ANALYSIS OF THE DECODING-COMPLEXITY OF COMPRESSED IMAGE-BASED SCENE REPRESENTATIONS

Ingo Bauermann and Eckehard Steinbach

Institute of Communication Networks, Media Technology Group,
Technische Universität München, Munich, Germany
{ingo.bauermann, eckehard.steinbach}@tum.de

ABSTRACT

Interactive navigation in image-based scenes requires random access to the compressed reference image data. When using state of the art block-based hybrid video coding techniques, the degree of inter and intra block dependencies introduced during compression has an impact on the effort required to access reference image data and therefore delimits the response time for interactive applications. In this work a theoretical model for the decoding-complexity of compressed image-based scene representations is presented and evaluated. Results show the validity of the model. Additionally, results for decoding-complexity constrained rate distortion optimization (RDC) using our model show the benefit of incorporating the computational power of client devices into the compression process.

Index Terms— decoding complexity, compression, image-based rendering

1. INTRODUCTION

Image-based scene representations such as light fields and others (see [1] for an overview) provide a fast and easy acquisition and rendering process. Based on sampling of the plenoptic function, the downside of these representations is the large amount of reference image data that has to be stored. Compression and streaming schemes for image-based rendering have been adopted from common video compression techniques (e.g. [2],[3],[4]).

While for video sequences sequential play out of whole frames is dominant, free navigation in image-based scenes requires random access to individual frames or even small image parts like pixel blocks to compose virtual views. When the reference image data is compressed using, e.g., motion compensated prediction, the amount of dependently encoded image data determines the number of pixels that have to be decoded with respect to the number of pixels actually needed for rendering. We call this ratio the decoding-complexity which is a strong measure for the performance (like frame rate or response time) of an image-based walkthrough system. A low decoding-complexity indicates that random access to image data is easily performed resulting in a high rendering speed while the compression efficiency is low. On the other hand a high decoding-complexity provides good compression efficiency as dependencies are exploited during encoding but random access is associated with a high effort and a low response time. The goal of this work is to provide and to evaluate a theoretical model for the decoding-complexity of compressed image-based scene representations encoded using block-based hybrid video coding concepts. In this context, pixel domain caching and basic data access patterns are investigated.

The remainder of this paper is structured as follows. Section 2 introduces the image-based compression and rendering system under consideration. Section 3 describes the procedure used in this work to measure the mean decoding-complexity of a compressed image sequence. Section 4 provides theoretical descriptions of the decoding-complexity in systems without a cache. Section 5 provides models for systems with pixel domain caching. In Section 6 the derived models are extended to full virtual views while in Section 7 the theoretical models and results on RDC optimization are evaluated. Section 8 concludes the paper.

2. CONSIDERED COMPRESSION SCHEME

The input to the system under consideration is a sequence of calibrated images of a static scene. We perform offline compression on group of pictures (GOP) of size N images. Consecutive frames are encoded using motion compensated prediction of BxB pixel blocks. Similar to block matching in video compression, a displacement of blocks in consecutive frames is calculated. Due to the epipolar constraint (e.g. [5]) these block displacements are described by a scalar motion displacement Δd (pixels) on a vector valued function $[\Delta x,\Delta y]^T=\boldsymbol{D}(x,y,\Delta d)$ which can be determined from the frame calibration as a static scene is assumed to be captured. For simplicity we assume a rectified image sequence as the input to our system. Consequently, only horizontal motion compensation has to be performed.

Pixel blocks can be encoded in intra and inter/skip modes. The first frame of a GOP is encoded entirely using the intra mode. Intensity values in intra mode undergo a typical encoding procedure (e.g., transform coding + quantization). The exact scheme is not important for this work as long as it is significantly more complex than simple pixel copying. For the inter block mode the residual error after motion compensated prediction (with the previous frame as reference) is encoded using the intra encoding scheme. For the skip mode only motion compensated prediction is performed.

Virtual views are rendered from decoded pixel blocks containing the needed pixel data. The corresponding reference blocks in the previous frame are decoded as well. Then references of these blocks are decoded. This procedure is continued until all reference blocks can be decoded using the intra mode decoding procedure. The smallest decodable unit in our scheme is a single pixel block.

3. MEASURING THE DECODING-COMPLEXITY

Independent from the intra encoding scheme used, which can be based on transform coding or any other technique, we define the decoding-complexity C for decoding one pixel block from the compressed data as the number of pixels that have to be decoded to reconstruct a pixels RGB values completely. This definition is chosen as intra and residual error decoding consumes most of the decoding time while skip modes only require copy operations. For random access to independently encoded blocks the decoding-complexity C of a pixel block at a position $\boldsymbol{p}=[x,y]^{\mathrm{T}}$ in a frame with frame index f (frame f=0 is entirely independently encoded) can now be written as

$$C(\boldsymbol{p},f)=1 \text{ [decoded pixel per rendered pixel]} \quad (1)$$

For blocks encoded dependent on blocks in previous frames, dependencies have to be resolved. To determine the reference block position we define the relative block displacement Δd_B (unit: block) which is calculated using the motion displacement Δd (unit: pixel):

$$\Delta d_B(\boldsymbol{p},f)=\left\lfloor \frac{\Delta d(\boldsymbol{p},f)}{B} \right\rfloor \quad (2)$$

This relative block displacement represents the motion in block units. The residual decoding-complexity is set to M=1 when encoding in inter mode and M=0 when encoding in skip mode. For the case with a single reference $\Delta d(\boldsymbol{p},l) \pmod B = 0$ (Δd is a multiple of B), C can be then determined as:

$$C(\boldsymbol{p},f)=M+C\left(\begin{bmatrix} x+B\cdot\Delta d_B(\boldsymbol{p},f) \\ y \end{bmatrix}, f-1\right) \quad (3)$$

and for the case that $\Delta d(\boldsymbol{p},l) \pmod B \neq 0$:

$$\begin{aligned} C(\boldsymbol{p},f)=M &+C\left(\begin{bmatrix} x+B\cdot\Delta d_B(\boldsymbol{p},f) \\ y \end{bmatrix}, f-1\right) \\ &+C\left(\begin{bmatrix} x+B\cdot(\Delta d_B(\boldsymbol{p},f)+1) \\ y \end{bmatrix}, f-1\right) \end{aligned} \quad (4)$$

When a block serving as reference already resides in the pixel domain cache then the corresponding $C(\boldsymbol{p},\mathrm{f})$ of this block is set to zero.

4. DECODING A BLOCK WITHOUT A CACHE

Assume that a fraction α of all encoded blocks (not counting the blocks in the first frame of the GOP) is to be encoded in intra mode. All other blocks are encoded using inter mode. The probability of a block to have dependent data is 1-α. Now, additionally, let a fraction of inter blocks have one reference block (Δd of these blocks is a multiple of B) while the rest have two reference blocks in the previous frame. This is reflected by the single reference ratio b which can be approximated from the probability distribution $p_{\Delta \mathrm{d}}$ of displacements Δd:

$$b=\sum_{k=-\infty}^{\infty} p_{\Delta d}(k\cdot B) \quad (5)$$

Assuming a uniform distribution of the displacements Δd, we can approximate b=1/B. Then, 2-b is the mean number of reference blocks for inter/skip encoded blocks. The mean decoding-complexity C for an access to an arbitrary block in the GOP can be modeled as:

$$C(N,\alpha,b)=\frac{1}{N}\cdot\sum_{f=0}^{N-1}\sum_{j=0}^{f}\left((1-\alpha)\cdot(2-b)\right)^{j} \quad (6)$$

5. DECODING A SINGLE BLOCK WITH A PIXEL DOMAIN CACHE

When using a sufficiently large pixel domain cache, modeling the random access decoding-complexity can be done using a statistical model based on the insights from Section 4. Figure 1 illustrates the approximated statistical relationship between dependent blocks in neighboring frames for the case that block (l,t)=(0,0) is needed for rendering. As the requested block has to be decoded we can write the decoding probability as $a_{0,0}$=1. When calculating the decoding probability $a_{1,1}$ we have to consider that block (0,0) and (0,1) might reference this one. We consider three cases:

1. (0,0) is encoded in inter mode and does not have a single reference; (0,1) has not been decoded;
2. (0,0) has not been decoded; (0,1) is in inter mode;
3. both, (0,0) and (0,1) have been decoded; (0,1) is inter or (0,0) is inter and has two references;

Equation (7) gives the generalized decoding probabilities $a_{m,n}$ for pixel blocks according to Figure 1 incorporating these three cases:

$$a_{m,n} = \begin{cases} 0 & \text{if } n > m \\ 1 & \text{if } m,n = 0 \\ a_{m-1,n} \cdot (1-\alpha) & \text{if } m \neq 0, n = 0 \\ a_{m-1,n-1} \cdot (1-\alpha) \cdot (1-b) \cdot (1-a_{m-1,n}) & \\ \quad + a_{m-1,n} \cdot (1-\alpha) \cdot (1-a_{m-1,n-1}) & \\ \quad + a_{m-1,n-1} \cdot a_{m-1,n} \cdot \left(\left(1-\alpha^2\right) - \alpha \cdot b \cdot (1-\alpha) \right) & \text{else} \end{cases} \quad (7)$$

Averaging over all possible random access points in a GOP we get the decoding-complexity C' for a single block random access event for systems with a sufficiently large pixel domain cache:

$$C'(N,\alpha,b) = \frac{1}{N} \cdot \sum_{f=0}^{N-1} \sum_{l=0}^{f} \sum_{t=0}^{l} a_{l,t} \quad (8)$$

Now we additionally consider that blocks can be encoded using the skip mode. Then, the decoding-complexity can be calculated similar to (8) except that the decoding-complexity caused by decoding non-intra blocks has to be scaled with respect to the skip rate β (ratio of blocks encoded in skip mode with respect to the number of blocks not encoded in intra mode). First the decoding-complexity C_I due to the intra-only frame at the GOP beginning has to be determined as those blocks do not contribute to skip blocks in question:

$$C_I(N,\alpha,b) = \frac{1}{N} \cdot \sum_{f=0}^{N-1} \sum_{t=0}^{f} a_{f,t} \quad (9)$$

With (8) the decoding-complexity C_C for a single block random access event with cache can be written as:

$$C_C(N,\alpha,b,\beta) = C_I + \left(C' - C_I\right) \cdot \left(1 - \beta \cdot (1-\alpha)\right) \quad (10)$$

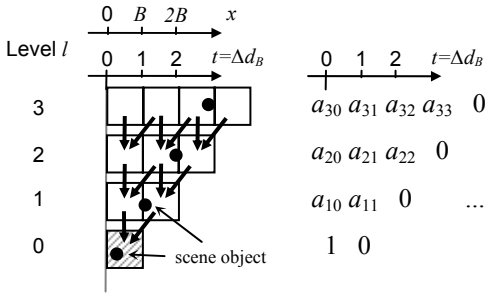

Figure 1: Derivation of a statistical model for the decoding-complexity in systems with pixel domain caching. Bold arrows indicate dependencies between blocks in neighboring frames. Block $(l,t)=(0,0)$ is requested. The probability that block (1,0) has to be decoded due to the request of block (0,0) is $a_{1,0}$ etc.

6. DECODING A VIRTUAL VIEW

Up to now only single block access has been considered. However, in IBR systems complete virtual views consist of image data from nearby frames and blocks. These neighboring blocks again share reference blocks. To approximate the decoding-complexity for a whole virtual view we assume that blocks from approximately the same position in neighboring frames are needed. For densely sampled image-based scene representations this assumption is valid. Consequently, the probabilities in (7) have to be changed to reflect the fact that all blocks with t=0 in Figure 1 have a decoding probability of 1:

$$a_{m,n} = \begin{cases} 0 & \text{if } n > m \\ 1 & \text{if } m,n = 0 \\ 1 & \text{if } m \neq 0, n = 0 \\ a_{m-1,n-1} \cdot (1-\alpha) \cdot (1-b) \cdot (1-a_{m-1,n}) & \\ \quad + a_{m-1,n} \cdot (1-\alpha) \cdot (1-a_{m-1,n-1}) & \\ \quad + a_{m-1,n-1} \cdot a_{m-1,n} \cdot \left(\left(1-\alpha^2\right) - \alpha \cdot b \cdot (1-\alpha) \right) & \text{else} \end{cases} \quad (11)$$

As not every pixel in a requested block is used for rendering, a correcting factor is introduced which depends on the rendering system and the block size B and has to be determined by experiment:

$$\gamma(B) = \frac{\text{number of pixels requested}}{\text{number of pixels needed for rendering}} \quad (12)$$

Equations (8) to (10) can now be applied for approximating the decoding-complexity for a virtual view C_{VV} using the probabilities from (11):

$$C_{VV}(N,\alpha,b,\beta,B) = \gamma(B) \cdot C_C(N,\alpha,b,\beta) \quad (13)$$

7. EXPERIMENTAL RESULTS

In the first experiment we consider decoding a single block with an initially empty cache. Several random access experiments are performed and evaluated using (1), (3) and (4) on different rate distortion optimized streams implementing different intra and skip ratios. A mean decoding-complexity is calculated and compared to the expected decoding-complexity from (10). Figure 2 shows the result. We assume a GOP size of N=10 frames and a block size of B=8. The model fits the measurements quite well. For evaluating our theoretical models for decoding arbitrary virtual views (13) we use an implementation of an image-based rendering system using a densely sampled concentric mosaic [6] to determine random access patterns for a real rendering system. Random views are generated and corresponding frame and block requests are evaluated. Results for different block sizes are shown in Figure 3 and 4 for different intra and skip ratios, respectively. Again, the model fits the measurements quite well. Mismatches are mainly due to oversimplification of the access pattern in (11). Finally, Figure 5 shows an application of the proposed model. Valid intra and skip ratio pairs are computed numerically from (13) for constrained C_{VV} (C_{VV}<C_{VVmax}). A maximum decoding-complexity C_{VVmax} corresponds to a fixed number of pixels rendered per time ensuring a mean response time for interactive rendering. The resulting RD plots with respect to the valid intra and skip ratios are

shown. The conclusion drawn from these results is that for the same desired response time and rendering quality the model can fit to different computational capabilities (C_{VVmax}) by adapting the coding efficiency.

8. CONCLUSION

In this paper we introduced and evaluated a theoretical model for the decoding-complexity of compressed image-based scene representation. Relevant data access patterns with the use of a pixel domain caching system are evaluated. Results for decoding-complexity constrained rate distortion optimization show the benefit of considering the computational power of client devices.

9. REFERENCES

[1] H.-Y. Shum, S.B. Kang, and S.-C. Chan, "Survey of Image-Based Representations and Compression Techniques," in IEEE Transactions on Circuits and Systems for Video Technology, pp. 1020–1037, Volume: 13, Issue: 11, Nov. 2003.

[2] C. Zhang and J. Li, "On the Compression and Streaming of Concentric Mosaic Data for Free Wondering in a Realistic Environment over the Internet," in IEEE Trans. on Multimedia. pp. 1170- 1182, Volume:7 Issue: 6, Dec. 2005.

[3] M.Magnor and B. Girod, "Data Compression for Light Field Rendering," IEEE Trans. on. Circuits and Systems for Video Technology, Vol. 10, No. 3, April 2000.

[4] P. Ramanathan and B. Girod, "Random Access for Compressed Light Fields Using Multiple Representations," Proc. IEEE International Workshop on Multimedia Signal Processing, MMSP 2004, Siena, Italy, September 2004.

[5] H. C. Longuet-Higgins, "A computer algorithm for reconstructing a scene from two projections," Nature, vol. 293, pp. 133–135, 1981.

[6] H.-Y. Shum, L.-W. He, "Rendering with Concentric Mosaics," ACM SIGGRAPH'99, Los Angeles, CA, pp.299–306, Aug. 1999.

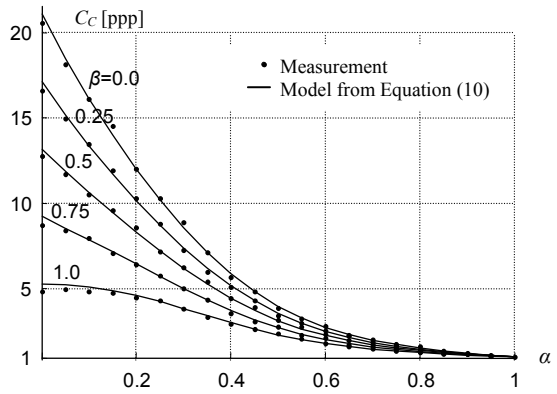

Figure 2: The decoding-complexity measured in decoded pixel per pixel needed for rendering as a function of the intra-rate α and the skip-rate β for a GOP size of N=10 frames assuming full-pel motion compensation and a block size of 8x8 pixel. A sufficiently large cache is available.

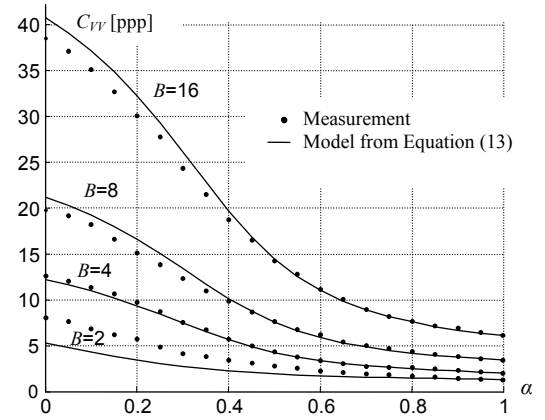

Figure 3: The decoding-complexity C_{VV} measured in decoded pixel per pixel needed for rendering as a function of the intra-rate α and the skip-rate β for a GOP size of N=13 frames assuming full-pel motion compensation and a sufficiently large cache. β=0.

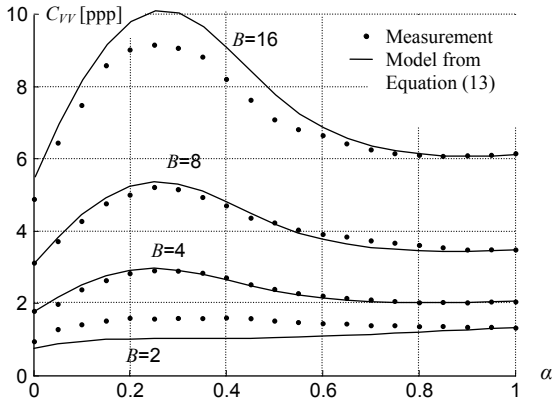

Figure 4: Same as Figure 3 except that β=1.0.

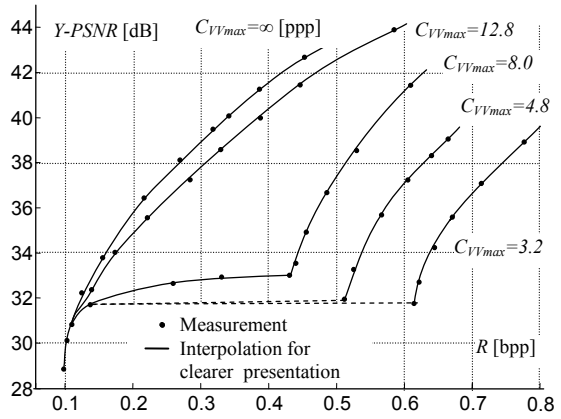

Figure 5: Rate distortion performance of decoding-complexity constrained RD optimization for random access to an arbitrary virtual view using a pixel domain cache. N=20, B=8. The dashed lines denote gaps in the curves due to not feasible areas in the parameter space.