An Empirical Model of Multiview Video Coding Efficiency for Wireless Multimedia Sensor Networks

Stefania Colonnese, *Member, IEEE*, Francesca Cuomo, *Senior Member, IEEE*, and Tommaso Melodia, *Member, IEEE*

***Abstract*—We develop an empirical model of the Multiview Video Coding (MVC) performance that can be used to identify and separate situations when MVC is beneficial from cases when its use is detrimental in wireless multimedia sensor networks (WMSN). The model predicts the compression performance of MVC as a function of the correlation between cameras with overlapping fields of view. We define the *common sensed area* (CSA) between different views, and emphasize that it depends not only on geometrical relationships among the relative positions of different cameras, but also on various *object-related phenomena*, e.g., occlusions and motion, and on *low-level phenomena* such as variations in illumination. With these premises, we first experimentally characterize the relationship between MVC compression gain (with respect to single view video coding) and the CSA between views. Our experiments are based on the H.264 MVC standard, and on a low-complexity estimator of the CSA that can be computed with low inter-node signaling overhead. Then, we propose a compact empirical model of the efficiency of MVC as a function of the CSA between views, and we validate the model with different multiview video sequences. Finally, we show how the model can be applied to typical scenarios in WMSN, i.e., to clustered or multi-hop topologies, and we show a few promising results of its application in the definition of cross-layer clustering and data aggregation procedures.**

***Index Terms*—Multiview video coding, MVC efficiency model, video sensor networks.**

I. Introduction

Wireless multimedia sensor networks (WMSNs) can support a broad variety of application-layer services, especially in the field of video surveillance [1], [2] and environmental monitoring. The availability of different views of the same scene enables multi-view oriented processing techniques, such as video scene summarization [3], moving object detection [4], face recognition [5], depth estimation [6], among others. Enhanced application-layer services that rely on these techniques can be envisaged, including multi-person tracking, biometric identification, ambience intelligence, and free-view point video monitoring.

Recent developments in video coding techniques specifically designed to jointly encode multiview sequences (i.e., sequences in which the same video scene is captured from different perspectives) can provide compact video representations that may enable more efficient resource allocation. Roughly speaking, cameras whose fields of view (FoV) are significantly overlapped may generate highly correlated video sequences, which can in turn be jointly encoded through multiview video coding (MVC) techniques. Nevertheless, MVC techniques introduce moderate signaling overhead. Therefore, if MVC techniques are applied to loosely (or not at all) correlated sequences that differ significantly (for instance, because of the presence of different moving objects), MVC may provide equal or even lower compression performance than encoding each view independently.

The MVC coding efficiency can be predicted by theoretical models (see for instance [10]). Nonetheless, the adoption of theoretical models allows a qualitative rather than a quantitative analysis of the coding efficiency. As far as the quantitative prediction is concerned, a simple theoretical model can introduce relatively high percentage errors. Thereby, investigation of theoretical models taking into account further parameters (number of moving objects, object-to-camera depths, occlusions, discovered areas, amount of spatial details) is still an open issue. Besides, a theoretical model dependent on several parameter could result useless when applied to networking problems in a MWSN, where the model parameters must be estimated at each node and periodically signaled among nodes.

For these reasons, we turn to an empirical model of the efficiency of MVC in order to accurately identify and separate situations when MVC coding is beneficial from cases when its use is detrimental. The empirical model provides an accurate quantitative prediction of the coding efficiency while it keeps low the processing effort and inter-node signaling overhead.

Based on these premises, we derive an empirical model of the MVC compression performance as a function of the correlation between different camera views and then discuss its application to WMSN. We define the *common sensed area* (CSA) between different views, and emphasize that the CSA depends not only on geometrical relationships among the relative positions of different cameras, but also on several real *object related phenomena*, (i.e., occlusions and motion), and on *low-level phenomena* such as illumination changes. With these premises, we experimentally characterize the relationship between MVC compression gain (with respect to the single view video coding, denoted in the following as AVC-Advanced Video Coding) and the estimated CSA between views. Our experiments are based on the recently defined standard [7] that extends H.264/AVC to multiview, while we estimate the CSA by means of a low-complexity inter-view common area estimation procedure. Based on

Manuscript received August 01, 2012; revised December 12, 2012 and March 28, 2013; accepted April 05, 2013. Date of publication June 27, 2013; date of current version November 13, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Charles D. (Chuck) Creusere.

S. Colonnese and F. Cuomo are with the DIET, Universitá "La Sapienza" di Roma, 00184 Roma, Italy (e-mail: colonnese@infocom.uniroma1.it; francesca.cuomo@uniroma1.it).

T. Melodia is with the Department of Electrical Engineering, The State University of New York at Buffalo, Buffalo, NY 14260 USA (e-mail: tmelodia@buffalo.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMM.2013.2271475

this experiments, we propose an empirical model of the MVC efficiency as a function of the CSA between views. Finally, we present two case studies that highlight how the model can be leveraged for cross-layer optimized bandwidth allocation in WMSNs.

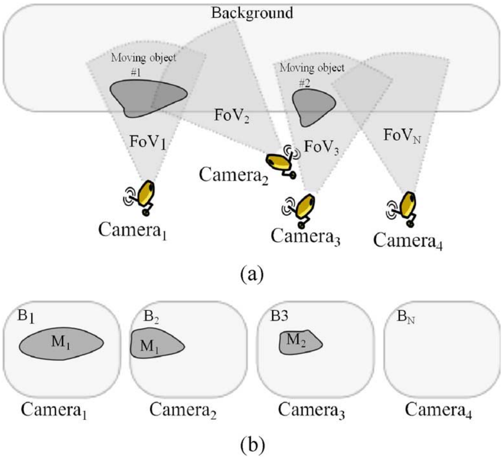
In a nutshell, our model summarizes the similarities between different views in terms of a single parameter that i) can be estimated through inter-node information exchange at a low signaling cost and ii) can be used to predict the relative performance of MVC and AVC in network resource allocation problems. The main contributions of the paper are therefore as follows:

- After introducing the notion of CSA between overlapped views, we provide an experimental study of the relationship between MVC efficiency and CSA. The core novelty of this study is that, unlike previous work, we evaluate the MVC efficiency as a function of a parameter related to the scene content rather than to the geometry of the cameras only. Preliminary studies on this relationship have been presented in [8]. In this paper we extend these studies to different video sequences.
- Based on the experimental data, we introduce a compact empirical model of the relative compression performance of MVC versus AVC as a function of the estimated CSA. In the proposed model, the MVC efficiency is factored in through i) a scaling factor describing the efficiency of the temporal prediction and ii) a factor describing the efficiency of the inter-view prediction; the latter is expressed as a function of the CSA only. Our model is the first attempt to predict the performance of MVC from an easy-to-compute parameter that goes beyond camera geometry considerations, and takes into account moving objects and occlusions.
- After discussing some practical concerns (signaling overhead, CSA estimation), we present two case studies (single hop clustering scheme and multi hop aggregation toward the sink) in which we show how the proposed model can be applied to WMSN to leverage the potential gains of MVC.

The structure of the paper is as follows. In Section II, we discuss the multimedia sensor network model, while in Section III we review the state of the art in MVC encoding for WMSNs. After introducing the notion of common sensed area in Section IV, in Section V we define the relative efficiency of MVC versus AVC and establish experimentally the relationships between the efficiency of MVC and the common sensed area. Based on this, in Section VI we propose an empirical model of the MVC efficiency, and evaluate its accuracy on different video sequences. Finally, Section VIII concludes the paper.

II. Multimedia Sensor Network Scenario

A WMSN is typically composed of multiple cameras, with possibly overlapping FoVs. The FoV of a camera can be formally defined as a circular sector of extension dependent on the camera angular width, and oriented along the pointing direction of the camera. A given FoV typically encompasses static or moving objects at different depths positioned in front of a far-field still background. An illustrative example is reported in Fig. 1(a).



(a)

(b)

Fig. 1. Example scenario (a) and different image planes (b).

The camera imaging device performs a radial projection of real-world object points into points of the camera plane where they are effectively acquired. According to the so-called pinhole camera model, those points can be thought of as belonging to a virtual plane, named image plane, located outside the camera at the same distance from the focal length as the camera plane.[1] For instance, the image planes corresponding to the scenario in Fig. 1(a) are shown in Fig. 1(b). Note that while every point in the image plane has a corresponding point in the FoV, not all the points in the FoV correspond to points in the image plane, due to the occlusions between objects at different depths.

We observe that while the FoVs depend on characteristics exclusively of the camera such as position, orientation, angular view depth, the effectively acquired images resulting from the projection of real-world objects on the image plane depend on the effectively observed scene. First, each near-field object partially occludes the effective camera view to an extent depending on the object size and on the object-to-camera distance. Besides, the same real-world object may be seen from different points of view and at different depths by different cameras. Therefore, the views provided by the nodes of a WMSN may correspond to image planes characterized by different degrees of similarity, depending both on the camera locations and on the framed scene. The view similarity can be exploited to improve the compression efficiency through MVC.

III. Related Work

The problem of compressing correlated data for transmission in WMSNs has been recently debated in the literature. Several papers have shown that the transmission of multimedia sensors towards a common sink can be optimized in terms of rate and energy consumption if correlation among different views is taken into account. In [9], highly correlated sensors covering the same object of interest are paired to cooperatively perform

[1] Each real-world point framed by the camera is mapped into the acquisition device, on the internal camera plane, where acquisition sensors are located on a grid. According to the so-called pinhole camera model, the camera plane on which the numerical image is formed is associated to a virtual plane, symmetric with respect to the camera pinhole. This virtual plane, called *image plane*, is regarded as a possibly continuous-domain representation of the numerical image acquired by the camera.

the sensing task by capturing part of the image each. The pioneering work of [1] demonstrated that a correlation-based algorithm can be designed for selecting a suitable group of cameras communicating toward a sink so that the amount of information from the selected cameras can be maximized. To achieve this objective, the authors design a novel function to describe the correlation characteristics of the images observed by cameras with overlapped FoVs, and they define a disparity value between two images at two cameras depending on the difference θ of their sensing direction. A clustering scheme based on spatial correlation is introduced in [10] to identify a set of coding clusters to cover the entire network with maximum compression ratio. The coefficient envisaged by Akyildiz *et al.* [1] measures, in a normalized fashion, the difference in sensing direction under which the same object is observed in two different camera planes. This measure is strongly related to the warping transformation that is established between the representations of the same object in the two considered image planes. The larger the sensing difference, the more sensible warping (object rotation, perspective deformation, shearing) is observed. Since canonical inter-frame prediction procedures (such as those employed in the various versions of the H.264 encoding standard) can efficiently encode only simple, translatory transformations, more general warping transformations observed between images captured at large sensing directions are not efficiently encoded by MVC procedures. Thereby, the performance of MVC encoding is expected to worsen as long as the difference in sensing direction increases.

The coefficient in [1] is related to the angular displacement between two cameras. Indeed, even with cameras characterized by low difference in the sensing direction, occlusions among real world foreground objects in the FoVs may cause the acquired images to differ significantly. Besides, motion of objects may result in time-varying inter-view similarity even if the WMSN nodes maintain the same relative positions. These observations motivate us to consider a different, scene-related, parameter accounting for key phenomena that affect the correlation between views.

There are several possible choices for the similarity measure to be used. Here, we propose to capture the two phenomena of occlusion and object motion by introducing the notion of Common Sensed Area (CSA) between views, and propose a low-complexity correlation-based CSA estimator. Based on this, we are able to measure and model the efficiency of MVC techniques as a function of the CSA value between two views. While the empirical studies presented in this paper are based on this simple estimator, the general framework presented in this paper is certainly compatible with more refined (but computationally more expensive) view-similarity based estimators of the CSA such as those recently discussed in [11]. Advanced feature-mapping techniques that appeared in the literature can also be adopted, e.g., [12]–[14]. In any case, the CSA estimation accuracy shall be traded off with the cost (computational complexity, signaling overhead) required to compute it within a WMSN.

IV. Common Sensed Area

In this Section, we characterize the view similarity by means of a parameter depending not only on geometrical relationships among the relative positions of different cameras, but also on several real *object related phenomena*, namely occlusions, motion, and on *low-level phenomena* such as illumination changes. To this aim, we start by formulating a continuous model of the images acquired by different cameras.

The acquisition geometry is given by a set of N cameras, with assigned angular width and FoV (as in Fig. 1(a)). Real-world objects framed by the video cameras are mapped into the image plane. Let us consider the luminance image $I^{(i)}(x,y)$ acquired at the i-th camera at a given time.[2] Each image point (x_i, y_i), represents the radial projection, on the i-th image plane, of a real point (u,v,z). We define the background B_i as the set of points (x_i, y_i) resulting from projections of real points (u,v,z) that belong to static objects. Similarly, we define the foreground M_i as the set of points resulting from projections of points belonging to real moving objects. Thus, the domain D_i of the the i-th acquired image $I^{(i)}(x,y)$ is partitioned as $\mathrm{D}_i = \mathrm{B}_i \cup \mathrm{M}_i$, $\mathrm{B}_i \cap \mathrm{M}_i = \emptyset$, and we can express the image $I^{(i)}(x,y)$ as

$$I^{(i)}(x,y) = b^{(i)}(x,y) + m^{(i)}(x,y), \tag{1}$$

with $b^{(i)}(x,y) = 0$, $(x,y) \notin \mathrm{B}_i$, and $m^{(i)}(x,y) = 0$, $(x,y) \notin \mathrm{M}_i$. The background B_i and moving object M_i supports can be estimated through existing algorithms [15], [16]. Partitions corresponding to the example scenario in Fig. 1(a) are reported in Fig. 1(b); cameras 1 and 2 capture points belonging to the same moving object M_1, camera 3 captures projections of moving object M_2 and camera 4 captures only background points.

Let us now consider a pair of images $I^{(i)}(x,y), I^{(j)}(x,y)$, acquired by cameras with possibly overlapping FoVs. We are interested in defining the CSA between the two views. To this aim, let us define the common background CB_{ij} between image i and j as the set of points in B_i representing static background points appearing also in B_j, namely

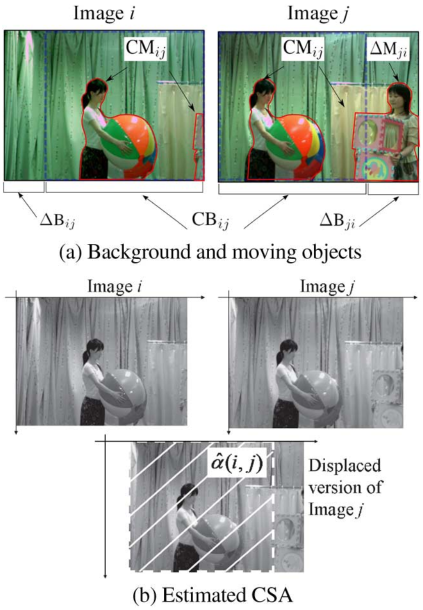
$$\begin{aligned} \mathrm{CB}_{ij} = \{ &(x_i, y_i) \in \mathrm{B}_i, (x_i, y_i) s.t. \\ &\text{if } (x_i, y_i) = P_i(u,v,z), \\ &\;\mathrm{then}\ (x_j, y_j) = P_j(u,v,z) \in \mathrm{B}_j \} \end{aligned} \tag{2}$$

Further, let CM_{ij} be defined as

$$\begin{aligned} \mathrm{CM}_{ij} = \{ &(x_i, y_i) \in \mathrm{M}_i, (x_i, y_i) s.t. \\ &\text{if } (x_i, y_i) = P_i(u,v,z), \\ &\;\mathrm{then}\ (x_j, y_j) = P_j(u,v,z) \in \mathrm{M}_j \} \end{aligned} \tag{3}$$

that is, CM_{ij} is the set of points in M_i representing real-world moving object points whose projections appear in M_j. Let us observe that, although originated by the same object, the luminance values assumed on the sets $\mathrm{CB}_{ij}, \mathrm{CM}_{ij}$ in different cameras' images differ because of the different perspective warping under which the scene is acquired and of several other acquisition factors, including noise and illumination. An example appears in Fig. 2(a), showing two images that include moving objects and a background; image i shares with image j a part of a common background and two common moving objects. The 2-D sets $\mathrm{CB}_{ij}, \mathrm{CM}_{ij}$ of the i-th image and $\mathrm{CB}_{JI}, \mathrm{CM}_{ji}$ of the

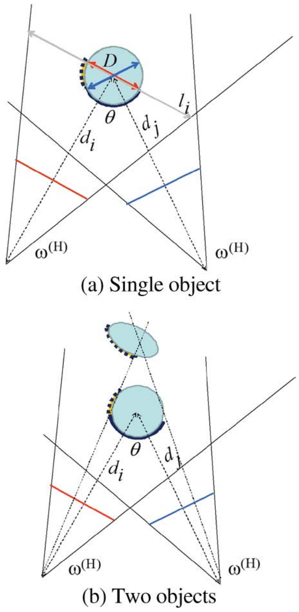
[2] For the sake of simplicity, we disregard the effect of discretization of the acquisition grid. Nevertheless, such a simplified model can be properly extended to take into account the discrete nature of the acquired image.



(a) Background and moving objects

(b) Estimated CSA

Fig. 2. Example of common backgrounds and moving objects and estimated CSA $\hat{\alpha}(i,j)$. (a) Background and moving objects; (b) Estimated CSA.



(a) Single object

(b) Two objects

Fig. 3. Single and multiple objects geometry (example). (a) Single object; (b) Two objects.

i-th image are shown. Note that the extension of the common sets differ in the two image planes, since they depend on the particular view angle.

Based on the afore defined sets, we can formally define the CSA between views. Let us consider the image $i^{(i)}[m,n]$ obtained by sampling $I^{(i)}(x,y)$ on a discrete grid with sampling interval $(\Delta x, \Delta y)$. We define the CSA $\alpha(i,j)$ as

$$\alpha(i,j) = \frac{|\mathrm{CB}_{ij} \cup \mathrm{CM}_{ij}|}{|\mathrm{D}_i|}, \tag{4}$$

where $|S|$ denotes the number of pixels of the i-th image $i^{(i)}[m,n]$ such that $(m\Delta x, n\Delta y) \in S$.

Hence, the CSA $\alpha(i,j)$ is formally defined as the ratio between the number of pixels belonging to the common areas of two images i and j and the overall number of pixels of the image captured by camera i.

The definition of $\alpha(i,j)$ in (4) allows us to identify the factors that affect the similarity between camera views, accounting for occlusions and uncovered background phenomena between different cameras.

A. CSA and Scene Geometry

The CSA $\alpha(i,j)$ between two cameras is indeed related to the 3D geometrical features of the framed scene, and it depends not only on the angular distance between cameras but also on the objects' positions, occlusions, etc.

To show an example of the relation between the CSA and the scene characteristics, let us sketch out a case in which only one object, cylindric in shape, is within the FOVs of two cameras C_i, C_j at distance d_i, d_j from the object itself. We assume that the cameras are placed at the same height, so as to develop the analysis of the CSA between cameras by taking into account only the horizontal dimension. We denote by D the object diameter, and by h the object height. We assume that both the cameras are pointed towards the object center. The geometry of the scene is sketched in Fig. 3(a).

According to definition (4), the CSA can be evaluated as

$$\alpha(i,j) = \frac{n^{(CB)} + n^{(CM)}}{n_r \times n_c}, \tag{5}$$

where n_r, n_c respectively denote the number of rows and columns of the i-th image, and $n^{(CB)}$, $n^{(CM)}$ denote the number of pixels in the sets $\mathrm{CB}_{ij}, \mathrm{CM}_{ij}$.

In order to put in evidence how the parameters describing the scene geometry affects the CSA, we now sketch out how the evaluation of $n^{(CM)}$ can be carried out. Firstly, we evaluate the size of the rectangular area occupied by the object in the image framed by C_1. To this aim, let us denote by l_i the spatial width framed by the camera C_i at the distance d_i. We recognize that the width is

$$l_i = d_i tg\left(\frac{\omega_i}{2}\right) \tag{6}$$

being $\omega_i^{(H)}$ the horizontal angular camera width. The horizontal size of the object equals to

$$n_c^{(M)} = n_c \frac{D}{l_i} = n_c \frac{D}{d_i} \cdot \cot\left(\frac{\omega_i^{(H)}}{2}\right) \tag{7}$$

Besides, we recognize that the vertical size of the object equals to

$$n_r^{(M)} = n_r \frac{h}{d_i} \cdot \cot\left(\frac{\omega_i^{(V)}}{2}\right) \tag{8}$$

being $\omega_i^{(V)}$ the vertical angular camera width. The remaining image pixels are occupied by projections of a subset B_I of the background points, that is the subset of background points which are not occluded by the object itself.

Let us now consider the second camera C_j, and let ϑ denote the angular distance between the cameras. The second camera C_j captures a different part of the cylindric surface. Specifically, C_j captures a new (i.e. not visible by C_i) surface corresponding to a sector of angular width ϑ, whereas a specular surface visible by C_i is not yet visible by C_j. In order to evaluate $n^{(CM)}$, we now compute the size of the rectangular area belonging to the object which is still visible in C_j.

Let us point out that, the projection of the cylindric surface corresponding to an elementary angular sector $\delta\vartheta$ has an extension that varies depending on the angle ξ formed by the normal to the surface and the direction of the camera axis. As ξ varies from $\pi/2$ to $\pi/2 - \vartheta$, the projection of the surface varies. From geometrical considerations only, we derive the horizontal width of the object visible in both cameras as:

$$n_c^{(CM)} = \frac{1}{2} n_c cot\left(\frac{\omega_i^{(H)}}{2}\right)\frac{D}{d_i}\cdot\left(1+\frac{\sin(2\vartheta)}{1-\frac{1}{2}\frac{d}{D}\sin(\vartheta)}\right) \quad (9)$$

for $\vartheta \leq \pi/2$. Since the camera are at the same height, we obtain $n^{(CM)} = n_r^{(M)} \times n_c^{(CM)}$.

The value $n^{(CM)}$ is then straightforwardly related to the object diameter, to the distances with respect two the cameras; more in general, also the object shape, the inclination of the object surface with respect to the cameras' axis should be taken are account for by the $n^{(CM)}$ value. With similar considerations, it is possible to carry out the computation of the common background pixels $n^{(CB)}$, which in turn requires additional hypotheses about the background surface shape (planar, curve).

The case of multiple objects is more complex. In Fig. 3(b), we illustrate an example in which the cameras are placed at the same angular distance as in the preceding example. There are two objects, and there is a partial occlusion between them in both the cameras. Due to this phenomenon, the second-plane object points that are visible by the first camera are not at all visible by the second camera, and the CSA drastically reduces to to inter-object occlusions. As far as a more realistic object and scene model is taken into account, the CSA depends by additional parameters describing the object and scene features.

Thereby, we recognize that the CSA depends not only on the depth, volume and shape of each object, but also on the relative positions between objects. The complexity of the analytical characterization of the CSA becomes rapidly sophisticated when more realistic scene settings are considered. Although the analysis can be useful in a particular constrained framework, such as a video surveillance in a fixed geometry framework (e.g. indoor, corridors), where a few parameter are constrained, a general solution is hardly found. Besides, the analysis looses relevance in a realistic WMSN framework, where the scene features are in general erratic and time-varying, so that dynamical and accurate estimation of the main time varying scene features is a critical task.

To sum up, the CSA, being related to the image content, is indeed related to the characteristics of the framed scene and of the cameras. Precisely, we recognize that the CSA depends not only on the camera positions but also on the actual objects positions, depths and relative occlusions. Thereby, the CSA can be regarded as a parameter summarizing those characteristics. For this reason, the CSA is better suited than other parameters in the literature, depending only on the camera geometry, to estimate and track the actual similarity between different views of the framed scene.

In the following, we face the problem of CSA evaluation by approximating the CSA $\alpha(i,j)$ with its estimate $\hat{\alpha}(i,j)$ computed by means of a low-complexity correlation-based estimator. This computation can be executed in real time and does not require segmentation and identification of moving objects from static background area for each view. Remarkably, our coarse, cross-correlation based, approach implicitly taken into account also low-level related phenomena may cause the views to differ, such as acquisition noise and so on, since these phenomena result in an increase of the mean square difference between the luminance of the images. This does not prevent further refinements in CSA estimation, using advanced similarity measures, such as advanced feature-mapping techniques [12]–[14]. or even resorting to more refined (but computationally more expensive) view-similarity estimators such as those recently discussed in [11].

With these positions, we present a simulative study and introduce an empirical model of the MVC efficiency with respect to AVC as a function of the CSA between views, thus providing the rationale for dynamic adoption of the MVC coding scheme in a WMSN.

V. Common Sensed Area and MVC Efficiency

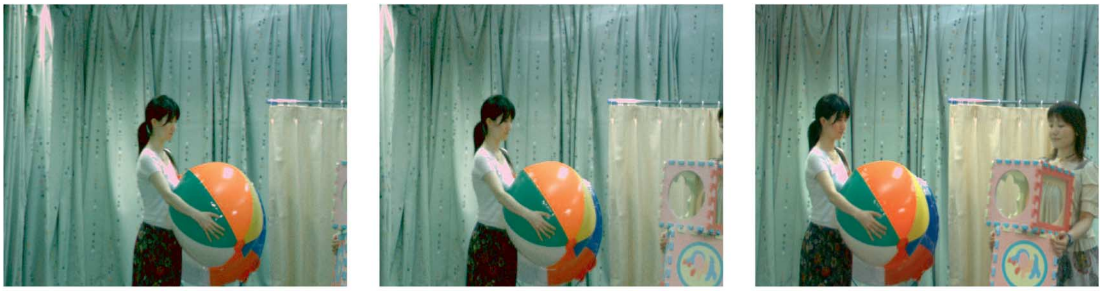
To quantify the benefits of MVC with respect to multiple AVC, we introduce here the *MVC relative efficiency parameter* $\eta(i,j)$, which depends on the bit rates generated by the encoder for the video sequence j when the sequence i is also observed (and therefore known at the codec).

Let us consider a pair of cameras i and j and let us denote by $r^{AVC}(i)$ the overall bit rate generated by the codec in case of independent encoding (AVC) of the sequence acquired by i-th camera, referred to in the following as i-th view; besides, let $r^{MVC}(i,j)$ denote the overall bit rate generated by the codec in case of joint encoding (MVC) of i-th and j-th views.

The efficiency $\eta(i,j)$ is defined as

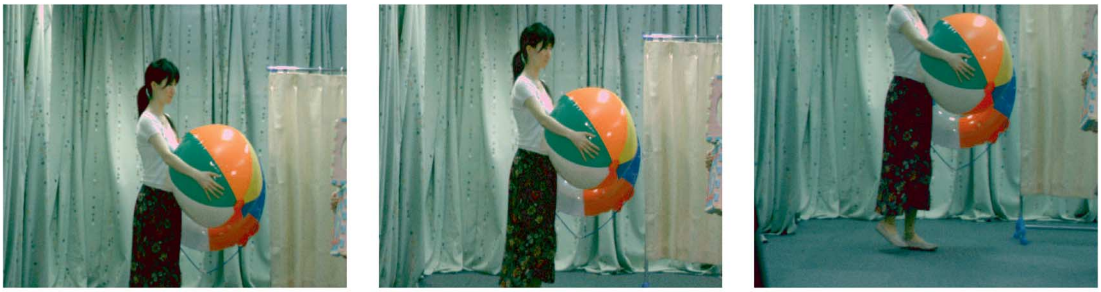
$$\eta(i,j) = 1 - \frac{r^{MVC}(i,j)}{r^{AVC}(i) + r^{AVC}(j)}, \quad (10)$$

and can be interpreted as the gain achievable by jointly encoding the sequences i and j with respect to separate encoding. In case of a pair of sequences, we can also denote as $\Delta r^{MVC}(j;i)$ the bit rate of the differential bit stream generated to encode the j-th view once the bit stream of the i-th view is known, i.e. $r^{MVC}(i,j) = r^{AVC}(i) + \Delta r^{MVC}(j;i)$. The bit rate generated by the MVC codec depends on the intrinsic sequence activity, which depends on the presence of moving objects in the framed video scene, as well as on the CSA between the considered camera views; for MVC to be more efficient than AVC, it must hold $\Delta r^{MVC}(j;i) \leq r^{AVC}(j)$. In the following, we show how this condition can be predicted given the CSA of the i-th and j-th cameras. Specifically, we will assess the behavior of $\eta(i,j)$ as a function of $\alpha(i,j)$ through experimental tests, and we will derive an empirical model $\eta(\alpha)$ matching these experimental results. Finally, we will discuss how this model can be leveraged in a WMSN.



(a) View 0 (b) View 5 (c) View 10

Fig. 4. Selected camera views of Akko&Kayo sequences, horizontal displacement. (a) View 0; (b) View 5; (c) View 10.



(a) View 20 (b) View 40 (c) View 80

Fig. 5. Selected camera views of Akko&Kayo sequences, vertical displacement. (a) View 20 (b) View 40; (c) View 80.

A. Experimental Setup

We consider the recently defined H.264 MVC [17]; the study can be extended to different, computationally efficient encoders [2] explicitly designed for WMSNs. The CSA is estimated on a per-frame basis as the number of pixels in the rectangular overlapping region between the view $I^{(j)}[m,n]$, and a suitably displaced version of the second view $I^{(j)}[m,n]$, as shown for instance in Fig. 2(b).[3] Despite the coarseness of this computation, according to which the CSA is at its best estimated as the area of the rectangular bounding box of the true CSA, the experimental results show that this fast estimation technique is sufficient to capture the inter-view similarity for the purpose of estimating the MVC efficiency.

The video coding experiments presented here have been conducted using the JMVC reference software. on different MPEG multiview test sequence. The first considered sequence is Akko&Kayo [19], acquired by 100 cameras organized in a 5 × 20 matrix structure, with 5 cm horizontal spacing and 20 cm vertical spacing. The experimental results reported here have been obtained using a subset of 6 (out of 100) camera views, i.e., views 0, 5, 10, 20, 40 and 80; the 0-th, 5-th and 10-th cameras are horizontally displaced in the grid while the 20-th, 40-th and 80-th cameras are vertically displaced with respect to the 0-th camera. The first frames corresponding to each of the selected cameras are shown in Figs. 4 and 5. The Akko&Kayo sequence present several interesting characteristics, since the FoVs of the cameras include different still and moving objects (a curtain, persons, balls, boxes), and movements and occlusions occur to different extent.

TABLE I
H.264 AVC/MVC ENCODING SETTING

AVC Bit-rate	Spatial resolution	GOP	Entropy coding	Basis QP
48 Kb/s	QCIF (176×144)	8	CAVLC	32

In the followings, we also consider the Kendo and Balloons multiview test video sequences [19]. For these two sequences, we have considered 7 and 6 views respectively, as acquired by uniformly separated cameras deployed on a line, with 5 cm horizontal spacing; the first frames corresponding to the different cameras are shown in Figs. 9 and 10.

B. Experimental Results

We first present simulations that quantify the relationship between the estimated CSA and the observed MVC efficiency.

We begin by presenting results that were obtained by resampling the sequences at QCIF spatial resolution, and at 15 frames per second, since such format is quite common in monitoring applications and it is compatible with the resource constraints of WMSNs. The different views were AVC-encoded and MVC-encoded using view 0 as a reference view. The basis QP is set to 32. A summary of the fixed encoder settings is reported in Table I.

We encoded the Akko&Kayo sequence through AVC and MVC with a Group of Picture (GOP) structure[4] of $M = 8$ frames; in the MVC case, the view of the camera #0 has been selected as the reference view. For fair comparison of the coding results, the MVC coding cost $r^{MVC}(i,j)$ and the multiple

[3] For a given couple of frames, the displacement is chosen so as to maximize the inter-view normalized cross-correlation $\rho(i,j)$, defined as

$$\rho(i,j) = \frac{\sum_{m,n} I^{(i)}[m,n] \cdot I^{(i)}[m+m_0, n+n_0]}{\sigma_i \cdot \sigma_j}$$

with $\sigma_i^2 = \sum_{m,n}(I^{(i)}[m,n] - \sum_{k,l} I^{(i)}[k,l])^2$ and $\sigma_j^2 = \sum_{m,n}(I^{(j)}[m,n] - \sum_{k,l} I^{(j)}[k,l])^2$.

[4] In the Multiview encoding scheme, in the reference view a picture every M is encoded using an INTRA coding mode for the reference view; in the dependent view a picture every M is encoded exploiting inter-view prediction from the contemporary reference view intra frame, using the anchor frame coding mode.

AVC coding cost $r^{AVC}(i) + r^{AVC}(j)$ have been compared under the constraint of equal average decoded sequence quality, measured in terms of Peak Signal-to-Noise Ratio[5] (PSNR), specifically, the PSNR_{AVC}, averaged on the 6 considered views, equals 32.24 dB with 1 dB standard deviation, whereas the PSNR_{MVC}, averaged on the 6 considered views, takes on the value of 32.59 dB with 0.34 dB standard deviation.[6] Therefore, the coding comparison is carried out on a fair, equal quality, basis.

In Figs. 7 and 8 we plot $\eta(i,j)$ for the Akko&Kayo sequence as a function of the frame index for both the horizontal and vertical pairs; for comparison, in Fig. 7 we also plot the MVC efficiency η_{00} of sequence 0 MVC-encoded using as a reference the sequence 0 itself.

As already observed in [20] under a different experimental setting, the MVC efficiency achieves its maximum value on frames which are encoded without motion compensation with respect to preceding frames; such frames, realizing both random access and error resilience functionalities, are named *intra frames* in the reference view and *anchor frames* in the dependent views. Besides, the MVC efficiency decreases mainly in the horizontal direction (pairs 0–5 and 0–10) rather than in the vertical one (pairs 0–10, 0–20 and 0–40), and it changes in time (apart from the 0-0 case) because of movement of real objects.

We now show that, while the absolute rates of MVC and AVC encoding significantly vary with encoding settings such as QP or the spatial resolution, the MVC efficiency, being related to the ratio between the MVC and AVC rates, is substantially independent of these settings.

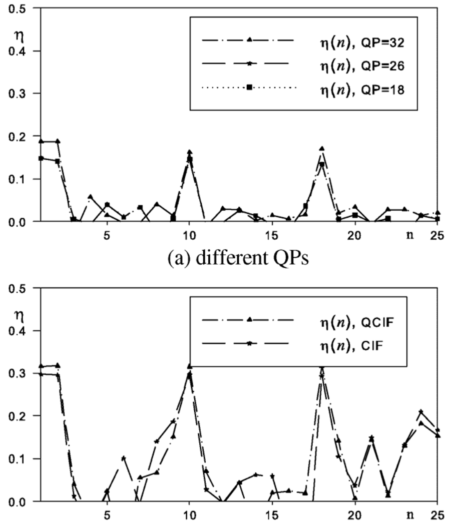
An example of this is shown in Fig. 6(a), where we plot the efficiency vs the frame index for two views of the QCIF version of the test sequence Akko&Kayo, encoded using different QP values (32, 26 and 18). The encoder settings are as in Table II. Although the actual bit-rates differ by a factor of two or four, the measured efficiency spans the same values and presents a common behavior in the different cases. By comparing Fig. 6(a) and Figs. 7 and 8, we observe that the ratio between the MVC and AVC rates depends on the dissimilarity between views, but is basically independent of the encoder settings.

The same trend is observed when the spatial resolutions varies. Fig. 6(b) plots efficiency vs frame index for two views of the test sequence Balloons, encoded at QCIF and CIF resolutions using the same basis $\mathrm{QP} = 24$, resulting in an average rate around 184 kbit/s and 514 kbit/s for the AVC encoded reference views, respectively (see Table III for encoder settings). This confirms that the efficiency analysis carried out for a given resolution extends to different resolutions.

To sum up, similar trends are observed at different values of resolution and QP. Therefore, without loss of generality, in the following we extensively and systematically study the impact of the estimated CSA on the observed MVC efficiency, and we assume the encoder settings to be fixed as in Table I.

On the Akko&Kayo sequence, we also calculated the low-complexity cross-correlation based CSA estimator between the k-th frames of reference view 0 and of the j-th sequence view, namely $\hat{\alpha}_k(0,j)$. The average of the values $\hat{\alpha}_k(0,j)$ over the temporal index k is reported in the second column of Table IV for different values of the view index j. In the same table, we report the corresponding MVC versus AVC relative efficiency (third column), as measured on each of the 6 encoded view pairs $(0,j)$ of the Akko&Kayo sequences.

As already observed in Figs. 7 and 8, we recognize a definite trend between those estimates of the relative H.264/MVC versus



(a) different QPs

(b) different resolutions

Fig. 6. Efficiency vs frame index: (a) at different QPs (Akko&Kayo sequence, QCIF resolution, QPs = 32, 26, 18) and (b) at different resolutions (Balloons sequence, QP = 24, QCIF and CIF resolutions.)

TABLE II
AKKO&KAYO SEQUENCE ENCODING AT DIFFERENT QPS: H.264 AVC/MVC SETTING

Basis QP	AVC Bit-rate	Spatial resolution	GOP	Entropy coding
18	226 kbit/s	QCIF (176×144)	8	CAVLC
26	98 kbit/s	QCIF (176×144)	8	CAVLC
32	48 kbit/s	QCIF (176×144)	8	CAVLC

TABLE III
BALLOONS SEQUENCE ENCODING AT DIFFERENT SPATIAL RESOLUTIONS: H.264 AVC/MVC SETTING

Spatial resolution	Basis QP	AVC Bit-rate	GOP	Entropy cod.
QCIF (176×144)	24	184 kbit/s	8	CABAC
CIF (352×288)	24	514 kbit/s	8	CABAC

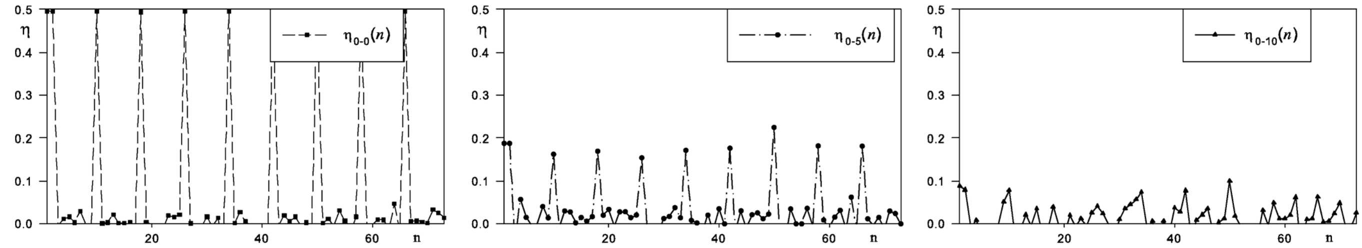
TABLE IV
MEASURED AVERAGE EFFICIENCY AND CORRESPONDING AVERAGE $\hat{\alpha}(i,j)$ FOR DIFFERENT VIEWS IN CASE OF AKKO&KAYO

View pairs	Average value of $\hat{\alpha}(i,j)$	Measured efficiency
0-0	1	0,251
0-5	0.811	0.092
0-10	0.618	0.034
0-20	0.855	0.132
0-40	0.702	0.067
0-80	0.542	0.023

[5] For an $M \times N$ original image $I[m,n]$ represented with l-bit luminance depth, and the corresponding encoded and decoded image $\tilde{I}[m,n]$, the PSNR is defined as

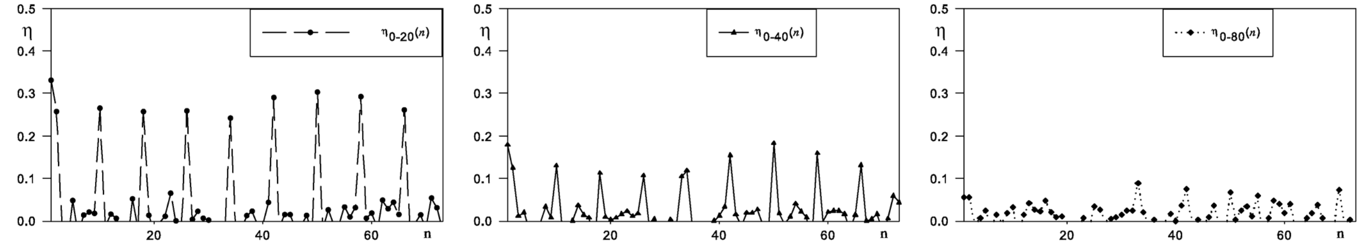
$$\mathrm{PSNR} = \frac{M \cdot N \cdot 2^{2l}}{\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \left(I[m,n] - \tilde{I}[m,n] \right)^2}$$

[6] These values have been obtained for an absolute coding cost of view-0 AVC coding of 48 Kb/s, using a minimum QP equal to 32.



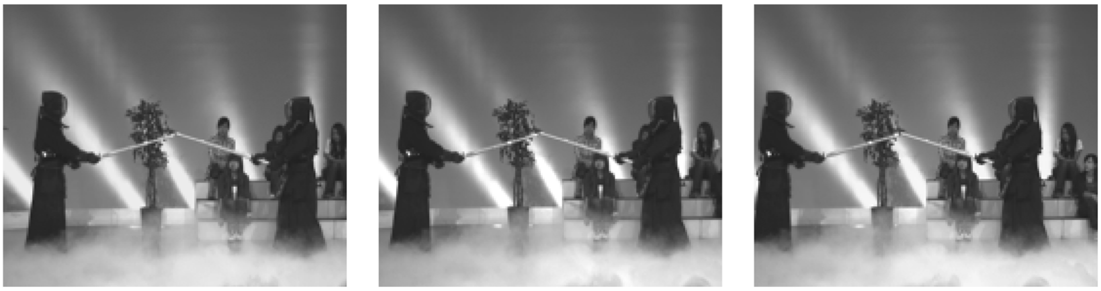
(a) Views 0, 0 (b) Views 0, 5 (c) Views 0, 10

Fig. 7. MVC efficiency as a function of time, using view 0 as a reference: sequence Akko&Kayo, pairs 0–0, 0–5 and 0–10. (a) Views 0; 0; (b) Views 0; 5 (c) Views 0, 10.



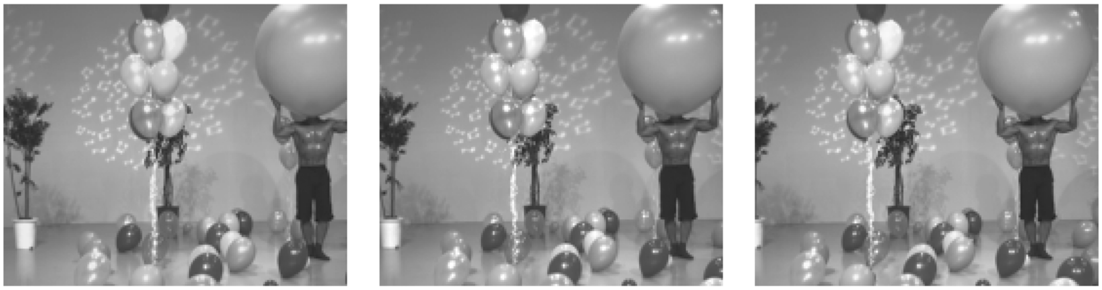
(a) Views 0, 20 (b) Views 0, 40 (c) Views 0, 80

Fig. 8. MVC efficiency as a function of time, using view 0 as a reference: sequence Akko&Kayo, pairs 0–20, 0–40 and 0–80. (a) Views 0, 20; (b) Views 0, 40; (c) Views 0, 80.



(a) View 0 (b) View 3 (c) View 6

Fig. 9. Selected camera views of Kendo sequences, horizontal displacement. (a) View 0; (b) View 3; (c) View 6.



(a) View 0 (b) View 3 (c) View 6

Fig. 10. Selected camera views of Balloons sequences, horizontal displacement. (a) View 0; (b) View 3; (c) View 6.

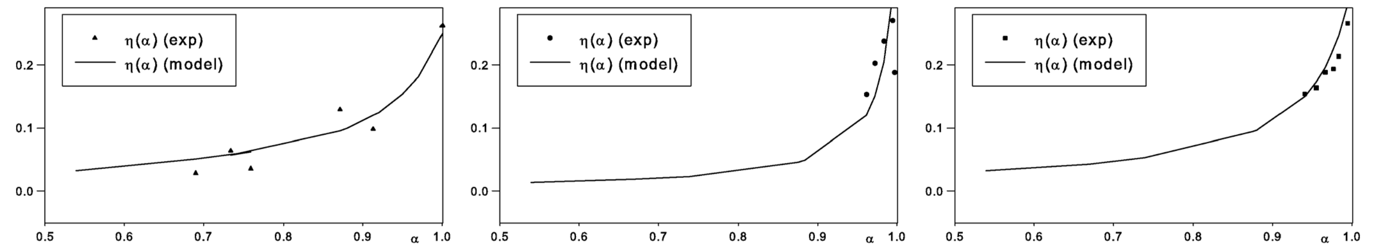
H.264/AVC efficiency $\eta(i,j)$ and of the CSA $\alpha(i,j)$. Let us now in detail carry out the analysis of the temporal evolution of the trend summarized in Table IV.

The MVC efficiency takes on values depending on the extension of the CSA between views, which, in turns, dynamically varies because of moving objects. Therefore, even for still cameras, the coding efficiency accordingly changes in time. For in-depth analysis of such dynamic behavior we need to define the MVC efficiency on a GOP time scale.

To elaborate, let us consider a sequence GOP and let us denote by $R_k^{AVC}(i)$ the cost in bits of AVC encoding the k-th frame of the GOP in the i-th view; besides, let $\Delta R_k^{MVC}(j;i)$ be the cost in bits of MVC encoding the same frame using the j-th view as reference. With this notation, the efficiency on a GOP is computed as:

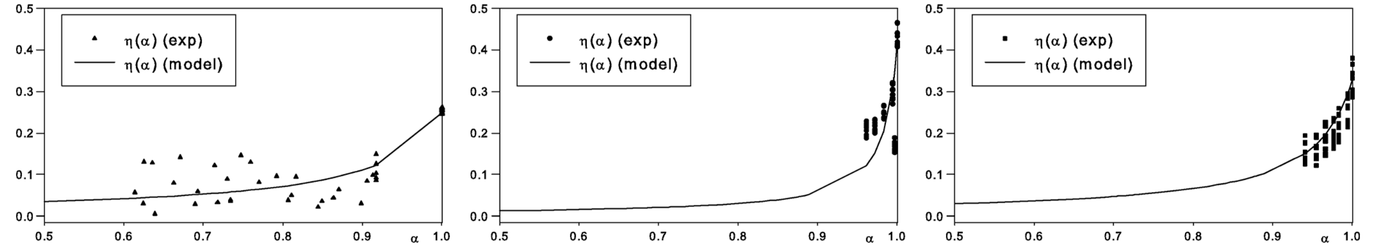
$$\overline{\eta}(i,j) = 1 - \frac{\sum_{k=1}^{M} R_k^{AVC}(i) + \sum_{k=1}^{M} \Delta R_k^{MVC}(j;i)}{\sum_{k=1}^{M} R_k^{AVC}(i) + \sum_{k=1}^{M} R_k^{AVC}(j)}, \quad (11)$$

where M is total number of frames in the GOP.



(a) Akko&Kayo (b) Balloons (c) Kendo

Fig. 11. Comparison of the efficiency of the empirical model (continuous line) with the actual efficiency measured in a single GOP for the three sequences Akko&Kayo ($\eta_{\rm MAX}(i,j) = 0.25, \epsilon = 0.08$), Balloons ($\eta_{\rm MAX}(i,j) = 0.43, \epsilon = 0.015$), Kendo ($\eta_{\rm MAX}(i,j) = 0.33, \epsilon = 0.05$). (a) Akko&Kayo; (b) Balloons; (c) Kendo.



(a) Akko&Kayo (b) Balloons (c) Kendo

Fig. 12. Comparison of the efficiency of the empirical model (continuous line) with the actual efficiency measured in several GOPs for the three sequences Akko&Kayo ($\eta_{\rm MAX}(i,j) = 0.25, \epsilon = 0.08$), Balloons ($\eta_{\rm MAX}(i,j) = 0.43, \epsilon = 0.015$), Kendo ($\eta_{\rm MAX}(i,j) = 0.33, \epsilon = 0.05$). (a) Akko&Kayo; (b) Balloons; (c) Kendo.

Let us now introduce the CSA on a GOP. In formulas, let us denote by $\bar{\alpha}(i,j)$ the average of the estimated CSA over the GOP frames, namely:

$$\overline{\alpha}(i,j) = \frac{1}{M}\sum_{k=1}^{M}\hat{\alpha}_k(i,j), \tag{12}$$

where $\hat{\alpha}_k(i,j)$ is the estimated CSA between the GOP k-th frames of the i-th and j-th views. With these positions, we can compare the pairs of values $(\eta(i,j), \bar{\alpha}(i,j))$ as observed along different GOPs of a video sequence.

We present now results pertaining to the temporal analysis of the three sequences Akko&Kayo, Balloons, Kendo. Specifically, in Figs. 12(a)–12(c) we report the scatter plots -compactly denoted as $\eta(\alpha)$ (exp) in the legends- of the MVC efficiency $\bar{\eta}(0,j)$ versus the estimated CSA $\bar{\alpha}(0,j)$ observed on several consecutive GOPs[7] for different view pairs of the three sequences. We observe that the multiview sequences span different ranges of $\bar{\alpha}(0,j)$.

Further, in Fig. 11(a)–11(c), for each of the three sequences we report the scatter plots of the measured pairs $(\bar{\eta}(0,j), \bar{\alpha}(0,j))$ as observed on a single GOP interval.

Since the encoding process depends on a large number of factor, including, but not limited, to illumination conditions, scene dynamics, moving objects and background textures, a random variability is observed when jointly encoding different pairs of frames, although they have a similar value of estimated CSA. Nonetheless, a statistical regularity is observed in the above data, which will be considered in the derivation of the empirical model.

VI. Empirical Model of MVC Efficiency

Based on the above presented results, we seek to develop an empirical model to predict the MVC efficiency $\eta(i,j)$ as a function of of the CSA $\alpha(i,j)$. The model $\eta(\alpha)$ expressing the efficiency $\eta(i,j)$ as a function of the CSA $\alpha(i,j)$ should properly take into account a few trends that, despite the erratic nature of the encoding results, stem out from the simulative study carried out until now. The trends can be summarized as follows:

- the model should account for an abrupt increase in the efficiency when CSA values approach 1;
- the model should describe a plateau of medium-to-low efficiency values for decreasing CSA;
- the model should account for the intrinsic video sequence activity, which ultimately determines the maximum MVC efficiency.

To satisfy these requirements with a compact set of parameters, we propose to use a hyperbolic model, which can be formulated as follows

$$\eta(\alpha) = \eta_{\rm MAX}\cdot\frac{\epsilon}{1-\alpha+\epsilon}. \tag{13}$$

The model in (13) depends on two parameters, namely the constant ϵ[8], which drives the hyperbole curvature, and $\eta_{\rm MAX}$. In the following, we relate the parameter $\eta_{\rm MAX}$ to the efficiency of the temporal prediction and we provide a rational to set it based on the video sequence activity.

The parameter $\eta_{\rm MAX}(i,j)$ can be set based on the encoding conditions as follows. We expect a high MVC efficiency for

[7] Specifically, in our experiments the encoded sequences correspond to a time interval of 5.9 s.

[8] Specifically, ϵ ranges in 0.015 $\div$ 0.08 in all the experimental results.

TABLE V
EFFICIENCY VALUES IN CASE OF $\alpha(i,j)a = 1$

Sequence	Activity	μ	$\eta_{\mathrm{MAX}}(i,j)$	Maximum $\overline{\eta}(i,j)$ (exp)
Akko&Kayo	High	1	0.25	0.261
Kendo	Medium	0.5	0.33	0.343
Balloons	Low	0.15	0.43	0.465

frames encoded without temporal motion compensation, namely Intra frames and anchor frames, where inter-view prediction is more beneficial. For full superposition of the to-be-encoded frames associated with different views (i.e., CSA $= 1$), the cost of the inter-view predicted anchor frame is expected to be very low with respect to single view encoding. With decreasing values of CSA, the inter-view prediction efficiency decays; on the other hand, the efficiency of temporal prediction, which is related to the sequence content only, does not decrease.

To take into account these observations, we compute $\eta_{\mathrm{MAX}}(i,j)$ by approximating the cost of the secondary view as follows:

$$\Delta R_k^{MVC}(j;i) \simeq \begin{cases} R^{AVC}(i), & k \neq 1 \\ 0, & k = 1. \end{cases} \tag{14}$$

In the limits of $\alpha(i,j) = 1$, that is for highly correlated views, we can also assume that the costs of AVC encoding the i-th and j-th views become comparable, i.e.:

$$R_1^{AVC}(i) + \sum_{k=2}^{M} R_k^{AVC}(i) \approx R_1^{AVC}(j) + \sum_{k=2}^{M} R_k^{AVC}(j) \tag{15}$$

so we can express $\overline{\eta}(i,j)$ in (11) as:

$$\overline{\eta}(i,j) \approx 1 - \frac{R_1^{AVC}(i) + 2\cdot\sum_{k=2}^{M} R_k^{AVC}(i)}{2\cdot\left(R_1^{AVC}(i) + \sum_{k=2}^{M} R_k^{AVC}(i)\right)} \tag{16}$$

Let us now denote by μ the ratio between the overall inter-coded frames cost versus the cost of the intra frame, namely:

$$\mu = \frac{\sum_{k=2}^{M} R_k^{AVC}(i)}{R_1^{AVC}(i)} \tag{17}$$

The ratio μ randomly varies depending on several factors but it is certainly related to the video content activity. Specifically, it tends to zero[9] for a perfectly still scene, and it assumes increasing values for increasingly dynamic scenes. With these approximations, we obtain

$$\eta_{\mathrm{MAX}}(i,j) = \lim_{\alpha(i,j)\to 1} \overline{\eta}(i,j) = 1 - \frac{1+2\mu}{2+2\mu} \tag{18}$$

The expression in (18) summarizes the joint effect of temporal and inter-view prediction, and allows us to express the expected relative MVC encoding efficiency performance based on the video sequence activity.

In Table V we report a comparison of the maximum value $\eta_{\mathrm{MAX}}(i,j)$ of the relative MVC efficiency predicted according to (18), for suitable setting of the activity parameter μ, and the same value the relative MVC efficiency $\overline{\eta}(i,j)$ experimentally measured on the three considered test sequences. From Table V, we observe that the maximum measured efficiency is quite close to the value $\eta_{\mathrm{MAX}}(i,j)$ computed as in (18).

TABLE VI
CLUSTER EFFICIENCY IN DIFFERENT SIMULATIONS

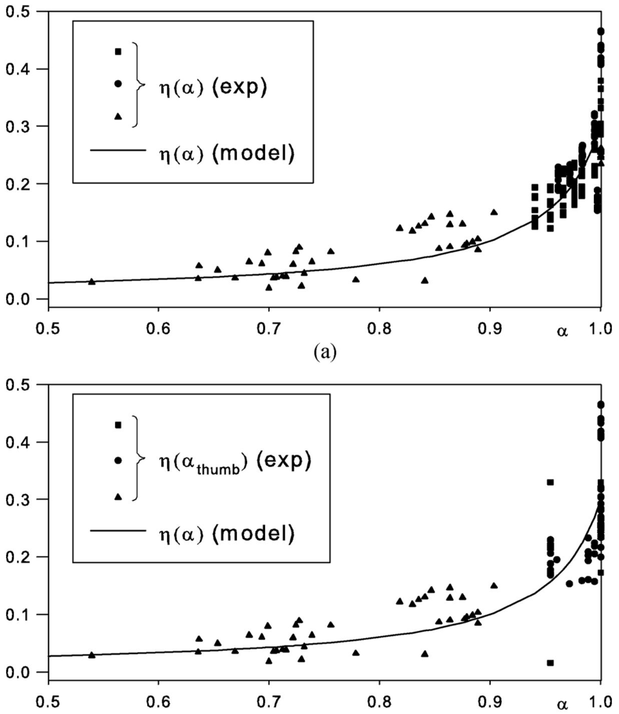
	Conf1	Conf2	Conf3
Mean efficiency	0.21	0.22	0.20
Max efficiency	0.33	0.40	0.33
Min efficiency	0.10	0.10	0.13
Number of clusters with 1 node	2	2	1

The MVC efficiency model introduced in (13) is now compared with the scatter plots summarizing the experimental results relating $\eta(i,j)$ to $\alpha(i,j)$. Specifically, in the Figs. 12(a)–12(c), we plot the MVC efficiency $\eta(\alpha)$ (continuous line) evaluated in accordance to the empirical model in (13). For all the sequences, within the limits of the random variability expected from the encoding results, we can appreciate that the model captures the relationship between $\overline{\eta}(0,j)$ and $\bar{\alpha}(0,j)$ observed on different GOPs. Also on the single GOP time scale, reported in Figs. 11(a)–11(c), the model matches the experimental results, within the limits of random fluctuations observed when a video sequence is coded at constant video quality.

To summarize the above results, in Fig. 13(a) we show the empirical model $\eta(\alpha)$ in (13) (continuous line) for parameter settings $\eta_{\mathrm{MAX}}(i,j) = 0.3$ and $\epsilon = 0.05$, together with the three scatter plots of the MVC efficiency values $\overline{\eta}(0,j)$ versus the corresponding estimated average value $\bar{\alpha}(0,j)$ corresponding to the Akko&Kayo (triangle), Balloons (circle) and Kendo (square) sequences. From Fig. 13(a) we recognize that, in spite of the differences among the sequences, the hyperbolic common model $\eta(\alpha)$ captures well the variations of $\overline{\eta}(0,j)$ versus $\bar{\alpha}(0,j)$ in all the experiments.

Finally, we report simulation results assessing the computational feasibility of estimating the CSA in WMSNs. In WMSN applications, the CSA needs to be estimated after signaling of suitable information among the nodes. Besides, this information shall be periodically updated to track the scene changes in the camera FoVs. We assessed the performance of the proposed model when the CSA is not estimated on the original frames but on subsampled version of the frames, namely image thumbnails. Thumbnails can be more easily exchanged among WMSN nodes, and reduce the computational complexity of verifying the view similarity by making it feasible real-time. Let us consider the case in which the CSA $\overline{\alpha}_{\mathrm{th}}(0,j)$ is estimated through the cross correlation based estimator on thumbnails of size 22 × 18 of the frames belonging to different views. In Fig. 13(b), we report a plot of the empirical model efficiency $\eta(\alpha)$ together with the scatter plot of the observed pairs $(\overline{\eta}(0,j), \overline{\alpha}_{\mathrm{th}}(0,j))$. We observe that, in the limit of the random variations due not only to the encoding efficiency but also to the coarser estimation stage, the model in (13) still captures the relationship between the efficiency and the CSA as estimated on thumbnails. We observe that if a signaling period of T s is considered, transmission of uncompressed thumbnails data requires a bandwidth overhead of $22 \times 18 \times l/T$ bit/s, being l the luminance depth. For $T = 10$ s and $l = 8$ bits this corresponds to an overhead of 316 bit/s. We observe that this value can be considered a maximum overhead achieved for signaling between nodes that are not performing MVC. On the contrary, when MVC is performed, each node is able to check the efficiency of MVC from the available reference view data, without the additional burden of signaling overhead.

[9] For any real video encoder, μ cannot take on the zero value due to unavoidable syntax overhead.



(a)

(b)

Fig. 13. Synoptic comparison of the efficiency of the empirical model $\eta(\alpha)$, ($\eta_{\mathrm{MAX}}(i,j) = 0.4, \epsilon = 0.05$), with the scatter plots of the efficiency $\overline{\eta}(i,j)\}$ vs estimated CSA $\overline{\alpha}_{\mathrm{th}}(0,j)$ (a), and vs thumbnail estimated CSA $\overline{\alpha}_{\mathrm{th}}(0,j)\}$ (b), as observed in several GOPs for the different views in the sequences Akko&Kayo (triangle), Balloons (circle), Kendo (square).)

In principle, the model we presented can be extended to predict the efficiency of MVC on more than two views. Nonetheless, given our focus on WMSN, we recognize that there is a trade off between performance gain, system complexity, and signaling overhead. As a consequence, practical considerations on the need of reference view availability suggest to limit the adoption of MVC two pairs of views.

Finally, as far as the modern multiview encoders are concerned, all the up-to-date standard multi-view encoders, from H.264 MVC to the upcoming HEVC are basically hybrid video encoders adopting motion compensation as well as disparity compensation techniques. Thereby, the trends herein obtained using the JMVC codec can be considered representative of a wide variety of multi view video encoders.

To recap, the adoption of MVC between adjacent camera nodes may be beneficial only under certain similarity conditions between camera views. Generally speaking, the relative efficiency of joint view coding is related to time-variant inter-view similarity, which depends not only on camera locations, but also on several characteristics of the framed scene, such as activity, moving objects to camera distances, occlusions. Despite the complexity of describing the real scene, the model introduced in (13) captures the relationship between MVC efficiency and CSA, being this latter computed through a low-complexity correlation based estimator. Since the model in (13) provides a tool for predicting the relative MVC efficiency given the CSA between camera views, it can be used on sensor nodes to decide and switch between MVC versus AVC modes. Careful selection of the most effective coding mode is especially needed in WMSN, which are well known to be resource-limited, dynamic and computationally constrained in nature; hence, a compact criterion for the adoption of inter-node MVC comes in handy in several network design problems. Examples of application of the MVC efficiency model are given in the followings.

VII. Wireless Multimedia Networking Through CSA

We now conclude our paper by presenting two case studies (single hop clustering scheme and multi hop aggregation toward the sink) that show how the proposed model can be applied to WMSN to leverage the potential gains of MVC. Consider a WMSN with one sink and N sensor nodes equipped with video cameras uniformly distributed in a square sensor field. The network of sensors is modeled as an undirected graph $G(\mathcal{N}, \mathcal{E})$, where $\mathcal{N}$ represents the set of vertexes (sensor nodes) $n_1, n_2, \ldots, n_N$, with $N = |\mathcal{N}|$, and $\mathcal{E}$ is the set of edges between nodes. An edge exists between two nodes if they fall within one another's transmission range (e.g., in the order of 100 m for an IEEE 802.15.4 WMSN). Each edge of the network graph between node i and j is associated with a weight equal to $\alpha(i,j)$ (4). Depending on the network topology and camera orientations, neighboring nodes may acquire overlapping portions of the same scene, leading to correlated views. Without loss of generality we randomly generated the $\alpha(i,j)$ of the edges in the range [0–1] to model the fact that, due to the orientation of the cameras, some close nodes may not have overlapping sensed area, while nodes at higher distance (within transmission range) may have overlapped FoVs. We assume that all views from all sensor nodes are transmitted to the sink. Each node can send to the sink video encoded either in AVC or in MVC mode. In the AVC mode, the i-th node generates video at a rate denoted as $r^{AVC}(i)$, i.e., the bit rate for the single-view encoding of the scene acquired by camera i. In the MVC mode the j-th node generates a rate $r^{MVC}(i,j)$ depending on the CSA with the i-th node. In this analysis we assumed that all the $r^{AVC}(i)$ have the same value $\overline{R}^{AVC}$.

The effects of CSA is analyzed in two different case studies:
1) a single hop topology;
2) a multi-hop topology.

In the first scenario, we assume that all nodes can communicate via a single-hop to the sink. A given number m of nodes at any given time become cluster-heads (e.g., this may happen because their cameras detect a target). The resulting topology will then include multiple clusters sending their videos to a common sink as in the case of tracking the same object in a wireless camera network [21]. In this case it may be desirable that nodes that observe the same event in the restricted range of the cluster head encode their views in MVC mode if a high coding gain is expected. This would facilitate in-network processing and reduce the wireless bandwidth needed to transmit the views to the sink. The single hop clustering scheme has been introduced in [8]. We report here the main results of this study. Thanks to these encouraging results we studied also the application of the proposed approach to a multi-hop case where we considered a network with multi-hop paths between sensors and sink. A challenging problem in WMSN is to identify optimal multi-hop paths from sensors to sink [22]. We show, how, by building multi-hop paths based on the $\alpha(i,j)$ parameter introduced in (4), MVC may provide significant performance gains with respect to AVC in terms of bit rate; therefore leading to substantial capacity savings.

A. Single-Hop Case: Performance Analysis

We consider a clustered topology, with a set M of cluster-heads ($m = |M|$). Without loss of generality, we randomly select the m cluster heads. The role of the cluster head is to enable nodes in the cluster to encode their views in MVC mode. To form the clusters we then consider the following scheme:

1) each of the m nodes, once active, broadcasts an image, denoted as thumbnail T_i with $1 \leq i \leq m$, used by the other nodes to compute the CSA with node i[10] each node j (with $j \neq i$ and $j \in M$) receiving the thumbnail T_i computes the $\hat{\alpha}(i,j)$;
2) each node j selects one cluster-head in accordance with the following criteria:
   - $\hat{\alpha}(i,j) < \alpha_{th}$, do not select node i as cluster-head;
   - $\hat{\alpha}(i,j) \geq \alpha_{th}$ select as cluster-head the node i with $\max_{1 \leq i \leq m} T_i$

where α_{th} is a threshold set in our correlation model. We consider a thumbnail of size 22 × 18. As a reference, if a signaling interval of T s is used, transmission of uncompressed thumbnail data would require a bandwidth overhead of $22 \times 18 \times l/T$ bit/s, with l being the luminance depth. For $T = 10$ s and $l = 8$ bits this corresponds to an overhead of 316 bit/s.

Based on (10), the overall rate of the cluster i to send all views of the k nodes in the cluster to the sink is

$$R_{i,tot}^{MVC} = \overline{R}^{AVC} + \overline{R}^{AVC} \sum_{j=1, j \neq i}^{k} \left(1 - 2\eta(i,j)\right). \qquad (19)$$

It can be observed that this rate depends on the $\eta(i,j)$ that are directly related to the $\alpha(j,i)$. In the following numerical analysis the $\eta(i,j)$ values are derived form the model curve reported in Fig. 13(a). On the contrary, the total rate in the case of single-view transmissions in the i-th cluster is

$$R_{i,tot}^{AVC} = \sum_{j=1}^{k} r_j^{AVC} = k \cdot \overline{R}^{AVC}. \qquad (20)$$

We analyzed three different WMSN configurations: Conf1 with $N = 50$ and $\overline{R}^{AVC} = 80$ kbit/s; Conf2 with $N = 70$ and $\overline{R}^{AVC} = 80$ kbit/s; Conf3 with $N = 90$ and $\overline{R}^{AVC} = 120$ kbit/s. In each network, we set $m = 10$. Each cluster includes at most N/m nodes and at least 1 node. In some cases, the cluster-head was not selected by any neighbors because the correlation between the views was too low ($\alpha < \alpha_{th}$). We set an $\alpha_{th} = 0.5$ and measured the efficiency as $1 - R_{i,tot}^{MVC}/R_{i,tot}^{AVC}$ and the mean, min and max values are reported in Table VI. We observe that the maximum efficiency achieves high values (33%–40%). This indicates that significant advantages can be obtained by using this approach in clustering schemes.

As a second set of experiments, we generated random networks of different sizes ($n = 2, \ldots 70$) and imposed that the cluster size be lower or equal than 3. This is to test if MVC performance gains persist with small cluster size. We considered $\alpha_{th} = 0.5$. Fig. 14 shows the MVC and AVC total rate generated by the network nodes as a function of the network size. The sum of the total rate of each cluster increases as the number of nodes increases and the AVC curve is always above the MVC curve. The rate performance gain offered by MVC can also be observed in Fig. 15(a), which depicts the total load of a network of 50 nodes where α_{th} is varied. In this case, the number of cluster-heads is assumed to be 10. The number of clusters varies and we observe that the total bit rate generated with MVC is always lower than with AVC but it increases as the α_{th} increases. Therefore, there is a tradeoff in selecting α_{th}: a high threshold allows producing a low overall cluster bit rate. However, at the same time a $\alpha_{th} \simeq 1$ leads to network partitioning with isolation of cluster-heads that cannot find any feasible connections. Consequently, the rate is the same as in AVC since clusters are composed of one member only.

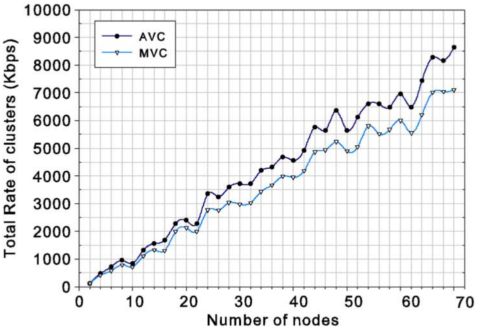
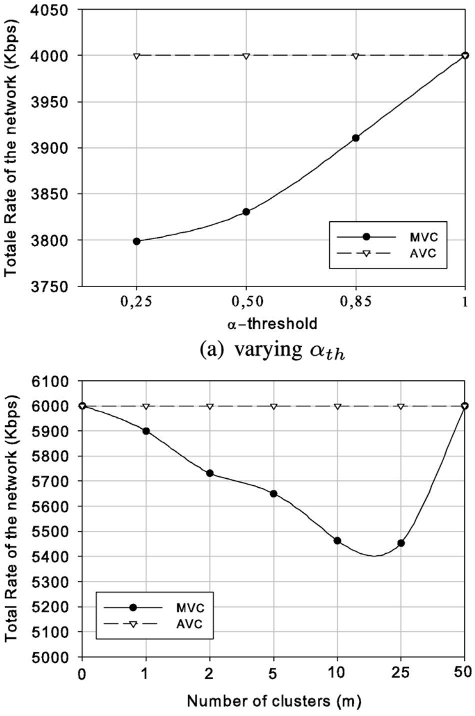


Fig. 14. Total AVC and MVC rates in case of $\overline{R}^{AVC} = 120$ kbit/s and $\alpha_{th} = 0.5$.



(a) varying α_{th}

(b) varying cluster sizes

Fig. 15. Total network rate for different α_{th} ($\overline{R}^{AVC} = 80$ kbit/s and $n = 50$) and cluster sizes ($\overline{R}^{AVC} = 120$ kbit/s and $\alpha_{th} = 0.5$). (a) varying α_{th}; (b) varying cluster sizes.

[10] Thumbnails are images at a low resolution entailing low bandwidth and allowing the $\hat{\alpha}(i,j)$ computation;

Then, we assess the role of the cluster size with the considered clustering scheme. We focus on a single network of 50 nodes randomly placed in a given area and with a random selection of the edge weights α. We set $\overline{R}^{AVC} = 120$ kbit/s and $\alpha_{th} = 0.5$. Fig. 15(b) shows how the rate varies as a function of the number of clusters. In case of AVC, the cluster size does not affect the rate, which is constant and equal to $50 \cdot 120$ kbit/s $= 6000$ kbit/s. By adopting MVC instead, we observe that the network load decreases as the number of clusters increases and reaches a minimum for a certain cluster size $(15 \leq m \leq 25)$. When the number of cluster-heads increases, the number of single-view coded videos increases, since each cluster-head sends its AVC version of the view. Finally, when the number of clusters is equal to the number of nodes $(m = 50)$ every node applies single-view coding. From Fig. 15(a), we observe that the total rate in the network can be optimized as a function of the number of clusters. The optimal value will depend on the sensor spatial distribution as well as on the transmission range. These issues are left for future investigations.

B. Multi-Hop Case: Performance Analysis

We considered N nodes, randomly distributed in a square area of $d \times d$ km^2 and we generated the initial network graph by setting a transmission range equal to r_{TX}. A node is randomly selected to act as sink, and indicated as s. Then, we randomly assigned to the l-th network link a weight α_l, representing the CSA between the pair of nodes constituting the link. Let us now consider a multi-hop path optimization scheme as follows:

1) for the i-th node, $i = 0, \dots N-1$, compute the set $\mathcal{P}^{(i)}$ of all the possible paths from i to s;to the k^{th} path in $\mathcal{P}^{(i)}$, let us say $p_k^{(i)}$, assign the utility function

$$U_k^{(i)} = \frac{1}{n_k^{(i)}} \sum_{l \in p_k^{(i)}} \alpha_l, \tag{21}$$

where $n_k^{(i)}$ represents the number of hops of $p_k^{(i)}$, and the index l spans the network links included in $p_k^{(i)}$;

2) in each set $\mathcal{P}_i$, choose the optimal path $p_{\mathrm{OP}}^{(i)}$ as the path providing the highest utility function

$$U_{\mathrm{OP}}^{(i)} = \max_{k \in \mathcal{P}^{(i)}} U_k^{(i)}. \tag{22}$$

We refer to the network topology obtained by superimposing all the optimal paths $p_{\mathrm{OP}}^{(i)}, i = 0, \dots N-1$ from each network node to the sink s as *MVC Optimal Path (MVC-OP)*.

As in the single-hop case, we can compute the total bit rate needed for nodes of a path to transmit their videos to the sink. Let us then consider a generic path p composed of k nodes and the sink (k hops). In case of MVC, the leaf node of the path p, let us say node 1, sends its own video to node 2 by means of an AVC-encoded bit stream at a rate $r^{AVC}(1) = \overline{R}^{AVC}$; in turn, node 2 sends the received AVC bit stream of the reference view and its own MVC encoded bit stream $\Delta r^{MVC}(2;1)$ to node 3. The latter sends the received AVC bit stream for the reference view, and the MVC bit stream $\Delta r^{MVC}(2;1)$ as well as its own MVC bit stream $\Delta r^{MVC}(3;2)$ to node 4, and so on until the sink is reached.

As a consequence, the overall bit rate $R_{p,tot}^{MVC}$ for the generic path p can be derived as

$$\begin{aligned} R_{p,tot}^{MVC} &= k \cdot r^{AVC}(1) + (k-1) \cdot \Delta r^{MVC}(2;1) \\ &\quad + (k-2) \cdot \Delta r^{MVC}(3;2) \dots + \Delta r^{MVC}(k;k-1) \\ &= k \cdot \overline{R}^{AVC} + (k-1) \cdot \overline{R}^{AVC} (1 - 2\eta(1,2)) \\ &\quad + (k-2) \cdot \overline{R}^{AVC} \cdot (1 - 2\eta(2,3)) \dots \\ &\quad + \overline{R}^{AVC} \cdot (1 - 2\eta(k-1,k)) . \end{aligned} \tag{23}$$

Then, we can express the total rate of the path p using MVC as

$$\begin{aligned} R_{p,tot}^{MVC} &= \overline{R}^{AVC} \cdot \left[\sum_{i=1}^{k} i - 2 \sum_{i=2}^{k} (k-i+1) \eta(i-1,i) \right] \\ &= \overline{R}^{AVC} \cdot \left[(k+1) \cdot k/2 - 2 \sum_{i=2}^{k} (k-i+1) \eta(i-1,i) \right]. \end{aligned} \tag{24}$$

Again the MVC overall rate depends on $\eta(i-1;i)$, which in turn depends on $\alpha(i-1,i)$.

Instead, in case of AVC, each node sends to its successor the AVC bit streams received from its predecessors and its own AVC bit stream. For a generic path p composed of k nodes and the sink (k hops), the overall bit rate spent to send the views to the sink using AVC is

$$\begin{aligned} R_{p,tot}^{AVC} &= k \cdot r^{AVC}(1) + (k-1) \cdot r^{AVC}(2) \\ &\quad + (k-2) \cdot r^{AVC}(3) + \dots + r^{AVC}(k) \\ &= \overline{R}^{AVC} .k(k+1)/2. \end{aligned} \tag{25}$$

We present an example by considering 25 nodes. Figs. 16(a) and 16(b) show the initial WMSN topology and the corresponding MVC-OP. The efficiency behavior is derived from the model of Section VI.

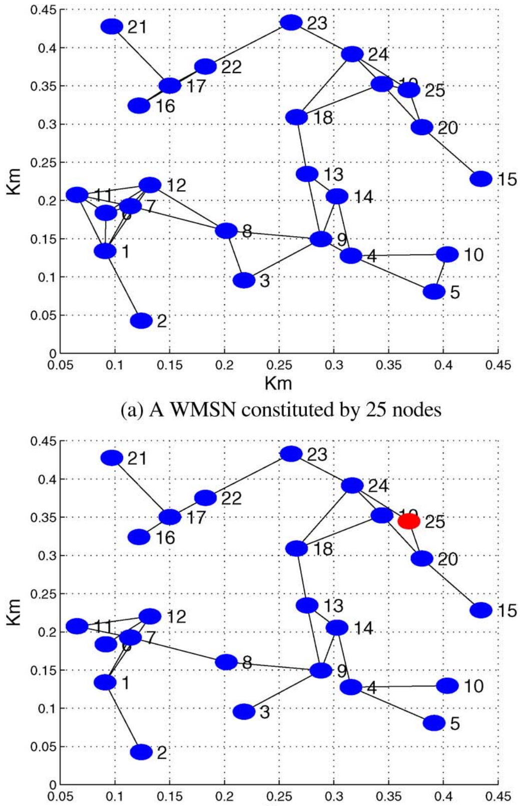
Fig. 17 presents a comparison between MVC and AVC bit rates for selected optimal paths in the MVC-OP in Fig. 16(b). It can be clearly appreciated from Fig. 17 that the MVC bit rate is always lower than the AVC rate. We also measured the path efficiency as $1 - R_{p,tot}^{MVC} / R_{p,tot}^{AVC}$, which is reported in Table VII for selected paths. The path efficiency is always around 20%–30%, with further reductions when the path is very short only. For example, the minimum value of efficiency is obtained in Fig. 17 for the shortest route (15-11-12).

Finally, we compare the MVC and AVC total rates with optimal paths. To compute these rates, we compute MVC-OP and sum the rates over all the paths starting from the leaf nodes toward the sink, as in (24) and (25), respectively, and denote these as R_{tot}^{MVC} and R_{tot}^{AVC}. In Fig. 18(a), we show R_{tot}^{AVC} and R_{tot}^{MVC} versus the number N of network nodes, randomly distributed[11] over an area of 1 km × 1 km.

Each plot point represents the sum over all the leaf network paths, averaged over 10 simulations, using $r_{TX} = 0.4$ km. In turn, in Fig. 18(b), we present the rates R_{tot}^{AVC} and R_{tot}^{MVC} versus the transmission radius r_{TX}, for $N = 50$.

In both cases, we can appreciate the bandwidth gain provided by MVC. It is worth pointing out that such gain is achieved through careful selection of the network paths along which

[11] The value of α_l is drawn in the range [0.25,1] for nodes that are within the transmission range and was set to 0 for non-connected nodes.



(a) A WMSN constituted by 25 nodes

(b) The resulting MVC-OP

Fig. 16. Initial network topology and the resulting MVC-OP one, case of 25 nodes. (a) A WMSN constituted by 25 nodes; (b) The resulting MVC-OP.

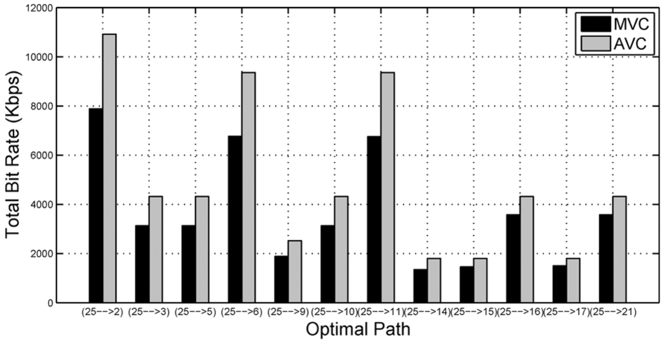
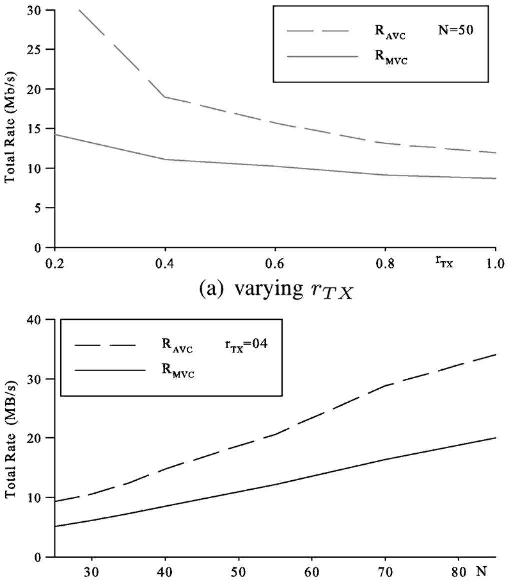


Fig. 17. Comparison between MVC and AVC bit rates of some optimal paths of the VC-OP in Fig. 16(b).

TABLE VII
PATH EFFICIENCY FOR SOME OPTIMAL PATHS FOR THE MVC-OP OF FIG. 16(b)

Path	Path efficiency	Path	Path efficiency
(25→2)	0.28	**(25→11)**	0.28
(25→3)	0.28	**(25→14)**	0.25
(25→5)	0.28	**(25→15)**	0.19
(25→6)	0.28	**(25→16)**	0.15
(25→9)	0.25	**(25→17)**	0.17
(25→10)	0.28	**(25→21)**	0.17

MVC is beneficial. This observation paves the way for further research, aimed at designing cross-layer optimized networking schemes. In this perspective, the herein presented MVC efficiency model provides a compact tool for optimized assignment of the scarce network resources in a WMSN.



(a) varying r_{TX}

(b) varying N

Fig. 18. Comparison of R_{tot}^{MVC} (solid line) and R_{tot}^{AVC} (dashed line) averaged on 10 simulations, on leaf nodes paths of the MVC-OP versus the transmission range $r_{TX}(N = 50)$, and versus the number of network nodes $N(r_{TX} = 0.4\ \text{km})$. (a) varying r_{TX}; (b) varying N.

VIII. CONCLUSIONS

We investigated the relationship between the efficiency of Multiview Video Coding and the common sensed areas between views through video coding experiments. We developed an empirical model of the Multiview Video Coding (MVC) compression performance that can be used to identify and separate situations when MVC is beneficial from cases when its use may be detrimental. The model, whose accuracy has been assessed on different multiview video sequences, predicts the compression performance of MVC as a function of the correlation between cameras with overlapping fields of view, and accounts not only for geometrical relationships among the relative positions of different cameras, but also for various *object-related phenomena*, e.g., occlusions and motion, and for *low-level phenomena* such as variations in illumination. Finally, we showed how the model can be applied to typical scenarios in WMSN, i.e., to clustered or multi-hop topologies, and highlighted some promising results of its application in the definition of cross-layer clustering and data aggregation procedures.

REFERENCES

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

[2] S. Pudlewski, T. Melodia, and A. Prasanna, “Compressed-sensing-enabled video streaming for wireless multimedia sensor networks,” *IEEE Trans. Mobile Comput.*, vol. 11, no. 6, pp. 1060–1072, Jun. 2012.

[3] Y. Fu, Y. Guo, Y. Zhu, F. Liu, C. Song, and Z.-H. Zhou, “Multi-view video summarization,” *IEEE Trans. Multimedia*, vol. 12, no. 7, pp. 717–729, Nov. 2010.

[4] A. C. Sankaranarayanan, R. Chellappa, and R. G. Baraniuk, “Distributed sensing and processing for multi-camera networks,” *Distrib. Video Sensor Netw.*, pt. 2, pp. 85–101, 2011.

[5] A. R. Vinod Kulathumani, S. Parupati, and R. Jillela, “Collaborative face recognition using a network of embedded cameras,” *Distrib. Video Sensor Netw.*, pt. 5, pp. 373–387, 2011.

[6] T. Montserrat, J. Civit, O. Escoda, and J.-L. Landabaso, "Depth estimation based on multiview matching with depth/color segmentation and memory efficient Belief Propagation," in *Proc. 2009 16th IEEE Int. Conf. Image Processing (ICIP)*, 2009, pp. 2353–2356.

[7] A. Vetro, T. Wiegand, and G. Sullivan, "Overview of the stereo and multiview video coding extensions of the h.264/mpeg-4 avc standard," *Proc. IEEE*, vol. 99, no. 4, pp. 626–642, Apr. 2011.

[8] S. Colonnese, F. Cuomo, and T. Melodia, "Leveraging multiview video coding in clustered multimedia sensor networks," in *Proc. IEEE Globecom 2012*, Dec. 2012, pp. 1–6.

[9] H. Ma and Y. Liu, "Correlation based video processing in video sensor networks," in *Proc. 2005 Int. Conf. Wireless Networks, Communications and Mobile Computing*, Jun. 2005, vol. 2, pp. 987–992.

[10] P. Wang, R. Dai, and I. F. Akyildiz, "A spatial correlation-based image compression framework for wireless multimedia sensor networks," *IEEE Trans. Multimedia*, vol. 13, no. 2, pp. 388–401, Apr. 2011.

[11] V. Thirumalai and P. Frossard, "Correlation estimation from compressed images," *J. Visual Commun. Image Represent., Special Issue on Recent Advances on Analysis and Processing for Distributed Video Systems*, vol. 24, no. 6, pp. 649–660, 2013.

[12] R. Arora and C. R. Dyer, "Projective joint invariants for matching curves in camera networks," in *Distributed Video Sensor Networks*, B. Bhanu, C. V. Ravishankar, A. K. Roy-Chowdhury, H. Aghajan, and D. Terzopoulos, Eds. London, U.K.: Springer, 2011, pp. 41–54.

[13] J.-N. Hwang and V. Gau, "Tracking of multiple objects over camera networks with overlapping and non-overlapping views," in *Distributed Video Sensor Networks*, B. Bhanu, C. V. Ravishankar, A. K. Roy-Chowdhury, H. Aghajan, and D. Terzopoulos, Eds. London, U.K.: Springer, 2011, pp. 103–117.

[14] J. Shen and Z. Cheng, "Personalized video similarity measure," *Multimedia Syst.*, pp. 1–13, 2010.

[15] A. Neri, S. Colonnese, G. Russo, and P. Talone, "Automatic moving object and background separation," *Signal Process.*, vol. 66, no. 2, pp. 219–232, 1998.

[16] H. Li and K. N. Ngan, "Image/video segmentation: Current status, trends, and challenges," in *Video Segmentation and Its Applications*, K. N. Ngan and H. Li, Eds. New York, NY, USA: Springer, 2011, pp. 1–23.

[17] Y. Chen, Y.-K. Wang, K. Ugur, M. Hannuksela, J. Lainema, and M. Gabbouj, "The emerging MVC standard for 3D video services," *EURASIP J. Adv. Signal Process.*, pp. 1–13, 2009.

[18] H. Kimata, A. Smolic, P. Pandit, A. Vetro, and Y. Chen, AHG Report: MVC JD & JMVM Text, Software, Conformance, Joint Video Team of ISO/IEC MPEG & ITU-T VCEG, in Doc. JVT-AD005. Lausanne, Switzerland, Jan. 2009.

[19] [Online]. Available: http://www.tanimoto.nuee.nagoya-u.ac.jp/fukushima/mpegftv/

[20] P. Merkle, A. Smolic, K. Muller, and T. Wiegand, "Efficient prediction structures for multiview video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 11, pp. 1461–1473, Nov. 2007.

[21] H. Medeiros, J. Park, and A. Kak, "Distributed object tracking using a cluster-based Kalman filter in wireless camera networks," *IEEE J. Select. Topics Signal Process.*, vol. 2, pp. 448–463, 2008.

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

Stefania Colonnese (M.S. 1993, Ph.D. 1997) was born in Rome, Italy. She received the Laurea degree in electronic engineering from the Universitá "La Sapienza", magna cum laude, Rome, 1993, and the Ph.D. degree in electronic engineering from the Universitá di Roma "Roma Tre" in 1997. She has been active in the MPEG-4 standardization activity on automatic Video Segmentation. In 2001, she joined the Universitá "La Sapienza", Rome, as Assistant Professor. Her current research interests lie in the areas of signal and image processing, multiview video communications processing and networking. She is currently associate editor of the Hindawi Journal on Digital Multimedia Broadcasting (2010). She served in the TPC of IEEE/Eurasip EUVIP 2011, (Paris, July 2011) and of Compimage 2012 (Rome, September 2012). She is Student Session Chair for IEEE/Eurasip EUVIP 2013. She has been Visiting Scholar at "The State University of New York at Buffalo" (2011), Erasmus Visiting teacher at Université Paris 13 (2012).

Francesca Cuomo received her "Laurea" degree in Electrical and Electronic Engineering in 1993, magna cum laude, from the University of Rome "La Sapienza", Italy. She earned the Ph.D. degree in Information and Communications Engineering in 1998. She is Associate Professor in Telecommunication Networks at the University of Rome "La Sapienza". Her main research interests focus on: Vehicular Networks, Wireless ad-hoc and Sensor networks, Cognitive Radio Networks, Green networking. Prof. Cuomo has advised numerous master students in computer science, and has been the advisor of 8 Ph.D. students. She has authored over 80 peer-reviewed papers published in prominent international journals and conferences. She is in editorial board of the Elsevier Ad-Hoc Networks and she has served on technical program committees and as reviewer in several international conferences and journals. She served as Technical Program Committee Co-Chair for the ACM PE-WASUN 2011, 2012 and 2013 2012 "ACM International Workshop on Performance Evaluation of Wireless Ad Hoc, Sensor, and Ubiquitous Networks". She is IEEE Senior Member.

Tommaso Melodia is an Associate Professor with the Department of Electrical Engineering at the State University of New York (SUNY) at Buffalo, where he directs the Wireless Networks and Embedded Systems Laboratory. He received his Ph.D. in Electrical and Computer Engineering from the Georgia Institute of Technology in 2007. He had previously received his "Laurea" (integrated B.S. and M.S.) and Doctorate degrees in Telecommunications Engineering from the University of Rome "La Sapienza", Rome, Italy, in 2001 and 2005, respectively. He is a recipient of the National Science Foundation CAREER award, and coauthored a paper that was recognized as the Fast Breaking Paper in the field of Computer Science for February 2009 by Thomson ISI Essential Science Indicators and a paper that received an Elsevier Top Cited Paper Award. He is the Technical Program Committee Vice Chair for IEEE Globecom 2013 and the Technical Program Committee Vice Chair for Information Systems for IEEE INFOCOM 2013, and serves in the Editorial Boards of IEEE Transactions on Mobile Computing, IEEE Transactions on Wireless Communications, and Computer Networks (Elsevier), among others. His current research interests are in modeling, optimization, and experimental evaluation of wireless networks, with applications to cognitive and cooperative networking, ultrasonic intra-body area networks, multimedia sensor networks, and underwater networks.