

Correlation-Aware Packet Scheduling for Multi-Camera Streaming

Laura Toni
Univ. Calif. at San Diego
email: ltoni@ucsd.edu

Thomas Maugey
LTS4, EPFL
email: thomas.maugey@epfl.ch

Pascal Frossard
LTS4, EPFL
email: pascal.frossard@epfl.ch

Abstract—In multi-view applications, multiple cameras acquire the same scene from different perspectives, which results in correlated video streams. It becomes important to exploit this correlation at the acquisition side (i.e., in the source coding) or at the receiver side (i.e., during error-concealment). In this work, we propose a correlation-aware scheduling algorithm for multi-camera sets, in which information from all views need to be sent over a bottleneck channel to clients that decode the 3D scene captured by the cameras. Based on a novel rate-distortion model, that takes into account the correlation between sources, we propose a solution that minimizes the distortion in the scene reconstruction and adapts to temporal variations in the scene content. Simulation results show the gain of the scheduling algorithm when the correlation model is known in the optimization, compared to scheduling policies with no information about the correlation or with a priori camera selection algorithms.

I. INTRODUCTION

Advances in both interactive services and multimedia sensor networks have paved the road to multi-view video applications, in which multiple sources acquire and transmit correlated media streams, [1]. The correlation between streams has led researchers to investigate distributed source coding (DSC) schemes, where media streams are encoded independently and decoded jointly. However, even if DSC permits to reduce bandwidth requirements, high-complexity decoding schemes are generally induced, and encoders are usually based on coding with side information (SI) methods, which require some a priori information about the correlation between sources. In addition, the application of DSC to many sources is still an open problem that rapidly reaches complexity limits. In [2], a clustered coding strategy has been proposed for multimedia sensor networks. Even if the complexity of the network is reduced, the proposed DSC technique still depends on the SI.

Rather than focusing on DSC schemes, recent works investigate multi-camera scenarios with practical solutions characterized by low-complexity coding and decoding processes. In [3], rate allocation (RA) techniques have been considered for video surveillance systems, however the correlation between sources has not been directly exploited. In [4], a spatial correlation model has been proposed for *camera selection* in wireless sensor networks. When the network resources do not permit all nodes to communicate to the receiver, the proposed camera selection method selects the cameras that minimize the joint entropy of a subset of cameras.

We propose, in this paper, a correlation-aware *packet scheduling* optimization problem for a multi-camera streaming in bandwidth-limited networks. We are interested in a feasible

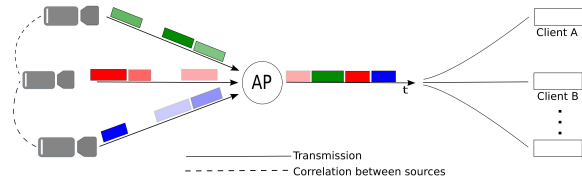


Figure 1. Multi-view streaming.

scenario in which each camera acquires part of a scene independently from the others. The encoded views are then transmitted with a correlation-aware optimized scheduling. As depicted in Fig. 1, multiple cameras acquire the same scene from different viewpoints and send it to a common access point (AP). This information is then provided to clients that might independently choose to decode (part of) the 3D scene. However, due to the limited bandwidth, only a portion of the acquired images can be scheduled for transmission to the AP. Thus, in order to improve the quality of these multi-view services, there is the need to optimize the scheduling policy. Classical RA techniques cannot solve our problem, since source correlation is usually not part of the optimization problem [3], [5]. In this work, we demonstrate the need of optimized correlation-aware scheduling policies, which are able to efficiently share resources in multi-view systems. We propose a novel rate distortion (RD) model that estimates the distortion in scene reconstruction from multiple correlated cameras. Based on the knowledge of the correlation model, we build a technique that minimizes the distortion in the scene reconstruction and adapts the transmission scheme to temporal variations of the scene content. Simulation results demonstrate that the proposed scheduling algorithm outperforms scheduling policies with no information about the correlation or a priori camera selection algorithms.

The remainder of this paper is organized as follows. In Section II, technical preliminaries are described and the RD model is introduced. The packet scheduling problem formulation is described in Section III. In Section IV, we provide simulation results and discussion, and we conclude in Section V.

II. FRAMEWORK

A. Multi-view System

We consider M cameras that acquire images and depth information from a 3D scene from different viewpoints. Based on both the geometry of the scenario and the video content, the frames acquired from the camera set might be correlated

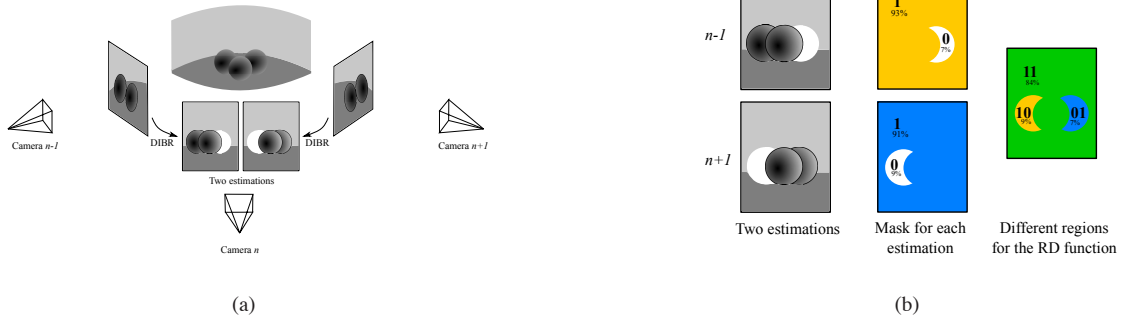


Figure 2. Example of image estimation. (a) the central camera n is estimated thanks to the two neighboring cameras $n - 1$ and $n + 1$. (b) the occlusion maps corresponding to the two estimations are merged in order to obtain 3 regions with their relative areas.

in both time and space. Each single camera acquires consecutive frames, which might be mutually correlated (*temporal correlation*), especially for static or low-motion 3D scenes. At the same time, neighboring cameras might acquire overlapping portions of the same scene, leading to correlated views (*spatial correlation*). For both the temporal and the spatial domains, two frames are correlated when their content is similar, and this correlation might help in reconstructing the scene information even if some views are missing. In the considered scenario, the frames acquired by the M correlated cameras need to be transmitted through a bottleneck channel to a common AP, that will need to provide (part of) the 3D scene to clients. Due to bandwidth constraints in realistic communication systems, network resources might not permit that all the cameras send their frames. Thus, at each transmission opportunity, it is important to accurately select which views have to be scheduled and which ones can be sacrificed (i.e., not transmitted), such that the average distortion is minimized.

We address this resource allocation problem by taking into account the level of correlation among cameras in a novel frame scheduling algorithm. This correlation information is estimated from cameras and periodically sent to the scheduler, which can exploit this knowledge in the scheduling decision. To estimate the influence that each camera can have in the reconstruction of the neighboring ones, only the information about the geometry of the camera set (e.g., cameras position) is required, and not full depth maps. We assume that each encoded view at a given time from a given camera is packetized into a data unit (DU) and stored in the camera buffer. Each data unit contains texture and depth information about the 3D scene. All the stored DUs are possible candidates for scheduling. We assume a Time Division Multiple Access (TDMA) transmission, based on which, at any TDMA slot, no more than one DU might be scheduled. Once a DU is scheduled, the channel will be busy for one or multiple time slots, until the current DU has been completely transmitted.¹ Due to a limited capacity of the buffer and also due to streaming delay constraints, the DU needs to be received before a playback deadline, denoted by T_D in order to be useful. This means that a DU acquired at the time t will expire

¹From here onwards, we assume the time axis discretized in slots (or scheduling slots) of length equal to the TDMA slot duration.

at the time $t + T_D$, after which it will not be a candidate for scheduling anymore. We also assume that each camera transmits over a lossless channel, and all the scheduled DUs are correctly received. Our goal is to propose a correlation-aware scheduling optimization, able to schedule DUs from different cameras in such a way that the overall distortion is minimized and yet the bandwidth constraint is met.

To better understand how the system takes advantage of the correlation between sources, we now describe the scene reconstruction process. At the receiver side, each transmitted frame is decoded independently. The non-transmitted images are estimated based on time and/or view interpolation algorithms using a given number of neighboring frames. More precisely, for the interpolation of a missing view n , the receiver uses images from neighboring cameras with help of depth image based rendering (DIBR) techniques (Fig. 2(a)). For the projection of a camera k , DIBR algorithms use its depth information in order to geometrically estimate the position of its pixel in the reconstructed camera n . The projected pixels are of a good precision (depending on the accuracy of the depth map [6]) but do not cover the whole estimated image, because of geometrical occlusions. As it is shown in Fig. 2 for an example of 2 neighboring cameras, one can imagine a binary mask that describes the occluded regions. While merging the different estimations coming from the projection of different neighboring cameras, we obtain different reconstructed regions in the interpolated image. In Fig. 2(b), we can observe an example where two estimations give three different regions where only one view or two views have been used for reconstruction. The principle for time error concealment is the same. The decoder uses the available past frames to reconstruct the current non-transmitted frame. The past frames cannot be used to estimate the whole missing image because of occlusions. Regions where the past frames could give some useful information are computed similarly to the view interpolation case. In particular, no motion compensation is employed and only the fixed background is considered for the scene reconstruction.

B. Rate-Distortion Model

We now translate the frame reconstruction model described above into a RD model, which will lead to the objective function of our scheduling optimization. The m -th camera,

at time t , acquires the image I_m^t , which is encoded in R_m^t bits, with $m = 1, \dots, M$. We introduce a view popularity parameter, which allows to weight the quality of each view in the mean distortion evaluation (i.e., some receivers might prefer the central camera than the lateral ones). For the reconstruction of a given frame I_m^t , if the frame is available at the decoder, the image distortion is driven by the source rate. If I_m^t is missing, it is reconstructed from neighboring frames (in time and space).

Under these assumptions, the RD model adopted in our problem formulation is the following. For each view acquired at the instant t , we fragment the frame into regions s_j^t ; $\alpha(s_j^t)$ represents the portion of the frame dedicated to the region s_j^t . The approximated RD function for the overall quality of the scene acquired at the instant t (i.e., the t -th frame) is expressed as

$$D^{(t)} = \sum_{m=1}^M \frac{1}{w_m} D_m^{(t)}(\mathbf{R}^t) \quad (1)$$

where w_m represents the relative popularity of the camera, $\mathbf{R}^t = [R_1^t R_2^t \dots R_M^t]^T$ is the size of the frame t received from the different cameras, and $D_m^{(t)}(\mathbf{R}^t)$ is the distortion of the m -th view. Thus, $D_m^{(t)}(\mathbf{R}^t)$ is given by

$$D_m^{(t)}(\mathbf{R}^t) = \sum_{s_j^t \in I_m^t} \alpha(s_j^t) d[\phi_{j,m,t} \cdot (\mathbf{R}^t)] \quad (2)$$

where $\phi_{j,m,t}$ is a mapping function, which allows to know the cameras that contribute to the reconstruction of the m -th view at time t . In case of no temporal correlation, only the spatially neighboring views can be considered for frame reconstruction. This means that $\phi_{j,m,t} = [\phi_{j,m,t}(1) \dots \phi_{j,m,t}(M)]$, where $\phi_{j,m,t}(k) = 1$ if the k -th camera contributes to the region s_j^t of the frame I_m^t . While in case of both spatial and temporal correlation, not only the spatially neighboring views, but also the past frames are used for the reconstruction of missing frames. This means that the $\phi_{j,m,t}$ matrix is given by

$$\phi_{j,m,t} = [\phi_{j,m,t}(1) \dots \phi_{j,m,t}(M) \quad \phi_{j,1,t-1}(1) \dots \phi_{j,m,t-1}(M) \quad \dots \quad \phi_{j,m,t-\rho_t}(1) \quad \dots \quad \phi_{j,m,t-\rho_t}(M)] \quad (3)$$

where ρ_t is the number of past frames that can be considered for the reconstruction of the current image. Note that, in case of both spatial and temporal correlation, the size vector needs to be updated as well, i.e., $\mathbf{R}^t = [R_1^t R_2^t \dots R_M^t R_1^{t-1} \dots R_M^{t-1} \dots R_1^{t-\rho_t} \dots R_M^{t-\rho_t}]^T$.

Finally, the distortion function in (2) can be evaluated from the general expression of the RD function of an intra-coded frame with high-rate assumption [7]:

$$d[R_I] = \mu_I \sigma_I^2 2^{-2R_I/S_f} \quad (4)$$

where R_I is the number of allocated bits, S_f is the number of pixels per frame, σ_I^2 is the spatial variance of the frame and μ_I is a constant depending on the source distribution. It is worth noting that we selected this theoretical model because it is quite simple and yet accurate; however, the scheduling algorithm presented in the following section can be extended to any other RD model.

A. Transmission Policy

We now discuss the transmission policy in our multi-camera system. At the time instant t , all the views of the frames acquired in the range $[t - T_D + 1, t]$ are possible candidates for being scheduled; they form the set \mathcal{S}_t , which has a cardinality L , with $L \in [0, T_D \cdot M]$. Let each l -th DU be characterized by its size R_l in terms of bits², its time acquisition slot $T_{A,l}$ (i.e., the instant at which the view is acquired), its expiration deadline $T_{TS,l} = T_{A,l} + T_D$, and its transmission policy $\pi_l = [\pi_l(1) \dots \pi_l(K)]$. A transmission policy π_l is a schedule according to which the DU l is allocated for transmission over a time horizon of $[t, t + K - 1]$, and $\pi_l(k) = 1$ means that the data unit l has to be sent at the transmission opportunity $t + k - 1$.

Denoting by $\boldsymbol{\pi} = [\boldsymbol{\pi}_1 \dots \boldsymbol{\pi}_L]^T$ the L -by- K matrix of transmission policies for all the candidate DUs, each $\boldsymbol{\pi}$ leads to an expected distortion $D(\boldsymbol{\lambda}^t, \boldsymbol{\pi}^t)$ evaluated over all the M views of the frames acquired in the range $[t - T_D + 1, t]$, where $\boldsymbol{\lambda}^t$ represents the state of the AP buffer at the time slot t , and $\boldsymbol{\pi}^t$ the transmission policy adopted in the range $[t, t + K - 1]$.³ Moreover, each transmission policy induces a transmission rate $\boldsymbol{\Gamma} = [\Gamma(1) \dots \Gamma(K)] = \mathbf{R} \cdot \boldsymbol{\pi}$, where $\Gamma(i)$ is the number of transmitted bits in each time slot for $i = 1, \dots, K$ and $\mathbf{R} = [R_1 \dots R_L]$. Since each DU is characterized by its own size R_l , the number of slots required for the transmission of each DU might differ from the others. Let assume that the channel capacity (in terms of *bps*) is C , then the time slots required to transmit all the DUs allocated in the range $[t, t + i - 1]$ is given by

$$\tau(i) = \left\lceil \frac{\sum_{k=1}^i \Gamma(k)}{C \cdot T_s} \right\rceil$$

where T_s is the slot duration and $\lceil x \rceil$ denotes the largest integer greater than or equal to x .

For a given scheduling policy $\boldsymbol{\pi}$, the RD function in (1), averaged over all the DUs in \mathcal{S}_t , can be expressed as

$$\begin{aligned} D[\mathbf{R}(\boldsymbol{\lambda}, \boldsymbol{\pi})] &= \sum_{\tau=t-T_D+1}^t \sum_{m=1}^M \frac{1}{w_m} D_m^{(\tau)}[\mathbf{R}^\tau(\boldsymbol{\lambda}, \boldsymbol{\pi})] \\ &= \sum_{l \in \mathcal{S}_t} \frac{1}{\tilde{w}_l} D_l[\mathbf{R}(\boldsymbol{\lambda}, \boldsymbol{\pi})] \end{aligned} \quad (5)$$

where \tilde{w}_l is the popularity of the l -th DU, and $\mathbf{R}(\boldsymbol{\lambda}, \boldsymbol{\pi})$ is the size of the all DUs in \mathcal{S}_t . Basically, the received bit budget (i.e., the number of received bits) depends on the scheduling policies both at the instant t and the past instants $[t - T_D + 1, t - 1]$, which are implicitly considered in $\boldsymbol{\lambda}$. Knowing the received bit budget, the distortion of the reconstructed DU l can be evaluated from (2) and (4), assuming the knowledge of the correlation model (i.e., the knowledge of the α and ϕ values).

²Each DU contains texture and depth information.

³From here onwards, since we refer to t as the current instant, we will omit the superscript t .

B. Problem Formulation

We aim at optimizing the DUs scheduling policy for the current instant t , assuming a time horizon of K time slots and taking into account both bandwidth and delay constraints. In particular, we seek the policy vector $\boldsymbol{\pi}$ that minimizes the expected distortion $D(\boldsymbol{\lambda}, \boldsymbol{\pi})$ for all the views in the set \mathcal{S}_t . We formulate the RD optimization with constraints as follows

$$\begin{aligned} & \min_{\boldsymbol{\pi}} \{D(\mathbf{R}(\boldsymbol{\lambda}, \boldsymbol{\pi}))\} & (6) \\ \text{subject to } & \sum_{l \in \mathcal{S}_t} \pi_{l,j} \leq 1, \quad \text{for } j = 1, \dots, K \\ & \sum_{k=1}^{\tau(i)} \Gamma(k) \leq \max\{\tau(i), K\} \cdot C \cdot T_s, \quad \text{for } i \in [1, K] \\ & T_{A,l} \geq t \wedge T_{TS,l} \leq t + K - 1, \quad \forall l \in \mathcal{S}_t \end{aligned}$$

where $\pi_{l,j}$ is the j -th element of the l -th row of the scheduling policy matrix $\boldsymbol{\pi}$. The first constraint in (6) imposes that the system cannot schedule two DUs in the same time slot. The second constraint imposes that, once the l -th DU has been allocated, it will occupy the channel for $\lceil r_l / (C \cdot T_s) \rceil$ time slots, and no other DUs will be scheduled in the meanwhile. The last constraint imposes that the set \mathcal{S}_t consists of views acquired no later than t with a playback deadline earlier than the considered time horizon. The RD optimization in (6) is the problem to be solved at time t , in order to optimize the scheduling policy for the current slot (i.e., the t -th time slot).

C. Greedy Allocation Algorithm

We consider here a simple scheduling policy that solves the problem in (6) and selects at each transmission opportunity the DU with maximal importance for the scene reconstruction. We propose a greedy algorithm that focuses on the current time instant (i.e., $K = 1$). The scheduling policy is assumed for all the time instants $t \in [0, T_{\text{acq}} \cdot N_f - 1]$, where N_f is the number of frames acquired in the video sequence and T_{acq} is the acquisition period. The transmission policy matrix for all the candidates DUs reduces to a vector $\boldsymbol{\pi} = [\pi_1 \dots \pi_L]'$, and the transmitted bit budget induced by the transmission policy $\boldsymbol{\pi}$ is $\Gamma = \mathbf{r} \cdot \boldsymbol{\pi}$. At each time slot t , the optimization problem (6), in the greedy case, becomes

$$\begin{aligned} & \min_{\boldsymbol{\pi}} \{D(\mathbf{R}(\boldsymbol{\lambda}, \boldsymbol{\pi}))\} & (7) \\ \text{subject to } & \sum_{l \in \mathcal{S}_t} \pi_l \leq 1 \\ & T_{A,l} \geq t \wedge T_{TS,l} \leq t + \tau, \quad \forall l \in \mathcal{S}_t \end{aligned}$$

where $\tau = \lceil \Gamma / (C \cdot T_s) \rceil$. As the search space is relatively small in the greedy scheduling algorithm, we consider an exhaustive search. Once the distortion minimization in (7) has been solved at the current slot t , the optimization policy is re-evaluated at the transmission opportunity $t \rightarrow t + \tau$.

IV. SIMULATION RESULTS

We now provide results for a multi-camera scenario, in which each camera acquires the scene in a resolution of 768×1024 pixel/frame with a frame rate of $F_R = 15$ fps leading to a source rate (in terms of *bps*) of each DU as

$r = R \cdot F_R$. In this work, we assume that the size of each compressed scene is constant across views and frames. If a frame rate of 15 fps is considered, results would experience the same qualitative behavior for a system with a higher frame rate. Experiments have been carried out with the ‘‘Ballet’’ and ‘‘Breakdancer’’ video sequences, consisting of $N_f = 100$ frames acquired with $T_{\text{acq}} = 1/F_R$ for each camera. Since both sequences led to similar results, for sake of brevity, we provide here performance results for the ‘‘Ballet’’ sequence only. The total number of camera ranges from 4 to 8. Denoting by ρ_s the number of spatially correlated cameras, we assume that each view is correlated to $\rho_s/2$ neighboring cameras, if available, on both the left and the right sides. The correlation in time, denoted by ρ_t , is related to the time interpolation at the decoder. In the following we will refer to time interpolation or ρ_t interchangeably. Since we are interested in reconstructing all the views of the camera set, results are provided in terms of mean PSNR, which is the PSNR averaged over all the cameras. This means that, even if some frames are decoded at high PSNR values, the average PSNR of the reconstructed scene might be in the low PSNR range in challenging transmission conditions.

We consider the case in which the source rate is constant for all the cameras (and thus for all the DUs) and denoted by r . We also impose that all the cameras (and DUs) have the same importance (i.e., $w_m = w$). We carry out experiments for different scenarios (i.e., with several channel capacities and encoding rates). Since once a DU is scheduled, it has to be completely transmitted before the allocation of a new DU, we assume that the encoding rate drives the scheduling optimization period (i.e., the time slot duration), which corresponds to r/C .

Our optimization algorithm will be evaluated for several levels of correlation known in the scheduling optimization: i) ‘‘Correlation Known’’, when the full correlation model is considered in the optimization; ii) ‘‘Space Corr Known’’, when only the spatial correlation is known; iii) ‘‘Time Corr Known’’, when only the temporal correlation is known; iv) ‘‘No corr known’’, when the scheduler ignores both ρ_s and ρ_t . In addition, we also consider a possible scenario in which the correlation model is coarsely estimated. In particular, we predict the spatial correlation and we neglect the correlation in time. The spatial correlation between the views k and k' is then modeled as

$$\rho(k, k') = \beta \cdot e^{-\Delta \cdot \mathcal{D}(k-k')} \quad (8)$$

where β is a normalization parameter, and $\mathcal{D}(k - k')$ is the distance between the two cameras. As baseline comparative algorithms, we consider a random allocation of the DUs (‘‘Baseline - RNDM’’) and a scheduling solution where cameras priorities are defined a priori based on [4] (labeled in the figures as ‘‘Baseline - Akyildiz’’).

We first study the gain that can be achieved when the correlation model is known by the scheduler. In the following figures, the PSNR of the reconstructed scene is evaluated from the rate-distortion model described in Sec. II-B. In Fig. 3, the performance of the scheduling algorithm is given as a function of the spatial correlation ρ_s for systems with 8 cameras,

Table I

PSNR FOR EACH RECONSTRUCTED VIEW FOR SYSTEMS WITH 8 CAMERAS AND $\rho_s = 8, \rho_t = 3, C = 23.5 \text{ Mbps}, r = 11.7 \text{ Mbps}$, AND $T_D = 5$.

Considered Method	View #							
	1	2	3	4	5	6	7	8
No Correlation known	24.95	25.32	26.97	27.44	26.88	26.69	25.80	25.26
Correlation known	26.19	26.26	24.13	28.08	26.23	25.18	26.87	26.18
Baseline - Akyildiz	22.28	23.07	24.87	24.52	24.64	25.84	23.84	22.55

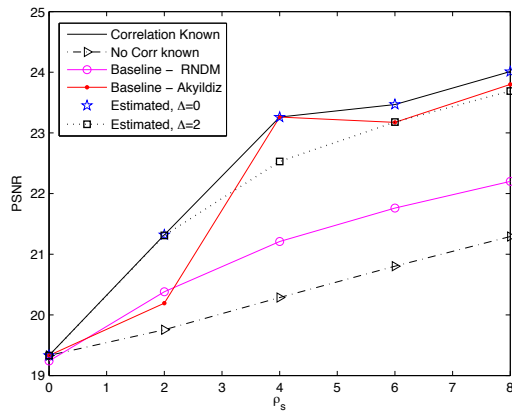


Figure 3. PSNR vs ρ_s for systems with 8 cameras, $C = 23.5 \text{ Mbps}$, $r = 11.7 \text{ Mbps}$, $T_D = 5$, and $\rho_t = 0$.

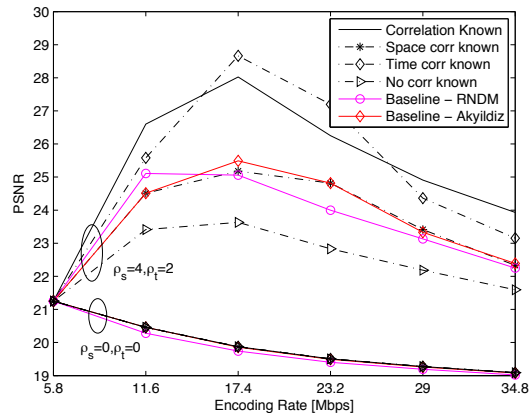


Figure 5. PSNR vs encoding rate for systems with 4 cameras $C = 23.5 \text{ Mbps}$, and $T_D = 5$.

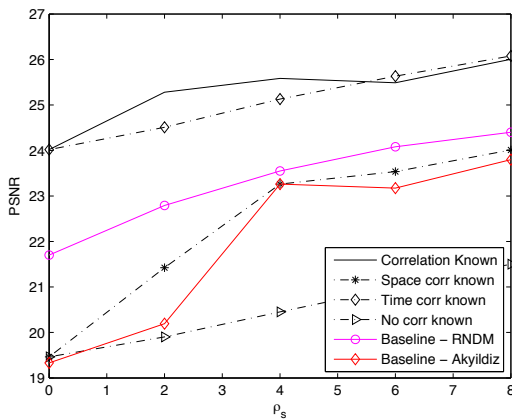


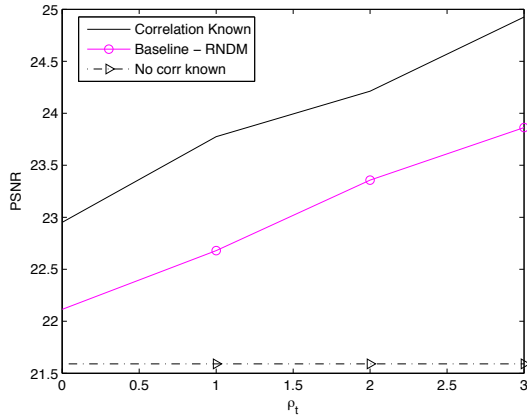
Figure 4. PSNR vs ρ_s for systems with 8 cameras, $C = 23.5 \text{ Mbps}$, $r = 11.7 \text{ Mbps}$, $\rho_t = 3$, and $T_D = 5$.

$T_D = 5, \rho_t = 0$, encoding rate constant of $r = 11.7 \text{ Mbps}$ and $C = 23.5 \text{ Mbps}$, which means that two frames can be allocated between each frame acquisition. The gain experienced by the “Correlation Known” algorithm compared to the case in which all the correlation levels are ignored is substantial and increases with ρ_s . Thus, the knowledge of the spatial correlation is able to considerably improve the efficiency of the scheduling decisions. Moreover, the proposed algorithm with a full knowledge of the correlation outperforms both baseline algorithms. This means that, even if there is no time correlation, the packet scheduling optimization leads to a better level of adaptation than the camera selection technique in [4]. It is worth noting that, by neglecting the correlation

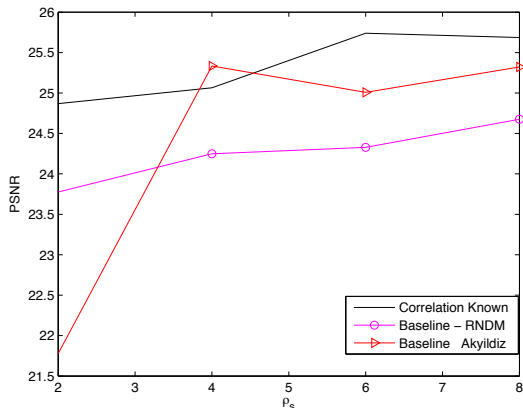
model (“No Correlation Known”) the performance becomes very bad and is even outperformed by the random allocation solution. The proposed algorithm has been tested also in the cases where the correlation is only coarsely estimated (Δ value of 0 and 2 in (8)). As expected, compared to the “Correlation Known” system, a PSNR decay occurs; however the experienced degradation is not substantial and smaller than the one occurring for the random or the “No corr known” scheduling.

In Fig. 4, the time interpolation is considered at the decoder. The PSNR is provided as a function of ρ_s for systems with 8 cameras, $C = 23.5 \text{ Mbps}$, $r = 11.7 \text{ Mbps}$ and a temporal correlation $\rho_t = 3$. It can be observed that the “Time Corr Known” curve is the closest one to the “Corr Known” case. Moreover, the greater the time correlation, the higher the gain of our scheduling optimization compared to the baseline one proposed in [4]. This is a consequence of the fact that, in the baseline algorithm there is no consideration of the time interpolation and the correlation between frames is assumed static. Note that even the gap between the “No corr known” scheduling and the random allocation increases accordingly with ρ_t , meaning that the greater the correlation level, the larger the penalty in neglecting it in the scheduling. In Table I, the PSNR for each reconstructed view is provided for the same systems of Fig. 4. It can be observed that most of the views achieve the highest PSNR when the correlation-aware scheduling is considered.

In Fig. 5, the PSNR as a function of the encoding rate is provided for a system with 4 cameras, $C = 23.5 \text{ Mbps}$, $T_D = 5$, and $K = 1$. It can be observed that the best encoding rate is a function of the available correlation level. If there is no known correlation neither in time nor space, which means no



(a) PSNR vs ρ_t for systems with 4 cameras, $C = 23.5 \text{ Mbps}$, $r = 23.5 \text{ Mbps}$, $\rho_s = 2$, and $T_D = 5$.



(b) PSNR vs ρ_s for systems with 8 cameras, $C = 23.5 \text{ Mbps}$, $r = 11.7 \text{ Mbps}$, $\rho_t = 3$, and $T_D = 5$.

Figure 6. Experimental Results for systems with 4 and 8 cameras for different encoding rates and levels of correlation.

possibility of reconstructing the missing frames, (i.e., $\rho_s = 0$, $\rho_t = 0$), it is better to reduce the encoding rate, so that there is a chance of increasing the number of DUs allocated for transmission. On the contrary, when the correlation can be exploited in both time and space for frame interpolation (i.e., $\rho_s = 4$, $\rho_t = 2$), the best encoding rate is 17 Mbps , which is not the lowest one. This means that, rather than scheduling all the frames at low rate (i.e., $r = 5.8 \text{ Mbps}$), it is better to transmit less frames but at higher rate (i.e., $r = 17 \text{ Mbps}$) and exploit the correlation for the reconstruction of the missing ones. It is worth noting that, for some configurations, knowing only the temporal correlation leads to a scheduling policy that is better than the one optimized with a full knowledge. An example given in the Fig. 5 for an encoding rate of 17 Mbps , and $\rho_s = 4$, $\rho_t = 2$. This is caused by the fact that the greedy allocation method is myopic. So, at a given instant t , the greedy algorithm schedules the DU that gives the

highest contribution to the distortion averaged over the frames acquired from the instants 1 to t . However, this scheduling not necessarily optimizes the overall distortion for all the N_f frames and all the M views. A less myopic scheduling might drive the scheduler to allocate more fairly all the views of the camera set, leading to a better distortion.

Finally, experimental results are provided in Fig. 6, where we considered a real reconstruction of the 3D scene from the received DUs. The ‘Baseline-Akyildiz’ performs better than a random scheduling most of the time, but it is in general outperformed by the proposed scheduling optimization, especially for the low ρ_s range.

V. CONCLUSIONS

We have investigated the impact of frame correlation in the scheduling algorithm of a multi-camera system. In particular, we have proposed an optimization algorithm, which optimizes the scheduling policy based on the channel capacity and both the temporal and spatial correlation among the encoded views. The proposed algorithm is able to handle the variations of the correlation among sources, adapting the transmission scheme to the level of correlation experienced by each camera. Simulation results have demonstrated the gain of the proposed method compared to classical resource allocation techniques. In addition, it is worth noting that i) when the level of correlation exists both in the time and space domains, knowing at least one of the two correlation levels leads to an improvement in the scheduling algorithm compared to the case of no correlation information; ii) the knowledge of the correlation level might help in selecting the best rate at which each camera should encode the images. In particular, the greater the level of correlation, the higher the encoding rate for each camera for optimal performance.

REFERENCES

- [1] G. Cheung, A. Ortega, and N.-M. Cheung, “Interactive streaming of stored multiview video using redundant frame structures,” *IEEE Trans. Image Processing*, vol. 20, no. 3, pp. 744–761, March 2011.
- [2] P. Wang, R. Dai, and I. 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.
- [3] H.-P. Shiang and M. van der Schaar, “Information-constrained resource allocation in multicamera wireless surveillance networks,” *IEEE Trans. on Circuits and Syst. for Video Tech.*, vol. 20, no. 4, pp. 505–517, April 2010.
- [4] R. Dai and I. 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.
- [5] L. Zhou and H.-H. Chen, “On distributed multimedia scheduling with constrained control channels,” *IEEE Trans. Multimedia*, vol. 13, no. 5, pp. 1040–1051, Oct. 2011.
- [6] K. Müller, P. Merkle, and T. Wiegand, “3D video representation using depth maps,” *Proc. IEEE*, vol. 99, pp. 643–656, April 2011.
- [7] T. A. Cover and J. A. Thomas, *Elements of Information Theory*, 1st ed. New York, NY, 10158: John Wiley & Sons, Inc., 1991.