmission in resource-constrained WMSNs is not completely solved yet. In this section, several research directions for multiview image coding and delivery mechanisms are discussed.

In WMSNs, significant number of novel image coding techniques with varying performance outcomes in terms of data compression rate, computational complexity or error recovery have been proposed. In the techniques that fall under the category of in-network processing with joint coding, camera sensors can communicate with each other and encode cooperatively. The redundant data is removed at the source nodes, and the data delivered in the network is minimized. In the methods that use distributed source coding, each camera sensor encodes its data independently without communicating with each other, while joint decoding is performed at receiver. We believe that the future research will follow the trends in these two kinds of mechanisms.

According to the in-network processing mechanisms surveyed in Section III, the embedded image processing algorithms adopted in WMSNs are mainly adapted from existing computer vision algorithms with little modification, and they rarely take the constraints of underlying wireless networks into consideration. Therefore, future research works should focus on reducing the computational complexity of on-board image processing. Algorithms of finding the disparity between multiview images should be lightweight. It means the number of feature points should be minimized while the disparities can be correctly determined. Thus, the selection procedure of the proper feature points is a promising research direction. On the other hand, in order to reduce the energy consumption on data exchange between camera sensors and fit bandwidth constraint, each camera sensor should obtain the ability of determining and only sending/broadcasting information which is unknown at the other sensors. This mechanism can be achieved by letting each camera sensor acquire the knowledge of other camera sensors' pose. It is also suggested in some works that data exchanged between camera nodes should be aggregated in-network at one of the camera sensors, and the decision about the most suitable data fusion center should be dynamic regarding the best field of view and the communication costs [44]. Finally, further research should explore the trade-offs between the energy consumption in on-board compression process and the energy consumption in communication process should be investigated and determined to minimize the overall energy consumption in the networks.

Based on the works surveyed in Section IV, distributed source coding can shift the complexity from source nodes to the receiver end, fitting well to the needs of WMSNs. However, problems still exist and keep distributed compression schemes from practice. One of the most challenging tasks in distributed source coding is identifying the correlation structure between multiview images at the decoder side, especially without the knowledge of network topology and the pose of camera sensors. A training phase can enable this knowledge at decoder, while more energy will be consumed. Moreover, accurate synchronization between packets transmitted from the source nodes is the prerequisite of distributed compression schemes. And the implementation of the distributed source coding will effect all the other network stack layers below [45]. This will entail a tight coupling between the distributed coding algorithms and the MAC, network, and link-layer protocols, organization, and channel conditions of the network, as well as the power supply, transmitter/receiver and kernel scheduler of each sensor. Therefore, future research requires to consider cross-layer designs and distributed compression schemes together to achieve optimal performance.

Even though a distributed compression scheme is usually regarded as the most suitable coding scheme for correlated images in WMSNs, we have reservations. Without a computationally resource intensive image compression through redundancy-removal at camera sensors, it is true that distributed compression mechanisms save energy in source nodes. However, compared with in-network processing which can minimize data delivered in the network, a distributed compression scheme suffers from consuming considerable amount of energy in the communication process. It should be added that, communications in distributed compression schemes are asymmetrical which can lead to fast energy draining in some sensors. Moreover, limited energy storage at camera sensors is not the only constraint for WMSNs, limited bandwidth, end-to-end delay and QoS requirements also influence the performance of the systems. Many in-depth research studies which consider vision-based cross-layer optimization offer a number of feasible methods for overcoming the shortages of in-network processing with joint encoding. Therefore, due to various intended purposes of different applications, there is not a universally adequate coding technique for all kinds of WMSNs, and the selection of the suitable coding scheme should consider application requirements and the characteristics of the network. Therefore, based on this consideration, future research could focus on developing adaptive coding schemes which can shift between joint compression and distributed compression schemes according to various scenarios.

V. Concluding Remarks

Image compression and transmission is a well-established area. The focus of this paper is to review the multiview images compression and transmission in WMSNs, and to point out the unique characteristics and constraints of spatially correlated image processing. We classified the state-of-the-art of the research works of multiview image processing strategies into two main approaches: joint compression schemes with in-network processing and distributed compression schemes without communication between sensors. We analyzed and compared those multiview image processing mechanisms. The drawbacks of the surveyed works were also discussed. Many problems still need to be addressed through future research. Finally, future research directions were presented. We believe that there is no preferred coding scheme for multiview image in WMSNs, since the requirement of the application and the characteristics of the network configurations will strongly affect the selection of the appropriate compression and transmission mechanism. And the breakthrough will germinate only through a comprehensive strategy which takes the image coding scheme with sensor deployment, network management and hardware issues into consideration.

References

[1] I. F. Akyildiz, T. Melodia, and K. R. Chowdhury, "A survey on wireless multimedia sensor networks," *Comput. Netw.*, vol. 51, no. 4, pp. 921–960, Mar. 2007.

[2] J. Zhao, S.-C. Cheung, and T. Nguyen, "Optimal camera network configurations for visual tagging," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 2, no. 4, pp. 464 –479, aug. 2008.

[3] J.-J. Gonzalez-Barbosa, T. Garcia-Ramirez, J. Salas, J.-B. Hurtado-Ramos, and J.-d.-J. Rico-Jimenez, "Optimal camera placement for total coverage," in *Robotics and Automation, 2009. ICRA '09. IEEE International Conference on*, may 2009, pp. 844 –848.

[4] A. Newell and K. Akkaya, "Self-actuation of camera sensors for redundant data elimination in wireless multimedia sensor networks," in *Communications, 2009. ICC '09. IEEE International Conference on*, june 2009, pp. 1 –5.

[5] T.-S. Chen, J.-J. Peng, D.-W. Lee, and H.-W. Tsai, “Prediction-based object tracking and coverage in visual sensor networks,” in *Proceedings of 7th International Conference on Wireless Communications and Mobile Computing (IWCMC), 2011*, july 2011, pp. 278 –284.

[6] D. G. Costa and L. A. Guedes, “A survey on multimedia-based cross-layer optimization in visual sensor networks,” *Sensors*, vol. 11, no. 5, pp. 5439–5468, 2011.

[7] C. Han, L. Sun, F. Xiao, J. Guo, and R. Wang, “Fovs correlation model for wireless multimedia sensor networks based on three-dimensional perception,” *Advances in Information Sciences and Service Sciences*, vol. 4, no. 3, pp. 27 –34, Feb. 2012.

[8] H. Ma and Y. Liu, “Correlation based video processing in video sensor networks,” in *Proceedings of 2005 International Conference on Wireless Networks, Communications and Mobile Computing.*, vol. 2, june 2005, pp. 987 – 992 vol.2.

[9] R. Dai and I. Akyildiz, “A spatial correlation model for visual information in wireless multimedia sensor networks,” *Multimedia, IEEE Transactions on*, vol. 11, no. 6, pp. 1148 –1159, oct. 2009.

[10] S. C and K. Tieu, “Automated multi-camera planar tracking correspondence modeling,” in *Proceedings of 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition.*, vol. 1, june 2003, pp. I–259 – I–266 vol.1.

[11] M. Wu and C. W. Chen, “Collaborative image coding and transmission over wireless sensor networks,” *EURASIP J. Appl. Signal Process.*, vol. 2007, no. 1, pp. 223–223, Jan. 2007.

[12] D. Devarajan, Z. Cheng, and R. Radke, “Calibrating distributed camera networks,” *Proceedings of the IEEE*, vol. 96, no. 10, pp. 1625 –1639, oct. 2008.

[13] P. Wang, R. Dai, and I. Akyildiz, “Collaborative data compression using clustered source coding for wireless multimedia sensor networks,” in *INFOCOM, 2010 Proceedings IEEE*, 2010, pp. 1–9.

[14] S. Belongie, J. Malik, and J. Puzicha, “Shape matching and object recognition using shape contexts,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 4, pp. 509 –522, apr 2002.

[15] J. Canny, “A computational approach to edge detection,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. PAMI-8, no. 6, pp. 679 –698, nov. 1986.

[16] P. J. Besl and R. C. Jain, “Three-dimensional object recognition,” *ACM Comput. Surv.*, vol. 17, no. 1, pp. 75–145, Mar. 1985. [Online]. Available: http://doi.acm.org/10.1145/4078.4081

[17] J. Duchon, “Splines minimizing rotation-invariant semi-norms in sobolev spaces,” in *Constructive Theory of Functions of Several Variables*, ser. Lecture Notes in Mathematics, W. Schempp and K. Zeller, Eds. Springer Berlin / Heidelberg, 1977, vol. 571, pp. 85–100, 10.1007/BFb0086566.

[18] J. Meinguet, “Multivariate interpolation at arbitrary points made simple,” *Zeitschrift fr Angewandte Mathematik und Physik (ZAMP)*, vol. 30, pp. 292–304, 1979, 10.1007/BF01601941.

[19] Y. Caspi, D. Simakov, and M. Irani, “Feature-based sequence-to-sequence matching,” *Int. J. Comput. Vision*, vol. 68, no. 1, pp. 53–64, 2006.

[20] L. Lee, R. Romano, and G. Stein, “Monitoring activities from multiple video streams: establishing a common coordinate frame,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 8, pp. 758 –767, aug 2000.

[21] X. San, H. Cai, J.-G. Lou, and J. Li, “Multiview image coding based on geometric prediction,” *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 17, no. 11, pp. 1536 –1548, nov. 2007.

[22] Y. Kima, S. Choia, S. Chob, and K. Sohn, “Efficient disparity vector coding for multiview sequences,” *Signal Processing, Image Communication*, vol. 19, no. 6, pp. 539 –553, may 2004.

[23] R. Wagner, R. Nowak, and R. Baraniuk, “Distributed image compression for sensor networks using correspondence analysis and super-resolution,” in *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on*, vol. 1, sept. 2003, pp. I – 597–600 vol.1.

[24] H. Wang, D. Peng, W. Wang, and H. Sharif, “Optimal rate-based image transmissions via multiple paths in wireless sensor network,” in *Multimedia and Expo, 2007 IEEE International Conference on*, july 2007, pp. 2146 –2149.

[25] W. C. Chia, L.-M. Ang, and K. P. Seng, “Multiview image compression for wireless multimedia sensor network using image stitching and spiht coding with ezw tree structure,” in *Intelligent Human-Machine Systems and Cybernetics, 2009. IHMSC '09. International Conference on*, vol. 2, aug. 2009, pp. 298 –301.

[26] M. Brown and D. Lowe, “Automatic panoramic image stitching using invariant features,” *International Journal of Computer Vision*, vol. 74, pp. 59–73, 2007, 10.1007/s11263-006-0002-3.

[27] C. Guillemot, F. Pereira, L. Torres, T. Ebrahimi, R. Leonardi, and J. Ostermann, “Distributed monoview and multiview video coding,” *Signal Processing Magazine, IEEE*, vol. 24, no. 5, pp. 67 –76, sept. 2007.

[28] D. Slepian and J. Wolf, “Noiseless coding of correlated information sources,” *Information Theory, IEEE Transactions on*, vol. 19, no. 4, pp. 471 – 480, jul 1973.

[29] A. Wyner and J. Ziv, “The rate-distortion function for source coding with side information at the decoder,” *Information Theory, IEEE Transactions on*, vol. 22, no. 1, pp. 1 – 10, jan 1976.

[30] X. Z. Anne, A. Aaron, and B. Girod, “Distributed compression for large camera arrays,” in *in Proceedings of the IEEE Workshop on Statistical Signal Processing*, pp. 30–33.

[31] D. Chen, D. Varodayan, M. Flierl, and B. Girod, “Wyner-ziv coding of multiview images with unsupervised learning of disparity and gray code,” in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, oct. 2008, pp. 1112 –1115.

[32] D. Varodayan, Y. chung Lin, A. Mavlankar, M. Flierl, and B. Girod, “Wyner-ziv coding of stereo images with unsupervised learning of disparity,” in *in: Proc. Picture Coding Symp*, 2007.

[33] N. Gehrig and P. Dragotti, “Dfferent - distributed and fully flexible image encoders for camera sensor networks,” in *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, vol. 2, sept. 2005, pp. II – 690–3.

[34] N. Gehrig and P. L. Dragotti, “Distributed compression of multi-view images using a geometrical coding approach,” in *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, vol. 6, 16 2007-oct. 19 2007, pp. VI –421 –VI –424.

[35] R. Shukla, P. Dragotti, M. Do, and M. Vetterli, “Rate-distortion optimized tree-structured compression algorithms for piecewise polynomial images,” *Image Processing, IEEE Transactions on*, vol. 14, no. 3, pp. 343 –359, march 2005.

[36] I. Tosic and P. Frossard, “Geometry-based distributed scene representation with omnidirectional vision sensors,” *Image Processing, IEEE Transactions on*, vol. 17, no. 7, pp. 1033 –1046, july 2008.

[37] I. Tošić and P. Frossard, “Distributed multi-view image coding with learned dictionaries,” in *Proceedings of the 5th International ICST Mobile Multimedia Communications Conference*, ser. Mobimedia '09. ICST, Brussels, Belgium, Belgium: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2009, pp. 27:1–27:7.

[38] I. Tosic and P. Frossard, “Geometry-based distributed coding of multi-view omnidirectional images,” in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, oct. 2008, pp. 2220 –2223.

[39] V. Thirumalai and P. Frossard, “Distributed representation of geometrically correlated images with compressed linear measurements,” *Image Processing, IEEE Transactions on*, vol. PP, no. 99, p. 1, 2012.

[40] J. Prades-Nebot, Y. Ma, and T. Huang, “Distributed video coding using compressive sampling,” in *Picture Coding Symposium, 2009. PCS 2009*, 2009, pp. 1–4.

[41] M. P. Tehrani, T. Fujii, and M. Tanimoto, “Distributed source coding of multiview images,” in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, ser. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, S. Panchanathan and B. Vasudev, Eds., vol. 5308, Jan. 2004, pp. 300–309.

[42] K. Ozonat. (2002) Lossless distributed source coding for highly correlated still images. [Online]. Available: http://www.stanford.edu/class/ee368b/Projects/ozonat/index.html

[43] A. Nakanishi, T. Fujii, T. Kimoto, and M. Tanimoto, “Ray-space data interpolation by adaptive filtering using locus of corresponding points on epipolar plane image,” *J. Inst. Image Inf. Telev. Eng. (ITE)*, vol. 56, no. 8, pp. 1321 –1327, 2002.

[44] S. Soro and W. Heinzelman, “A survey of visual sensor networks,” *Advances in Multimedia*, vol. 2009, no. 640386, 2009. [Online]. Available: http://www.hindawi.com/journals/am/2009/640386/ref/

[45] Z. Xiong, A. Liveris, and S. Cheng, “Distributed source coding for sensor networks,” *Signal Processing Magazine, IEEE*, vol. 21, no. 5, pp. 80 – 94, sept. 2004.