VIEW SELECTION POLICY FOR MULTI-VIEW VIDEO DELIVERY

Jacob Chakareski

Signal Processing Laboratory - LTS4, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

ABSTRACT

We derive an optimization framework for computing a view selection policy for streaming multi-view content over a bandwidth constrained channel. The optimization allows us to determine the decisions of sending the packetized data such that the end-to-end reconstruction quality of the content is maximized, for the given bandwidth resources. Two prospective multi-view content representation formats are considered: MVC and video plus depth. For each, we formulate directed graph models that characterize the interdependencies between the data units comprising the content. For the video plus depth format, we develop a spatial error concealment strategy that reconstructs missing content at the client based on received data from other views. We design multiple techniques to solve the optimization problem of interest either exactly or approximatively, at lower complexity. In conjunction, we derive a spatial model of the reconstruction error for the multi-view content that we employ to reduce the computational requirements of the optimization. We study the performance of our framework via simulation experiments. Significant gains in terms of rate-distortion efficiency are observed over a content-agnostic reference technique.

1. INTRODUCTION

Multimedia applications that exploit content captured simultaneously via multiple cameras are increasingly becoming popular. They include 3D and free-view point TV, immersive teleconferencing, virtual worlds and gaming, and many others. To date, a considerable effort has been placed into investigating efficient methods for representation and rendering of multi-camera signals. However, the development of corresponding transmission strategies for multi-view content delivery has been notably missing.

The present paper aims at addressing the above dichotomy by developing an optimization framework for transmission policy selection of multi-camera video signals. We consider computing view transmission decisions such that the end-to-end multi-view reconstruction quality is maximized for the given bandwidth resources. The framework features directed graph models that characterize the data unit dependencies of multi-view content and a collection of techniques for computing the optimal transmission policies, either exactly or approximatively. The computational requirements of the optimization can be reduced by a spatial distortion model that we design as part of the framework. By formulating the policy selection problem within a rate-distortion framework, we derive significant benefits in transmission efficiency against a conventional reference technique.

Related work includes [1] that studies resource allocation in video surveillance networks via game-theoretic methods, however, with no spatial correlation consideration, and [2] that applies distributed source coding principles to multi-view content. In addition, [3] investigates stereoscopic content adaptation to varying network conditions, by dynamically changing encoding parameters. Similarly, [4] considers user head position tracking for dynamic allocation of encoding resources across the captured views.

This work has been supported by the Swiss NSF Ambizione Career Award PZ00P2-126416.

2. SYSTEM SETUP

2.1. Multi-view Content

The content features N views captured by their respective cameras. We consider that the content has been encoded using one of the following two approaches that exist at present. In particular, the recent Multiview Video Coding (MVC) extension of the H.264 standard [5] employs inter- and intra-view prediction simultaneously in order to maximize compression efficiency. On the other hand, the increasingly popular "video plus depth" multiview format [6] employs intra-view prediction only, in order to limit its encoding complexity, however, it encodes in addition a depth signal for each camera.

2.1.1. MVC representation

The introduced intra- and inter-view dependencies between the encoded data units (video frames) associated with the different cameras can be modeled via a directed acyclic graph. Then, a directed edge from node j to node i in the graph signifies that data unit i needs to be decoded first, in order to decode data unit j. Symbolically, this partial ordering property of the data units can be represented as $i \prec j$, i.e., i is an ancestor of j in the encoding hierarchy.

To each data unit l, we assign the following quantities. B_l represents the size of the data unit in bytes and $t_{d,l}$ its decoding deadline. Precisely, this is the time by which l needs to be decoded in order to be displayed. ΔD_l represents the corresponding reduction in reconstruction distortion of the content that the *useful* decoding of l will contribute to. Otherwise, the decoder will employ the closest decodable ancestor data unit $j \prec l$ to conceal the absence of l by $t_{d,l}$ at the receiving client, when the content is reconstructed. Thus, $\Delta D^c_{l,j}$ denotes the reduction in reconstruction error when missing data unit l is replaced by data unit j at display.

From the reconstruction distortion reduction values associated with the individual data units, we can compute the corresponding quantities at the view level. Specifically, $D_i = \sum_{l \in V_i} \Delta D_l$ represents the reconstruction distortion reduction associated with decoding view i, where V_i denotes the set of data units comprising this view. Similarly, $D^c_{i,j} = \sum_{l \in V_i} \max_{\substack{\{m \prec l\} \\ m \in V_j}} \Delta D^c_{l,m}$ is the reduction in reconstruction error for the multi-view content when missing view i is concealed by view j, at decoding. Lastly, we define D_0 to represent the reconstruction error of the content when no view is available to be decoded. For instance, D_0 can be computed as $\sum_i D_i$. Note that the data unit dependency graph also induces a partial decoding order at the view level. Hence, we can write $i \prec j$, if view i always precedes view j in the encoding hierarchy of data units.

2.1.2. Video plus depth representation

The same formalism from the earlier section carries over. To each data unit l of view v in the content graph we can again assign the quantities $B_{l,v}$, $t_{d,l}$, and $\Delta D_{l,v}$, as before. Similarly, each l will feature a set of ancestor $\{j \prec l\}$ and descendant $\{j \succ l\}$ data units in the graph. Without loss of generality, we assume that the same data unit interdependencies have been employed to encode the video and depth signals of every view. Thus, there is no need to subscript l and its delivery deadline and concealment set with the view index v.

The presence of depth information motivates us to devise a somewhat different error concealment strategy in this case. In particular, let v denote the index of the view for which data unit l is missing at

decoding and let v_1 and v_2 be two other views for which the corresponding data unit l is decodable. Then, $\Delta D^{c(\cdot,v_1,v_2)}_{l,v}$ represents the reduction in reconstruction error for the content if missing data unit l of view v is recovered at decoding via joint spatial concealment from data units l in views v_1 and v_2. The recovery procedure operates as follows. First, the image content associated with data units l in views v_1 and v_2 is mapped to view v using the corresponding depth signals. The two projections are then blended together to generate an approximation to the content associated with data unit l in view v.

An algorithmic description of the concealment procedure is included in Algorithm 1 below. The mapping $F^l_{v_i\to v}$ is carried out by a back projection of the image content of data unit l of view v_i to the 3D scene coordinates, followed by a projection to the camera location of view v [7]. The operator $\oplus$ in Line 10 signifies a pixel-wise OR combination between the two projections $F^l_{v_1\to v}$ and $F^l_{v_2\to v}$. Specifically, a pixel in the recovered content $\widehat{F}^l_v$ is declared present if it is available at least in one of the two projections. Otherwise, it is declared missing. Present pixel $p \in \widehat{F}^l_v$ takes on the single available value in one of the projections or is computed as the average of the two corresponding pixels available in both $F^l_{v_1\to v}$ and $F^l_{v_2\to v}$.

Algorithm 1 Data recovery via spatial mapping

Input: Missing data unit l of view v
1: Check for decodable spatial neighbors l in views $v_i \neq v$
2: **if** None **then**
3: Replace l by a gray-level frame; Exit
4: **else if** One (v_1) **then**
5: Map l in v_1 to v using depth signal Z_1 ($\widehat{F}^l_v = F^l_{v_1\to v}$)
6: **else**
7: Find two nearest such views ($\Longrightarrow v_1, v_2$)
8: Map l in v_1 to v using depth signal Z_1 ($\Longrightarrow F^l_{v_1\to v}$)
9: Map l in v_2 to v using depth signal Z_2 ($\Longrightarrow F^l_{v_2\to v}$)
10: Blend the two projections ($\widehat{F}^l_v = F^l_{v_1\to v} \oplus F^l_{v_2\to v}$)
11: **end if**

Finally, the view level distortion reduction values are computed as follows. We define $D_v = \sum_l \Delta D_{l,v}$, analogously to Section 2.1.1. Then, $D^c_{v,(v_1,v_2)} = \sum_l \Delta D^{c(\cdot,v_1,v_2)}_{l,v}$ represents the reduction in reconstruction distortion for the content if view v is interpolated from views v_1 and v_2, at decoding.

2.2. Transmission Policy

We consider that the server can only communicate to the client a complete subset of the N views. That is, each of the views can be either fully sent or omitted (dropped) at the server. The server needs to decide then which specific views should be selected for transmission. This is a network flow type analysis where it is assumed that the forward channel is error-free, but rate-constrained. The decision policy of the server is denoted as $\mathbf{M} = (M_1, \ldots, M_N)$, where M_i is a binary variable representing the choice to omit view i at transmission. That is, $M_i = 1$ signifies the decision not to communicate view i over the forward channel, while $M_i = 0$ denotes the opposite.

3. RECON. QUALITY & TRANS. RATE

Let $D_F(\mathbf{M})$ denote the overall reconstruction distortion of the multiview content. Then, $R_F(\mathbf{M})$ denotes the corresponding transmission rate at which the server is sending the content to the client.

3.1. MVC representation

We first formulate an expression for the expected distortion in (1). This is followed by an expression for the expected transmission rate in (2). The first product term in (1) represents the event of decoding view i at the client. The following sum term accounts for the alternative event of concealing its absence at decoding, with other decodable views comprising the content.

$$D_F(\mathbf{M}) = D_0 - \sum_{i=1}^{N} \prod_{j\preceq i} (1-M_j) D_i + \sum_{j\prec i} \left(\prod_{k\preceq j} (1-M_k) \right) M_{l\succ j} D^c_{i,j}\,, \tag{1}$$

$$R_F(\mathbf{M}) = \sum_{i=1}^{N} (1-M_i) \sum_{l\in V_i} B_l\,. \tag{2}$$

3.2. Video plus depth representation

Deriving $D_F(\mathbf{M})$ is simpler in this case, due to the absence of inter-view encoding dependencies. Note the equivalence between the expressions for the expected transmission rate in (2) and (4).

$$D_F(\mathbf{M}) = D_0 - \sum_{v=1}^{N} (1-M_v) D_v + M_v D^c_v(\mathbf{M}_{-v})\,, \tag{3}$$

$$R_F(\mathbf{M}) = \sum_{v=1}^{N} (1-M_v) \sum_{l=1}^{L} B_{l,v}\,. \tag{4}$$

In (3), $D^c_v(\mathbf{M}_{-v})$ represents the distortion reduction for the content, if missing view v is interpolated from other views at decoding. This quantity depends on the transmission policy of the server for views $j \neq v$, hence the notation $\mathbf{M}_{-v} = (M_1, \ldots, M_{v-1}, M_{v+1}, \ldots, M_N)$. We define $D^c_v(\mathbf{M}_{-v}) = D^c_{v,(v_1,v_2)}$, for $v_1 = \arg\min_{\substack{\{j\neq v\}\\ M_j=0}} ||L_j - L_v||$ and $v_2 = \arg\min_{\substack{\{j\neq v,v_1\}\\ M_j=0}} ||L_j - L_v||$. Here, L_j are the camera coordinates of view j and $\|\cdot\|$ measures the magnitude of the relative difference between two camera locations. In words, v_1 and v_2 represent the two transmitted views closest to v.

4. POLICY OPTIMIZATION

We consider that the communication channel has a finite capacity C. We are interested in solving the following optimization problem

$$\min_{\mathbf{M}} D_F(\mathbf{M})\,, \quad \text{s.t. } R_F(\mathbf{M}) \leq C.$$

We reformulate it into an unconstrained optimization using the Lagrange multiplier method [8]. Thus, we aim to minimize $J_F(\mathbf{M}) = D_F(\mathbf{M}) + \lambda R_F(\mathbf{M})$, for some Lagrange multiplier $\lambda > 0$. We design exact algorithms and greedy heuristics to carry out this task. Computing λ for a given C can be done via iterative techniques, e.g., the bisection search [9]. Alternatively, the whole lower-convex hull of solutions $\mathbf{M}^*(\lambda)$ can be computed by sweeping λ from very small to very large values. Then, we select the point $\mathbf{M}^*$ on the lower-convex hull that exhibits the largest $R_F(\mathbf{M}^*) \leq C$.

The space of prospective policies is not large. Thus, we compute the optimal solution by enumerating them and selecting the policy $\mathbf{M}^*$ that exhibits the smallest Lagrangian $J_F(\mathbf{M})$. This approach applies uniformly to both content representation formats that we consider. In contrast, the different form of $D_F(\mathbf{M})$ in (1) and (3) motivated us to develop two separate greedy optimization techniques. We employ them instead to compute an approximative solution $\widehat{\mathbf{M}}^*$.

We proceed by describing the greedy heuristic in the case of MVC encoding. First, we derive the impact that not sending view l will have on the reconstruction quality of the content, given the transmission decisions for the other views. This quantity is obtained from (1) by grouping terms

$$S_l = \sum_{j\succeq l} \prod_{\substack{k\preceq j\\ k\neq l}} (1-M_k) D_k + \sum_{j\succ l} \sum_{\substack{i\succeq l\\ i\prec j}} \prod_{\substack{k\preceq i\\ k\neq l}} (1-M_k) M_{p\succ i} D^c_{j,i} - \sum_{j\succ l} \prod_{\substack{k\preceq i\\ i\prec l}} (1-M_k) D^c_{j,i}\,. \tag{5}$$

The first summation term in (5) accounts for the impact of view l on decoding descendant view $j \succeq l$. The second accounts for the impact of view l on reconstructing missing view j via another view $i \succeq l$, for $i \prec j$. Finally, the third summation term accounts for the fact that missing view j may also be reconstructed via view $i \prec l$. Hence, S_l should be reduced by this factor (note the minus sign in front).

For every view l, we compute S_l, assuming $M_i = 0$, $i \neq j$. Then, we compute the factors $\lambda_l = S_l / \sum_{j \in V_l} B_j$ that describe the reconstruction quality impact per unit of data of view l. Finally, we sort λ_l in decreasing order and select to transmit the corresponding first $j^* < N$ views that have an aggregate data rate that does not exceed the capacity of the forward channel. An algorithmic description of the greedy heuristic is provided in Algorithm 2.

Algorithm 2 Compute policy $\widehat{\mathbf{M}}^*$ (MVC format)

1: Set $M_i = 0$, $\widehat{M}_i^* = 1$, for $i = 1, \ldots, N$
2: **for** $l = 1$ to N **do**
3: Compute S_l using (5); $\lambda_l = S_l / \sum_{j \in V_l} B_j$
4: **end for**
5: Sort λ_l in decreasing order ($\Longrightarrow \lambda_{l_j}$)
6: Find index $j^* = \arg\max_j \sum_{i=1}^{j} \sum_{m \in V_{l_i}} B_m < C$
7: **for** $j = 1$ to j^* **do**
8: Set $\widehat{M}_{l_j}^* = 0$
9: **end for**

Next, we describe the greedy optimization technique that applies to content represented in the video plus depth format. Initially, we set $\widehat{M}_j^*$ to one for every view j. Then, we select to send the view with the biggest impact on (3) per unit of transmitted data and data rate smaller than C. The procedure is iteratively repeated by selecting the next best view that contributes to the highest incremental improvement in reconstruction quality per unit of additional transmitted rate, as long as the cumulative data rate of all selected views does not exceed C. In Algorithm 3, we provide a formal description of the optimization. The symbols $\setminus$ and $\cup$ in Lines 5 and 13 signify the operations of set difference and set union, respectively.

Algorithm 3 Compute policy $\widehat{\mathbf{M}}^*$ (video plus depth format)

1: Set $\widehat{M}_j^* = 1, \forall j$, $S = \emptyset$, $V = \{1, \ldots, N\}$, $D_F^{(0)} = D_0$
2: **for** $i = 1$ to N **do**
3: $\mathbf{M} = \widehat{\mathbf{M}}^*$
4: Set $p = 0$, $X_j = 0$, $Y_j = 0$, for $j = 1, \ldots, N - |S|$
5: **for** $l \in V \setminus S$ **do**
6: Set $M_l = 0$; Compute $D_F(\mathbf{M})$ using (3); $p = p + 1$
7: Set $X_p = l$, $Y_p = (D_F^{(i-1)} - D_F(\mathbf{M})) / \sum_{j \in V_l} B_j$
8: Set $M_l = 1$
9: **end for**
10: Find $k = \arg\max_j Y_j$; Set $l = X_k$
11: **if** $\sum_j^N (1 - \widehat{M}_j^*) \sum_{m \in V_j} B_m + \sum_{m \in V_k} B_m < C$ **then**
12: Set $\widehat{M}_l^* = 0$; Compute $D_F(\widehat{\mathbf{M}}^*)$ using (3)
13: Set $D_F^{(i)} = D_F(\widehat{\mathbf{M}}^*)$; $S = S \cup l$
14: **else**
15: Exit
16: **end if**
17: **end for**

5. SPATIAL DISTORTION MODEL

The characterization of the video plus depth representation of multiview content in Section 2.1.2 considered that a missing data unit can be reconstructed from up to two spatial neighbors. This choice has been inspired by the observation that the reconstruction quality of such data only marginally improves when additional adjacent views are employed to carry out the depth signal-based recovery. We exploited this fact to limit the complexity of our formalism of the content reconstruction process and ensure its high degree of accuracy, simultaneously. We made additional observations of the performance of depth signal-based data recovery on actual content that allowed us to develop a simple model that precisely characterizes the reconstruction error as a function of the missing view's location. Concretely, we noted that the reconstruction quality of view v is inversely proportional to its spatial position relative to views v_1 and v_2. That is,

$$D^c_{v,(v_1,v_2)} = \alpha \exp\left(-\beta_1 ||L_v - L_{v_1}|| - \beta_2 ||L_v - L_{v_2}||\right) . \quad (6)$$

The parameters of the model in (6) depend on the geometry of the three-dimensional scene captured by the multiple cameras and their relative positions in the space, as well as on the characteristics of the content itself. We design another greedy heuristic that exploits the model to compute the policy $\widehat{M}^*$. It assumes that the view-level encoding quality and data rate are (approximately) uniform across all views. Then, different views can be differentiated in terms of importance solely by how well they can be utilized to recover other, missing views at the client. Using (6), the optimization computes these impact factors for each view j for which $\widehat{M}_j^* = 1$ yet. Then, it selects the highest impact factor view and sets its policy entry to zero. The procedure is repeated until the channel capacity limit is reached. Due to space limits, we omit its algorithmic description here.

6. EXPERIMENTS

In our experiments, we employ the two commonly used test sequences *Ballet* and *Breakdancer* [10]. In the case of each, the content comprises eight camera views recording the scene of interest. There are 100 frames associated with a view, captured at a rate of 15 frames per second. Each view is encoded at an average Y-PSNR video quality of around 38 dB. We set the GOP size to eight frames.

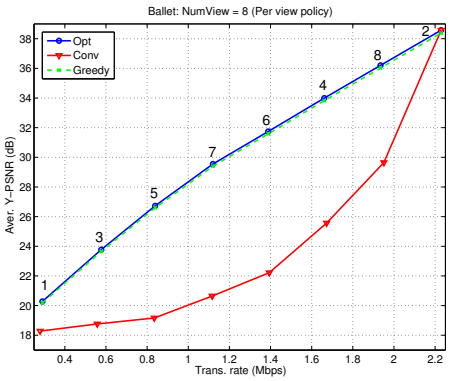

Fig. 1. Quality (dB) versus rate (Mbps) for MVC content *Ballet*. Views to send by *Opt* ($\mathbf{M}^*$) are incrementally denoted at each point.

Let *Opt* and *Greedy* denote the performances of the full optimization and its greedy approximation. In addition, we investigate the rate-distortion efficiency of a conventional system that selects views to transmit in no particular order, denoted *Conv*. In Figure 1, we show the video quality versus transmission rate (capacity) dependencies for the three systems under examination, for the MVC content *Ballet*. It can be seen that *Opt* substantially outperforms the baseline system *Conv* that is content agnostic in its operation. For instance, at a rate of 1.4 MBps, an improvement of 10 dB in video quality is registered. The gains achieved by the optimization are due to the fact that it takes into account the importance of each view for the overall reconstruction quality of the content, when making view selection decisions. The indices of the views transmitted by *Opt* are indicated in an incremental fashion next to its operating points in Figure 1. It is encouraging to see that the performance of *Greedy* is

practically identical to that of the full optimization. This is because the GOP encoding structure at the view level is simple, implying a strong view priority order that is easy to capture, and each view is encoded at roughly the same distortion-rate efficiency. Specifically, the first few view selection decisions of *Greedy* are identical to those of *Opt* and follow the relative importance of each view, as imposed by the GOP hierarchy.

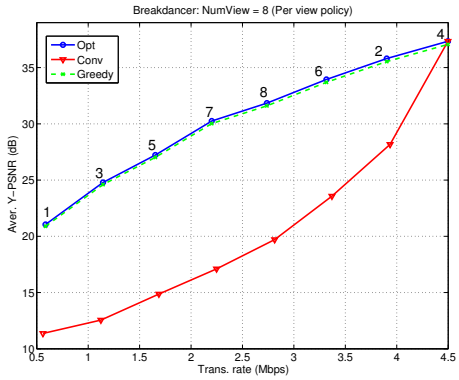

Fig. 2. Quality (dB) versus rate (Mbps) for MVC content *Breakdancer*. Views to send by *Opt* ($\mathbf{M}^*$) are incrementally denoted.

We observe an analogous outcome in the case of the MVC content *Breakdancer*. The only notable difference is that a different range of transmission capacity values needed to be used in this case due to the more dynamic nature of this content that makes it less distortion-rate efficient to encode. Still, we again observe a substantial improvement over the content-agnostic system *Conv*. For example, at a transmission rate of 2.8 Mbps, a gain of 12 dB in video quality is achieved by both *Opt* and *Greedy*. Similarly to the case of *Ballet* studied above, the performances of the latter two systems are practically overlapping over the whole range of transmission capacities considered in Figure 2. Interestingly, we can see that *Opt* chose to transmit the camera views in a different order in this case, as their indices denoted in Figure 2 indicate. This is due to the divergent rate-distortion characteristics of the video signals associated with the camera views, across the two multi-view sequences.

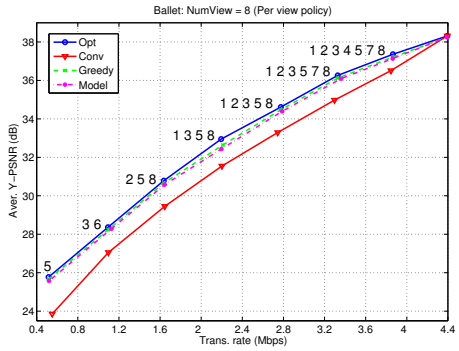

Fig. 3. Quality (dB) versus rate (Mbps) for VpD content *Ballet*. Views to send by *Opt* ($\mathbf{M}^*$) are fully denoted at each operating point.

Next, we study view selection policies in the case of video plus depth content. Due to the fact that no inter-view encoding dependencies are employed now, the policies[1] $\mathbf{M}^*$ do not exhibit any longer the property of being embedded. For the same reason, we witness less degradation in video quality in this case, when only a subset of views can be transmitted, aided further by the prospect of depth signal-based concealment from adjacent views. These observations are well noted in Figures 3 and 4 that show the performances of *Opt*, *Greedy*, and *Conv* in the cases of *Ballet* and *Breakdancer*, respectively. In addition, we examine in the figures the transmission efficiency of the view selection policies computed by the optimization, with the aid of the spatial distortion model from Section 5. The performance of this system is denoted as *Model*.

As seen from Figure 3, *Opt* again outperforms the baseline system *Conv* in the case of VpD *Ballet*, however, with a smaller margin. Specifically, the Y-PSNR video quality gains do not exceed 2 dB now. In Figure 3, we denote the view selection policies by *Opt* next to its operating points. It is encouraging to see that the performances of both *Model* and *Greedy* closely follow that of the full optimization.

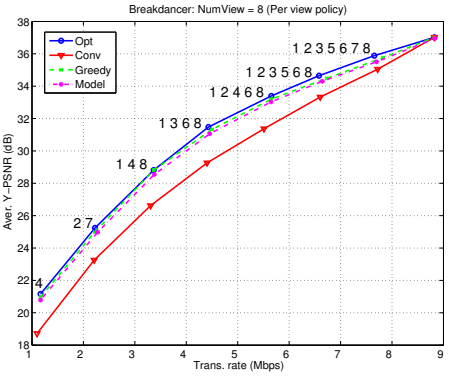

Fig. 4. Quality (dB) versus rate (Mbps) for VpD content *Breakdancer*. Views to send by *Opt* ($\mathbf{M}^*$) are fully denoted at each point.

Lastly, in Figure 4 we witness analogous performances of the four systems under comparison, in the case of the VpD content *Breakdancer*. Still, the view selection policies executed by *Opt* are different here, relative to those indicated in Figure 3. This is due to the divergent view-level rate-distortion characteristics of the two multi-view sequences, as explained earlier. Furthermore, the more dynamic nature of *Breakdancer* allows for somewhat larger video quality gains over *Conv*, as evident from Figure 4.

7. CONCLUSION

The building blocks of our framework are the models of the packetized multi-view source and the spatial error recovery procedure that is employed at the client to reconstruct missing data. Our optimization techniques exploit them in a synergistic manner such that superior rate-distortion efficiency is achieved over content-agnostic systems, when scheduling multi-view data for transmission over rate-limited channels. It is encouraging from the perspective of practical deployment that the greedy and model-based instances of our optimization still deliver significant performance gains over the reference system that we investigated.

8. REFERENCES

[1] H.-P. Shiang and M. van der Schaar, "Information-constrained resource allocation in multicamera wireless surveillance networks," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 20, no. 4, pp. 505–517, Apr. 2010.

[2] 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, Mar. 2011.

[3] A. Aksay, S. Pehlivan, E. Kurutepe, C. Bilen, T. Ozcelebi, G. B. Akar, M. R. Civanlar, and A. M. Tekalp, "End-to-end stereoscopic video streaming with content-adaptive rate and format control," *Signal Processing: Image Communications*, vol. 22, no. 2, pp. 157–168, Feb. 2007.

[1] Denoted next to every operating point of *Opt*.

[4] E. Kurutepe, M. R. Civanlar, and A. M. Tekalp, "Client-driven selective streaming of multiview video for interactive 3DTV," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 17, no. 11, pp. 1558–1565, Nov. 2007.

[5] A. Vetro, T. Wiegand, and G. J. Sullivan, "Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC standard," *Proceedings of the IEEE*, vol. 99, no. 4, pp. 626–642, Apr. 2011.

[6] P. Merkle, A. Smolic, K. Müller, and T. Wiegand, "Multi-view video plus depth representation and coding," in *Proc. Int'l Conf. Image Processing*, vol. 1. San Antonio, TX, USA: IEEE, Sep. 2007, pp. 201–204.

[7] P. Merkle, Y. Morvan, A. Smolic, D. Farin, K. Müller, P. H. N. de With, and T. Wiegand, "The effects of multiview depth video compression on multiview rendering," *Signal Processing: Image Communication*, vol. 24, no. 1–2, pp. 73–88, Jan. 2009.

[8] H. Everett, "Generalized Lagrange multiplier method for solving problems of optimum allocation of resources," *Operations Research*, vol. 11, no. 3, pp. 399–417, May-June 1963.

[9] B. L. Fox and D. M. Landi, "Searching for the multiplier in one-constraint optimization problems," *Operations Research*, vol. 18, no. 2, pp. 253–262, Mar.-Apr. 1970.

[10] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality video view interpolation using a layered representation," in *Computer Graphics Proc., Ann. Conf. Series*. Los Angeles, CA: ACM, Aug. 2004, pp. 600–608.