EFFICIENT BIT ALLOCATION FOR MULTIVIEW IMAGE CODING & VIEW SYNTHESIS

Gene Cheung

National Institute of Informatics

Vladan Velisavljević

Deutsche Telekom Laboratories

ABSTRACT

The encoding of both texture and depth maps of a set of multiview images, captured by a set of spatially correlated cameras, is important for any 3D visual communication systems based on depth-image-based rendering (DIBR). In this paper, we address the problem of efficient bit allocation among texture and depth maps of multi-view images. We pose the following question: for chosen (1) coding tool to encode texture and depth maps at the encoder and (2) view synthesis tool to reconstruct uncoded views at the decoder, how to best select captured views for encoding and distribute available bits among texture and depth maps of selected coded views, such that visual distortion of a "metric" of reconstructed views is minimized. We show that using the monotonicity assumption, sub-optimal solutions can be efficiently pruned from the feasible space during parameter search. Our experiments show that optimal selection of coded views and associated quantization levels for texture and depth maps can outperform a heuristic scheme using constant levels for all maps (commonly used in the standard implementations) by up to 2.0dB. Moreover, the complexity of our scheme can be reduced by up to 66% over full search without loss of optimality.

Index Terms— Multiview image, bit allocation, monotonicity.

1. INTRODUCTION

In a typical multiple view imaging scenario, a multiview image sequence is captured by a set of spatially correlated cameras simultaneously. Besides texture maps, depth information of a particular viewpoint (distance between camera and each captured pixel) can also be estimated or captured by special hardware, so that additional intermediate views can be synthesized using texture and depth maps of neighboring captured images via depth-image-based rendering (DIBR) [1]. Conveying both texture and depth maps of captured views for a large multiview image sequence to a decoder, however, means a large amount of data must be transmitted. Hence, compression of texture and depth maps of multiview images is important.

In response, compression strategies for texture and depth maps of multiview images have recently been proposed for DIBR [2, 3]. What is missing, however, is a comprehensive strategy to optimize coding parameters for DIBR in a rate-distortion (RD) sense. More specifically, given the decoder possesses a known view synthesis tool, how should the encoder decide the quantization levels of coded texture and depth maps of captured views $\mathcal{N}$ for a pre-defined DIBR metric? By metric, we assume here that given original captured views $\mathcal{N}$, a superset[1] $\mathcal{V} = \mathcal{N} \cup \mathcal{M}$, containing both the captured views $\mathcal{N}$ and *designated intermediate views* $\mathcal{M}$ (at least one between two consecutive captured views), is specified to evaluate the quality of the compressed texture and depth maps. Thus, the total distortion of the encoding would be distortion of decoded views using compressed texture maps of $\mathcal{N}$ compared to original $\mathcal{N}$, *and* distortion of synthesized views using compressed texture and depth maps of $\mathcal{N}$ compared to original $\mathcal{M}$.

Note that in general, not all captured views $\mathcal{N}$ need to be encoded for a given desired RD tradeoff. If the capturing cameras are sufficiently close spatially and the scene sufficiently interpolatable, then coding only a *subset* of captured views $\mathcal{N}$ (with finer quantization levels for texture and depth maps for high quality view synthesis), while relying on decoder to synthesize skipped captured views, may offer better RD tradeoff than encoding all captured views at coarser quantization levels. Hence finding the optimal subset of captured views $\mathcal{N}$ for encoding is also of critical importance.

In this paper, we propose a bit allocation algorithm that finds the optimal subset among captured views $\mathcal{N}$ for encoding, and assigns quantization levels for texture and depth maps of the selected coded views. We first establish that the optimal selection of coded views and associated quantization levels is equivalent to the shortest path in a specially constructed trellis. Given that the state space of the trellis is nonetheless enormous, we then show that using lemmas derived from monotonicity property in predictor's quantization level and distance, sub-optimal states and edges in the trellis can be pruned during shortest path calculation without loss of optimality. Experimental results show that optimal selection of captured views and associated quantization levels for texture and depth maps outperformed a heuristic scheme that selects all captured views for coding and assigns a fixed constant for all maps by up to 2.0dB. Further, our algorithm can reduce computation complexity over full trellis calculation by up to 66% without loss of optimality.

The paper is organized as follows. After describing related work in Section 2, we formulate our bit allocation problem in Section 3. Then, we introduce the monotonicity property and propose an efficient bit allocation algorithm in Section 4. We present our experimental results in Section 5. Finally, we conclude in Section 6.

2. RELATED WORK

New compression methods of depth maps [2, 3] for DIBR have been proposed, and formal analyses of resulting error in synthesized view due to lossy depth-map coding [4, 5] have been reported. Our bit allocation work is notably orthogonal to these proposals; no matter what coding methods are employed for texture and depth map and how a compressed depth map error manifests to a synthesized view, an optimizer must ultimately decide how to best distribute bits among texture and depth maps of coded views in a sequence for a desired RD tradeoff. To the best of our knowledge, our work is the first attempt in the literature to address this question formally.

Optimal bit allocation among independent [6] and dependent [7] quantizers in an operational sense has been studied thoroughly for RD optimized media compression. Our work differs in that bit allocation for both texture and depth maps are considered simultaneously, such that the resulting distortion of both encoded and synthesized views is minimized for a desired RD tradeoff.

[1] If the metric contains no intermediate views $\mathcal{M}$, the encoder would not encode any depth maps. Given depth maps provide decoder the flexibility to synthesize any number of intermediate views, this is a desired property and the metric should promote good quality depth maps for view synthesis.

3. FORMULATION

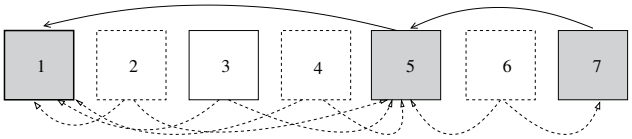


Fig. 1. Dependencies among Coded (Gray) and Synthesized (White) Frames. $\mathcal{J} = \{1, 5, 7\}$, $\mathcal{J}' = \{3\}$, $\mathcal{M} = \{2, 4, 6\}$.

The setting of our bit allocation problem is as follows. A *metric* of views $\mathcal{V} = \{1, \ldots, V\}$ in a 1D-camera-array arrangement, containing both camera-captured views $\mathcal{N}$ of size N and designated synthesized views $\mathcal{M}$, is specified *a priori* to evaluate quality of encoded texture and depth maps. Captured views $\mathcal{N}$ are divided into K *coded views*, $\mathcal{J} = \{j_1, \ldots, j_K\}$, and $N - K$ *uncoded views* $\mathcal{J}' = \mathcal{N} \setminus \mathcal{J}$. The first and last view in $\mathcal{V}$ are captured views and must be selected as coded views; i.e., $1, V \in \mathcal{J} \subseteq \mathcal{N}$. Texture and depth maps of a coded view j_i are encoded using quantization level q_{j_i} and p_{j_i}, respectively. q_{j_i} and p_{j_i} take on discrete values from quantization level set $\mathcal{Q} = \{1, \ldots, Q_{\max}\}$ and $\mathcal{P} = \{1, \ldots, P_{\max}\}$, respectively, where we assume the convention that a larger q_{j_i} or p_{j_i} implies a coarser quantization.

Uncoded views are not encoded at the encoder, but are synthesized at the decoder, along with designated synthesized views $\mathcal{M}$ in metric $\mathcal{V}$, each using texture and depth maps of the closest left and right coded views, denoted as $l, r \in \mathcal{J}$ for each $j' \in \mathcal{J}'$.

We assume inter-view differential coding is used for coded views as done in [8] and shown in Fig. 1. The first view is always coded as an I-frame. Each subsequent coded view j_i—frames 5 and 7 in Fig. 1—are coded as P-frame using previous coded view j_{i-1} as predictor for motion/disparity compensation.

3.1. Visual Distortion

Given the coded view dependencies, we can now write the distortion D^c of the coded views as a function of the texture map quantization levels, $\mathbf{q} = [q_{j_1} \ldots, q_{j_K}]$:

$$D^c(\mathbf{q}) = d_1^c(q_1) + \sum_{i=2}^{K} d_{j_i,j_{i-1}}^c(q_{j_i}, q_{j_{i-1}}) \tag{1}$$

(1) states that distortion of the starting I-frame d_1^c depends only on its own texture quantization level q_1, while the distortion of a P-frame $d_{j_i}^c$ depends on both its own texture quantization level q_{j_i} and its predictor j_{i-1}'s quantization level $q_{j_{i-1}}$. A more general model [7] is to have P-frame j_i depends on its own q_{j_i} and all previous quantization levels $q_1, \ldots, q_{j_{i-1}}$. We assume here that truncating the dependencies to $q_{j_{i-1}}$ only is a good first-order approximation.

Similarly, we now write the distortion of the synthesized views D^s (including uncoded views $\mathcal{J}'$ and designated synthesized views $\mathcal{M}$) as a function of $\mathbf{q}$ and depth quantization levels, $\mathbf{p} = [p_{j_1}, \ldots, p_{j_K}]$:

$$\begin{aligned} D^s(\mathbf{q}, \mathbf{p}) &= \sum_{j' \in \mathcal{J}' \cup \mathcal{M}} d_{j',l,r}^s(q_l, p_l, q_r, p_r) \qquad (2) \\ l &= \arg\min_{j \in \mathcal{J}} |j' - j| \qquad \text{s.t. } j < j' \\ r &= \arg\min_{j \in \mathcal{J}} |j' - j| \qquad \text{s.t. } j > j' \end{aligned}$$

where l and r are indices of the closest coded views to the left and right of synthesized view j'. In words, distortion of synthesized view j' depends on both the texture and depth map quantization levels of the two spatially closest coded views l and r.

3.2. Encoding Rate

As done for distortion, we can write the rate of texture and depth maps of coded views, R^c and R^s, respectively, as follows:

$$R^c(\mathbf{q}) = r_1^c(q_1) + \sum_{i=2}^{K} r_{j_i,j_{i-1}}^c(q_{j_i}, q_{j_{i-1}}) \tag{3}$$

$$R^s(\mathbf{q}, \mathbf{p}) = r_1^s(q_1, p_1) + \sum_{i=2}^{K} r_{j_i,j_{i-1}}^s(q_{j_i}, p_{j_i}, q_{j_{i-1}}, p_{j_{i-1}}) \tag{4}$$

(3) states that the encoding rate for texture map of a coded view, $r_{j_i}^c$, depends on its texture map quantization level, q_{j_i}, and its predictor's level, $q_{j_{i-1}}$. In contrast, (4) states that the encoding rate for depth map, $r_{j_i}^s$, depends on both the texture and depth map quantization levels, q_{j_i} and p_{j_i}, and its predictor's texture and depth map levels, $q_{j_{i-1}}$ and $p_{j_{i-1}}$. Our model hence includes the case when depth maps are differentially coded using texture maps as predictors.

3.3. Rate-distortion Optimization

Given the above formulation, the optimization we are interested in is to find the coded view indices $\mathcal{J} \subseteq \mathcal{N}$, and associated texture and depth quantization vector, $\mathbf{q}$ and $\mathbf{p}$, such that the Lagrangian objective is minimized for given Lagrangian multiplier $\lambda \geq 0$:

$$\min_{\mathcal{J}, \mathbf{q}, \mathbf{p}} \Phi_\lambda = D^c(\mathbf{q}) + D^s(\mathbf{q}, \mathbf{p}) + \lambda \left[R^c(\mathbf{q}) + R^s(\mathbf{q}, \mathbf{p})\right] \tag{5}$$

4. BIT ALLOCATION OPTIMIZATION

Previous work [7] has shown that using monotonicity property of dependent quantizers, efficient algorithms and heuristics can be constructed for optimal or near-optimal bit allocation. Our work can be viewed as a generalization of [7] to include synthesized views. We first discuss the useful monotonicity property along different dimensions. We then derive two lemmas based on monotonicity, and construct a fast optimization algorithm using the lemmas.

4.1. Monotonicity in Predictor's Quantization Level

Let $\phi_{j_i,j_{i-1}}(q_{j_i}, p_{j_i}, q_{j_{i-1}}, p_{j_{i-1}})$ be the Lagrangian term for coded view j_i given quantization levels of view j_i and its predictor view j_{i-1}, i.e., the sum of distortion $d_{j_i,j_{i-1}}^c(q_{j_i}, q_{j_{i-1}})$ and penalties $\lambda r_{j_i,j_{i-1}}^c(q_{j_i}, q_{j_{i-1}})$ and $\lambda r_{j_i,j_{i-1}}^s(q_{j_i}, p_{j_i}, q_{j_{i-1}}, p_{j_{i-1}})$ for texture and depth maps encoding. Motivated by a similar empirical observation in [7], we assume here also the *monotonicity in predictor's quantization level* for both Lagrangian $\phi_{j_i,j_{i-1}}$ of coded view j_i, and distortion $d_{j',l,r}^s$ of synthesized view j'; i.e., for any $\lambda \geq 0$:

$$\begin{aligned} \phi_{j_i,j_{i-1}}(q_{j_i}, p_{j_i}, q_{j_{i-1}}, p_{j_{i-1}}) &\leq \phi_{j_i,j_{i-1}}(q_{j_i}, p_{j_i}, q_{j_{i-1}}^+, p_{j_{i-1}}) \qquad (6) \\ \phi_{j_i,j_{i-1}}(q_{j_i}, p_{j_i}, q_{j_{i-1}}, p_{j_{i-1}}) &\leq \phi_{j_i,j_{i-1}}(q_{j_i}, p_{j_i}, q_{j_{i-1}}, p_{j_{i-1}}^+) \\ d_{j',l,r}^s(q_l, p_l, q_r, p_r) &\leq d_{j',l,r}^s(q_l^+, p_l, q_r, p_r) \qquad (7) \\ d_{j',l,r}^s(q_l, p_l, q_r, p_r) &\leq d_{j',l,r}^s(q_l, p_l^+, q_r, p_r) \end{aligned}$$

where q_n^+ (or p_n^+) implies a larger (coarser) quantization level than q_n (or p_n). In words, (6) states that if predictor view j_{i-1} uses a coarser quantization level in texture or depth map, it will lead to worse prediction in view j_i, resulting in larger distortion and/or coding rate, and hence a larger Lagrangian cost $\phi_{j_i,j_{i-1}}$, $\lambda \geq 0$.

(7) makes a similar statement for monotonicity of the synthesized view distortion $d_{j_i',l,r}^s$ with respect to the texture and depth quantization levels q_l and p_l of the closest left coded view l. We assume also monotonicity in the texture and depth quantization levels q_r and p_r of the closest right coded view r as well.

4.2. Monotonicity in Predictor's Distance

We can also express monotonicity of Lagrangian cost $\phi_{j,k}$ of coded view j, or synthesized view distortion $d^s_{j',l,r}$ of synthesized view j', with respect to the *predictor's distance* to a coded view used for differential coding or synthesis. Assuming further-away predictor view k^- for coded view j, $k^- < k$, has the same quantization levels as view k, and further-away predictor views l^- and r^+ have the same levels for synthesized view j' as respective levels of views l and r, we can write:

$$\phi_{j,k}(q_j, p_j, q_k, p_k) \leq \phi_{j,k^-}(q_j, p_j, q_k, p_k) \quad (8)$$

$$d^s_{j',l,r}(q_l, p_l, q_r, p_r) \leq d^s_{j',l,r^+}(q_l, p_l, q_r, p_r) \quad (9)$$

$$d^s_{j',l,r}(q_l, p_l, q_r, p_r) \leq d^s_{j',l^-,r}(q_l, p_l, q_r, p_r).$$

Here, $r^+ > r$ or $l^- < l$ implies a further-right coded view r^+ or further-left coded view l^- is used to synthesize view j'. In words, (8) and (9) say that using a further-away predictor to differentially encode or synthesize a view, given the quantization levels of texture and depth maps of the further-away predictor are the same, results in no smaller Lagrangian cost or synthesized distortion. These inequalities hold true under an assumption of Lambertian scenes.

4.3. Full Trellis & Viterbi Algorithm

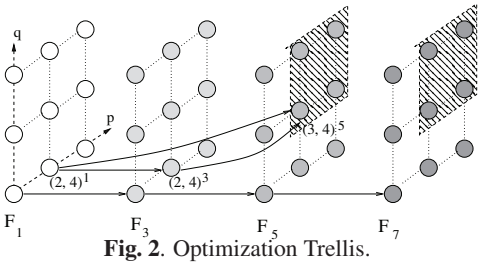


Fig. 2. Optimization Trellis.

We first show that the optimal solution to (5) can be computed by first constructing a trellis, and then finding the shortest path from the left end of the trellis to the right end using the famed Viterbi Algorithm (VA). Nevertheless, the complexity of constructing the full trellis is large, and hence we will discuss methods to reduce the complexity using the monotonicity property described earlier.

We can construct a trellis—one corresponding to the earlier example is shown in Fig. 2—for the selection of coded view indices $\mathcal{J}$, texture and depth quantization levels $\mathbf{q}$ and $\mathbf{p}$, as follows. Each captured view $j_i \in \mathcal{N}$ is represented by a *plane* of states, where each state represents a pair of levels $(q_{j_i}, p_{j_i})^{j_i}$ for texture and depth maps. States in the first plane corresponding to the first view 1 will be populated with Lagrangian costs $\phi_1(q_1, p_1)$'s for different level pairs $(q_1, p_1)^1$'s. Each directed edge from a state $(q_1, p_1)^1$ in the first plane to a state in the second plane $(q_j, p_j)^j$ of neighboring captured view $j \in \mathcal{N}$ will carry a Lagrangian cost $\phi_{j,1}(q_j, p_j, q_1, p_1)$ *and* designated synthesized view distortions $\sum_{1<j'<j} d^s_{j',1,j}(q_1, p_1, q_j, p_j)$. Selecting such edge would mean captured views 1 and j are both selected as coded views in $\mathcal{J}$. Each directed edge from a state $(q_1, p_1)^1$ in the first plane to a state $(q_k, p_k)^k$ in a further-away plane of captured view $k \in \mathcal{N}$ will carry similar Lagrangian cost $\phi_{k,1}(q_k, p_k, q_1, p_1)$ and synthesized view distortions $\sum_{1<j'<k} d^s_{j',1,k}(q_1, p_1, q_k, p_k)$. Selecting such edge would mean captured view 1 and k are both selected as coded views in $\mathcal{J}$ with no coded views in-between.

We state without proof that the shortest path from any state in the left-most plane to any state in the right-most plane, found using VA, corresponds to the optimal solution to (5). However, the number of states and edges in the trellis alone are prohibitive: $O(|\mathcal{Q}||\mathcal{P}|N)$ and $O(|\mathcal{Q}|^2|\mathcal{P}|^2N^2)$, respectively. Hence the crux to reduce complexity is to find the shortest path by visiting only a small subset of states and edges. We discuss this next.

4.4. Reducing Complexity

Let $\Phi_j(q_j, p_j)$ be the shortest sub-path from any states of first view to state $(q_j, p_j)^j$ of captured view j. The first lemma eliminates *sub-optimal states* $(q_j, p_j)^j$'s, given computed $\Phi_j(q_j, p_j)$'s, using monotonicity in quantization level.

Lemma 1 *For given p_{j_i}, if at state plane of captured view j_i, $\Phi_{j_i}(q^+_{j_i}, p_{j_i}) > \Phi_{j_i}(q^*_{j_i}, p_{j_i})$, $\forall q^+_{j_i} > q^*_{j_i}$, then sub-paths up to states $(q^+_{j_i}, p_{j_i})^{j_i}$, $\forall q^+_{j_i} > q^*_{j_i}$, cannot belong to shortest path.*

Proof of Lemma 1 *We prove by contradiction. Suppose shortest sub-path up to state $(q^+_{j_i}, p_{j_i})^{j_i}$, $q^+_{j_i} > q^*_{j_i}$, is part of an end-to-end shortest path. If we replace sub-path to $(q^+_{j_i}, p_{j_i})$ with sub-path to $(q^*_{j_i}, p_{j_i})$, a synthesized view j' to the right of j_i and coded view j_{i+1} that depend on view j_i's texture map will have no larger distortion d^s_{j',j_i} or Lagrangian cost ϕ_{j_{i+1},j_i}, if $q^*_{j_i}$ is used instead of $q^+_{j_i}$, by monotonicity in quantization level (6) and (7). Given $\Phi_{j_i}(q^+_{j_i}, p_{j_i}) > \Phi_{j_i}(q^*_{j_i}, p_{j_i})$, we see that replacing sub-path to $(q^+_{j_i}, p_{j_i})^{j_i}$ with sub-path to $(q^*_{j_i}, p_{j_i})^{j_i}$ will yield strictly lower Lagrangian cost. A contradiction.* □

Lemma 1 also holds true for depth quantization level p_{j_i}: given q_{j_i}, if $\Phi_{j_i}(q_{j_i}, p^+_{j_i}) > \Phi_{j_i}(q_{j_i}, p^*_{j_i})$, $\forall p^+_{j_i} > p^*_{j_i}$, then states $(q_{j_i}, p^+_{j_i})^{j_i}$'s, $\forall p^+_{j_i} > p^*_{j_i}$, are sub-optimal and can be eliminated.

The second lemma eliminates *sub-optimal edges* from state $(p_j, q_j)^j$ of captured view j to a state in further-away coded view k using monotonicity in prediction distance.

Lemma 2 *Given start state $(q_{j_i}, p_{j_i})^{j_i}$ of view j_i, end state $(q_k, p_k)^k$ of view k, and intermediate view j_{i+1}, $j_i < j_{i+1} < k$, if cost of traversing state $(q_{j_i}, p_{j_i})^{j_{i+1}}$ of view j_{i+1}, $\phi_{j_{i+1},j_i} + \sum_{j_i<j'<j_{i+1}} d^s_{j',j_i,j_{i+1}}$ is smaller than a lower-bound cost of skipping view j_{i+1}, $\sum_{j_i<j'\leq j_{i+1}} d^s_{j',j_i,k}$, then edge $(q_{j_i}, p_{j_i})^{j_i} \rightarrow (q_k, p_k)^k$ cannot belong to an end-to-end shortest path.*

Proof of Lemma 2 *We prove by contradiction. Suppose an optimal end-to-end path includes edge $(q_{j_i}, p_{j_i})^{j_i} \rightarrow (q_k, p_k)^k$. If we replace it with two edges $(q_{j_i}, p_{j_i})^{j_i} \rightarrow (q_{j_i}, p_{j_i})^{j_{i+1}} \rightarrow (q_k, p_k)^k$, the cost of traversing state $(q_{j_i}, p_{j_i})^{j_{i+1}}$ for views j''s, $j_i < j' \leq j_{i+1}$, is smaller than not traversing it by assumption. Moreover, Lagrangian cost of coded view k and distortion of a synthesized view to the right that depended on coded view j_i will not increase using view j_{i+1} instead with same quantization levels due to monotonicity of prediction distance (8) and (9). Hence a path using the two replacement edges will yield lower cost. A contradiction.* □

The corollary of Lemma 2 is that if the said condition holds, then edges $(q_{j_i}, p_{j_i})^{j_i} \rightarrow (q_{k^+}, p_{k^+})^{k^+}$, $\forall q_{k^+} \geq q_k, p_{k^+} \geq p_k$, where k^+ means all indices larger than k, also cannot belong to the shortest path. The reason is: synthesized distortion $d^s_{j_{i+1},j_i,k}$ using view j_i and k as predictors is surely no larger than $d^s_{j_{i+1},j_i,k^+}$ using same j_i and further-away k^+ with same or coarser quantization levels. Hence the said condition must hold also for $(q_{k^+}, p_{k^+})^{k^+}$ as well, and the same argument as proof 2 follows to rule out edge $(q_{j_i}, p_{j_i})^{j_i} \rightarrow (q_{k^+}, p_{k^+})^{k^+}$. As an example, in Fig. 2 if the cost of traversing state $(2,4)^3$, $\phi_{3,1} + d^s_{2,1,3}$, is smaller than $d^s_{3,1,5} + d^s_{2,1,5}$, then edges from $(2,4)^1$ to all states on the shaded region, including $(3,4)^5$ of view 5, can be eliminated.

4.5. Bit Allocation Algorithm

We now describe a bit allocation algorithm, shown in Fig. 3, exploiting the lemmas derived in previous section to reduce complexity from the full trellis. The basic idea is to try to prune as many states and edges in the trellis as early as possible. Starting from

1. Initialize $i = 1$. Compute $\phi(q_1, p_1)$'s as $\Phi(q_i, p_1)$'s for all states $(q_1, p_1)^1$'s of F_1.
2. For each p_i of view i, find q_i^* s.t. $\Phi_i(q_i^+, p_i) > \Phi_i(q_i^*, p_i), \forall q_i^+ > q_i^*$. Eliminate states $(q_i^+, p_i)^i$'s from consideration.
3. For each q_i of view i, find p_i^* s.t. $\Phi_i(q_i, p_i^+) > \Phi_i(q_i, p_i^*), \forall p_i^+ > p_i^*$. Eliminate states $(q_i, p_i^+)^i$'s from consideration.
4. For each survived state $(q_i, p_i)^i$ of view i, evaluate forward sub-paths to states $(q_j, p_j)^j$'s of neighboring captured view j, $j > i$.
5. For each survived state $(q_i, p_i)^i$ of view i, using state $(q_i, p_i)^j$ of neighboring captured view j, evaluate sub-paths forward:
   (a) Initialize k to be neighboring captured view of j, $k > j$, and length-$P_{\max}$ vector $\mathbf{Q}_{\lim}$ to $[Q_{\max}, \ldots, Q_{\max}]$.
   (b) for each state $(q_k, p_k)^k$, $q_k \leq \mathbf{Q}_{\lim}(p_j)$, if $\phi_{j,i} + \sum_{i<j'<j} d^s_{j',i,j} > \sum_{i<j'\leq j} d^s_{j',i,k}$, then evaluate possible path with edge $(q_i, p_i)^i \rightarrow (q_k, p_k)^k$ to state $(q_k, p_k)^k$. If not, $\mathbf{Q}_{\lim}(p_k^+) = q_k - 1, \forall p_k^+ \geq p_k$.
   (c) If $k < N$ and $\mathbf{Q}_{\lim}$ is non-zero vector, increment k to next neighboring captured view and repeat step 5(b).
6. If $i < N$, increment i to next neighboring captured view and repeat step 2 to 5.

Fig. 3. Bit Allocation Algorithm

the left-side of trellis, for each captured view i, using computed sub-paths to states $(q_i, p_i)^i$'s with Lagrangian costs $\Phi_i(q_i, p_i)$'s[2], we first eliminate states with larger Lagrangian costs Φ_i's and coarser texture quantization levels q_i^+'s than a minimum state (q_i^*, p_i), given p_i. Same procedure is applied for the depth quantization levels p_i's given fixed q_i. These sub-optimal states are eliminated due to lemma 1.

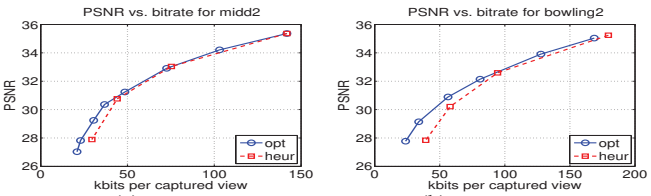
In step 4, for each survived state $(q_i, p_i)^i$ of view i, we *evaluate* all forward sub-paths to states $(q_j, p_j)^j$'s of the next captured view j. By "evaluate", we mean comparing the sum of $\Phi_i(q_i, p_i)$ and $\phi_{j,i} + \sum_{i<j'<j} d^s_{j',i,j}$ to the cost of the best sub-path to $(q_j, p_j)^j$ to date, $\Phi_j(q_j, p_j)$, for each state $(q_j, p_j)^j$. If the former is smaller, $\Phi_j(q_j, p_j)$ will be updated accordingly.

In step 5, for each survived state $(q_i, p_i)^i$, we next evaluate feasible edges to states $(q_k, p_k)^k$'s of captured views k's, $k > j$. Feasible edges are ones that satisfy $\phi_{j,i} + \sum_{i<j'<j} d^s_{j',i,j} > \sum_{i<j'\leq j} d^s_{j',i,k}$. We stop when there are no more forward feasible edges. We can identify the shortest end-to-end path by finding the minimum cost state $(q_N, p_N)^N$ of view N and tracing its back to view 1.

5. EXPERIMENTATION

To test the effectiveness of our proposed bit allocation scheme, we used H.264 JM16.2 video codec to encode texture and depth maps (texture and depth maps were encoded independently from each other), and used ViSBD 2.1 as view synthesis tool at the decoder. For test sequences, we used two Middlebury still image sequences [9], `midd2` and `bowling2`, of size 1366×1110 and 1330×1110, respectively. We assumed captured camera views were $\mathcal{N} = \{0, 2, 4, 6\}$, and designated synthesized views were $\mathcal{M} = \{1, 3, 5\}$, whereas the available quantization levels for both texture and depth maps were $\mathcal{Q} = \mathcal{P} = \{10, 15, \ldots, 50\}$. Rate controls were disabled in JM16.2, and software modifications were made so that a particular quantization level can be specified for each individual frame.

[2]Lagrangian costs of first view 1 are simply $\phi_1(q_1, p_1)$'s.



(a) `midd2` (b) `bowling2`

Fig. 4. Performance Comparison between Optimal and Heuristic Coded View and Quantization Level Selection Schemes

We tested the performance of our proposed scheme `opt` with a simple heuristic scheme `heur` that selects all captured views $\mathcal{N}$ for coding, i.e., $\mathcal{J} = \mathcal{N}$, and assigns a constant quantization level to all texture and depth maps of coded views. In Fig.4, we see the performance of both schemes, shown as PSNR (quality) versus bitrate per captured view (including both texture and depth maps) for the two test sequences. First, we see that `opt` has better RD performance than `heur`—by up to 1.2dB and 2.0dB for `midd2` and `bowling2`, respectively. This shows that correct selection of quantization levels per frame is important. Second, as bitrate decreased, `opt` selected fewer captured views for coding and relied instead on decoder's view synthesis (five left-most points of `opt` in `midd2` represented selections of uncoded views). This is also the region where `opt` out-performed `heur` the most, hence selection of captured views for coding is also important for best RD performance.

When generating `opt` curves, we tracked the amount of computation performed using our scheme over a full trellis search approach. We found the computation savings ranged from 40% to 66%, with the maximum saving occurring at the right-most RD point.

6. CONCLUSIONS

Towards the goal of efficient depth-image-based rendering (DIBR), in the paper we presented an algorithm to select captured views for coding and quantization levels of corresponding texture and depth maps in a rate-distortion (RD) optimal manner. We show that using monotonicity in predictor's quantization level and distance, search complexity can be drastically reduced without loss of optimality. Experiments show that our selection scheme outperformed a heuristic scheme by up to 2.0dB in PSNR for the same bitrate.

7. REFERENCES

[1] P. Merkle, A. Smolic, K. Mueller, and T. Wiegand, "Multi-view video plus depth representation and coding," in *IEEE International Conference on Image Processing*, San Antonio, TX, October 2007.

[2] Y. Morvan, D. Farin, and Peter H.N. de With, "Depth-image compression based on an R-D optimized quadtree decomposition for the transmission of multiview images," in *IEEE International Conference on Image Processing*, San Antonio, TX, September 2007.

[3] M. Maitre, Y. Shinagawa, and M.N. Do, "Wavelet-based joint estimation and encoding of depth-image-based representations for free-viewpoint rendering," in *IEEE Transactions on Image Processing*, June 2008, vol. 17, no.6, pp. 946–957.

[4] Y. Liu et al, "Compression-induced rendering distortion analysis for texture / depth rate allocation in 3D video compression," in *Data Compression Conference*, Snowbird, UT, March 2009.

[5] W.-S. Kim, A. Ortega, P. Lai, D. Tian, and C. Gomila, "Depth map distortion analysis for view rendering and depth coding," in *IEEE International Conference on Image Processing*, Cairo, Egypt, November 2009.

[6] Y. Shoham and A. Gersho, "Efficient bit allocation for an aibitrary set of quantizers," in *IEEE Transactions on Acoustics, Speech, and Signal Processing*, September 1988, vol. 36, no.9, pp. 1445–1453.

[7] K. Ramchandran, A. Ortega, and M. Vetterli, "Bit allocation for dependent quantization with applications to multiresolution and MPEG video coders," in *IEEE Transactions on Image Processing*, September 1994, vol. 3, no.5.

[8] P. Merkle, A. Smolic, K. Muller, and T. Wiegand, "Efficient prediction structures for multiview video coding," in *IEEE Transactions on Circuits and Systems for Video Technology*, November 2007, vol. 17, no.11, pp. 1461–1473.

[9] "2006 stereo datasets," http://vision.middlebury.edu/stereo/data/scenes2006/.