Rate-distortion control in wavelet-based video compression systems with memory restriction

Eugeniy Belyaev
ebelyaev@vu.spb.ru

Andrey Turlikov
turlikov@vu.spb.ru

Anna Ukhanova
anja@vu.spb.ru

Abstract

This paper presents algorithm for wavelet-based video compression in video transmission systems with the constraint memory budget that is insufficient to store a single video frame. The proposed algorithm of rate-distortion control provides with the equal acceptable level of visual quality over the whole frame taking into consideration the channel throughput and transmitter/receiver memory constraints.

I. Introduction

Video transmission systems over wired/wireless channels are widely used in many spheres of the everyday life. However, such systems still have high memory and computational requirements. This paper focuses on the development of algorithm for video compression with high level of visual quality for a given channel throughput that could be realized with sufficiently small transmitter/receiver memory consumption (the whole memory of the device is much smaller than the number of bits for one frame of video sequence).

Standard JPEG2000 [1] is the most adapted solution for this task as it provides bit rate control with a high degree of precision. On the other hand, following problems could be caused. As there are memory restrictions, the original image is separated into non-overlapping tiles that are compressed independently. Rate control mechanism in standard JPEG2000 provides that compressed tiles have approximately equal bit sizes. However, since statistical properties of each tile vary a lot, some tiles are compressed with a high level of visual quality, and some - with a low one.

The algorithm proposed in this paper provides with the acceptable level of visual quality for all tiles for a given channel throughput.

II. Bitrate control mechanism in JPEG2000

The general scheme of the algorithm JPEG2000 [2] is described below. At first, color space transformation is applied to the tile-component data. Then a discrete wavelet transformation is used to decompose each image tile into a hierarchy of subbands [3]. Each component is represented as a collection of resolution levels, and each level consists of some subbands. Each subband is further partitioned into code blocks, and entropy coding is performed on each code block independently. The entropy coding scheme in JPEG2000 is based on the classical context-adaptive arithmetic coder. Each bitplane in each codeclock (except the most significant non-zero bitplane from which the coding starts) is encoded in the three passes. The bitstream generated by each pass is referred to as coding pass. After entropy coding a rate control mechanism could be applied.

Rate control in JPEG2000 guarantees that the resulting tile size is not higher than the given threshold value providing image quality is as high as possible. The tile is composed of an integer number of code blocks with numbers $i = 1, 2, 3, \ldots$. Each block is subsequently composed of an integer number of coding passes. Each additional pass contributes some bytes to the block and, in turn, decreases the overall image distortion. Any block may be truncated after any integer number of passes to meet the budget constraint. The task of the rate control is to decide which passes are to be included in the resulting code stream, or, more specifically, to find the set of truncation points $\{n_i\}$ that satisfy the above requirements.

We will proceed with the mathematical formulation of the problem to find the optimal set of truncation points which is also known as the rate-distortion optimization task. Consider $\mathbf{n} = \{n_i\}$ to be *truncation vector*, where n_i signifies that i-th block is truncated in the point number n_i, and $\mathbf{N}$ is the set of truncation vectors. Denote the overall tile distortion after the truncation as $d(\mathbf{n}) = \sum_i d_i^{n_i}$, where quantity $d_i^{n_i}$ corresponds to the distortion value of i-th block truncated in the point number n_i. Similarly, denote resulting tile rate as $r(\mathbf{n}) = \sum_i r_i^{n_i}$. Vectors from the set $\mathbf{N}$ possess the following property: the number of bits for the tile decreases while the distortion value increases. The optimization task is to find $\mathbf{n}_{opt} \in \mathbf{N}$, so that

$$\begin{cases} d(\mathbf{n}_{opt}) = \min_{\mathbf{n}\in\mathbf{N}} d(\mathbf{n}) \\ r(\mathbf{n}_{opt}) \leq r_{max}. \end{cases} \tag{1}$$

In JPEG2000 standard the solution of this optimization task is based on the Lagrange multipliers technique [1] that is used to find the local extreme of a function.

III. VIDEO TRANSMISSION SYSTEM ON JPEG2000 BASIS

Consider the system timing is discrete and slotted. The slot time is a unity of the system time $[t, t+1)$, and time moment t refers to the end of this slot. The video source gives coder a tile at the certain time slots. Coder works in the real-time. After compressing tile into $r(\mathbf{x}_t)$ bits, $\mathbf{x}_t \in \mathbf{N}_t$, coder places it into the *transmitter buffer* [4], [5] at the $[t, t+1)$ slot end. Depending on the number of bits in the transmission buffer, rate-distortion controller forms the requirements for the next tile. The number of bits in the buffer $b(t)$ after placing there a new compressed tile and transmitting over the channel with the constant throughput rate r, changes as follows:

$$b(t) = \max\{0, b(t-1) - r\} + r(\mathbf{x}_t). \tag{2}$$

Solution of the optimization task (1) guarantees, that the transmitter buffer of the size of $B_0 = r$ will never be overflowed if for each transmitted tile the following truncation vector $\mathbf{x}_t \in \mathbf{N}_t$ is chosen, so that

$$\begin{cases} d(\mathbf{x}_t) = \min_{\mathbf{n}\in\mathbf{N}_t} d(\mathbf{n}) \\ r(\mathbf{x}_t) \leq r. \end{cases}$$

Hence, in the transmission system on JPEG2000 basis the number of bits for each compressed tile is close to the channel rate r. If statistical properties of tiles vary a lot, this rate control scheme can bring to the following effect: some compressed tiles could have the unnecessary high visual quality, but the quality of other tiles could be low.

IV. VIDEO TRANSMISSION SYSTEM ON MODIFIED JPEG2000 BASIS

In this paper it is proposed to control not only the number of bits of the compressed tile, but also the visual quality of each tile. In other words, it is suggested to compress tiles with the equal acceptable level of visual quality taking into account the channel throughput constraints. To provide the distortion not more than d it is needed, similarly to (1), to find the truncation vector $\mathbf{x}_t \in \mathbf{N}_t$, so that

$$\begin{cases} r(\mathbf{x}_t) = \min_{\mathbf{n}\in\mathbf{N}_t} r(\mathbf{n}) \\ d(\mathbf{x}_t) \leq d. \end{cases}$$

It is obvious that bit sizes of each tile can vary a lot and exceed the channel throughput r. Therefore, the size of the transmitter buffer $B_0 > r$. Consider that statistical properties of all frames in video sequence vary insignificantly. Formulating the optimization task more exact, it is needed to select the truncation vector $\mathbf{x}_t$ for each tile t, so that

$$\begin{cases} minimize \max_t d(\mathbf{x}_t) \\ b(t) \leq B_0. \end{cases} \quad (3)$$

A. Consecutive search algorithm

If there aren't any memory or computational resources constraints, this task can be solved as follows. Consider $\mathbf{D}$ to be the set of the possible distortion values for all tiles in ascending order, complemented with 0 and ∞. Assume d_i to be the element of this set. The algorithm is staged as follows. At the i-th stage the threshold $\tilde{d} = d_i$ is chosen. For each tile t it is needed to choose truncation vector $\mathbf{x}_t \in \mathbf{N}_t$, so that

$$\mathbf{x}_t = \arg\max_{\mathbf{n} \in \mathbf{N}_t} \{d(\mathbf{n}) : d(\mathbf{n}) \leq \tilde{d}\},$$

and compute the number of bits in the buffer $\tilde{b}(t)$ according to (2). If the buffer of size B_{con} is overflowed at the i-th stage, the same operations will be applied at the stage $i+1$, otherwise the solution is found and the threshold $\tilde{d}$ is the required solution for the optimization task. The algorithm described above is called the *consecutive search algorithm*.

Lemma 1: There is no sequence of truncation vectors, that does not lead to the buffer overflow and has the maximum distortion value less than $\tilde{d}$, where $\tilde{d}$ is the maximum distortion value, found by the consecutive search algorithm.

It is evident from this lemma, that the consecutive search algorithm can be used for solving the optimization task (3).

B. Memory constrained algorithm of rate and visual quality control

The method described above implies no memory constraints. This paper proposes algorithm that allows to find the estimation of $\tilde{d}$ for the memory constrained consecutive search algorithm. Consider $\hat{d}(t)$ to be the estimation of $\tilde{d}$ value and $\hat{d}(0) < \tilde{d}$. It is supposed to estimate $\tilde{d}$ value as follows. All tiles are compressed with the distortion not more than $\hat{d}(t)$ until the number of bits in the buffer will not exceed some threshold B_H. Threshold B_H crossing means that it is impossible to hold the $\hat{d}(t)$ level for the distortion value for the given channel throughput. Consequently, the buffer is firstly emptied and then the estimation of distortion $\hat{d}(t)$ is increased. Algorithm consists of the following three steps.

Step 0. (*Initialization*)

0.1 Set initial value $\hat{d}(0) = d_0$, $t = 0$, $b(0) = 0$.

0.2 Go to step 1.

Step 1. (*Buffer accumulation*)

1.1 $t = t + 1$, $\min\{b(t-1), r\}$ bits are transmitted.

1.2 t-th tile is lossless compressed, the set of truncation vectors $\mathbf{N}_t$ is found.

1.3 Search for $\mathbf{x}_t \in \mathbf{N}_t$, so that

$$\begin{cases} r(\mathbf{x}_t) = \min_{\mathbf{n} \in \mathbf{N}_t} r(\mathbf{n}) \\ d(\mathbf{x}_t) \leq \hat{d}(t). \end{cases}$$

1.4 If $\max\{0, b(t-1) - r\} + r(\mathbf{x}_t) > B_H$ then go to step 2.3.

1.5 Append compressed tile to the transmitter buffer according to the vector $\mathbf{x}_t$ and go to step 1.1.

Step 2. (*Buffer emptying*)

2.1 $t = t + 1$, $\min\{b(t-1), r\}$ bits are transmitted.

2.2 t-th tile is lossless compressed, the set of truncation vectors $\mathbf{N}_t$ is found.

2.3 Compute the number of bits that are needed to transmit the tile with distortion d_{empty} by searching for the vector for $\mathbf{x}_t \in \mathbf{N}_t$, so that

$$\begin{cases} r(\mathbf{x}_t) = \min_{\mathbf{n} \in \mathbf{N}_t} r(\mathbf{n}) \\ d(\mathbf{x}_t) \leq d_{empty}. \end{cases}$$

2.4 Search for $\mathbf{y}_t \in \mathbf{N}_t$, so that

$$\begin{cases} d(\mathbf{y}_t) = \min_{\mathbf{n} \in \mathbf{N}_t} d(\mathbf{n}) \\ r(\mathbf{y}_t) \leq \min\{B_0 - b(t), r(\mathbf{x}_t)\}. \end{cases}$$

2.5 If $\max\{0, b(t-1) - r\} = 0$ then $\hat{d}(t+1) = \hat{d}(t) + \Delta d$ and go to step 1.3.

2.6 Append compressed tile to the transmitter buffer according to the vector $\mathbf{y}_t$ and go to step 2.1.

Lemma 2: Consider that consecutive search algorithm with transmitter buffer of size B_{con} finds the maximum distortion value $\tilde{d}$. Then for the proposed algorithm with threshold $B_H = B_{con}$ and initial value of distortion threshold estimation $\hat{d}(0) < \tilde{d}$, the following inequality $\hat{d}(t) \leq \tilde{d} + \Delta d$ holds true for any time moment t.

V. Experimental results

For practical results was used the reference JPEG2000 codec implementation Jasper, v. 1.701.0. The performance of the discussed algorithm was tested on the set of video sequences. For simplifying results presentation each test video sequence was formed of the identical frames. Figure 1 presents PSNR/tile dependence for the original JPEG2000 algorithm with 1920x8 tiling and the proposed algorithm with 1920x8 tiling and transmitter buffer of size 30kB. Figures 2–4 show PSNR/rate curves for the original JPEG2000 algorithm with and without tiling and the proposed algorithm with 1920x8 tiling and transmitter buffer of size 30kB. For proposed algorithm PSNR/rate dependence varies from frame to frame until the system is adapted to the channel throughput. Therefore, graphs present PSNR/rate dependencies for the first frame and the frame after system adaptation to the throughput channel (for original JPEG2000 algorithm PSNR/rate dependence doesn't vary from frame to frame).

VI. Conclusion

Practical results given above shows that in memory restricted conditions bit rate control mechanism in JPEG2000 algorithm is not suitable for images with synthetic fragments. This is due to that fact that distortion fluctuation for tiles with the equal bit size for synthetic images is much more than for natural images. Therefore, while transmitting images with synthetic fragments it is necessary to control distortion value and use transmitter buffer. This approach significantly improves the visual quality of the transmitted frames with synthetic fragments. For video sequences with natural frames only this approach shows the same results as the original JPEG2000 algorithm.

References

[1] ITU-T and ISO/IEC JTC 1, "JPEG 2000 Image Coding System: Core Coding System, ITU-T Recommendation T.800 and ISO/IEC 15444-1" *JPEG 2000 Part 1*, 2000.

[2] D. Taubman and M. W. Marcellin, "JPEG2000: Image Compression Fundamentals, Practice and Standards" *Massachusetts: Kluwer Academic Publishers*, 2002.

[3] S. Mallat, "A Theory of Multiresolution Signal Decomposition: The Wavelet Representation" *IEEE Transactions on Pattern Anal. Mach. Intell.*,vol. 11, pp. 674-693, 1989.

[4] F. Jelinek, "Buffer overflow in variable-length coding of fixed rate sources" *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 490–501, 1968.

[5] A.R. Reibman, B.G Haskell, "Constraints on variable bit-rate video for ATM networks" *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 2, Issue 4,pp. 361 – 372, 1992.

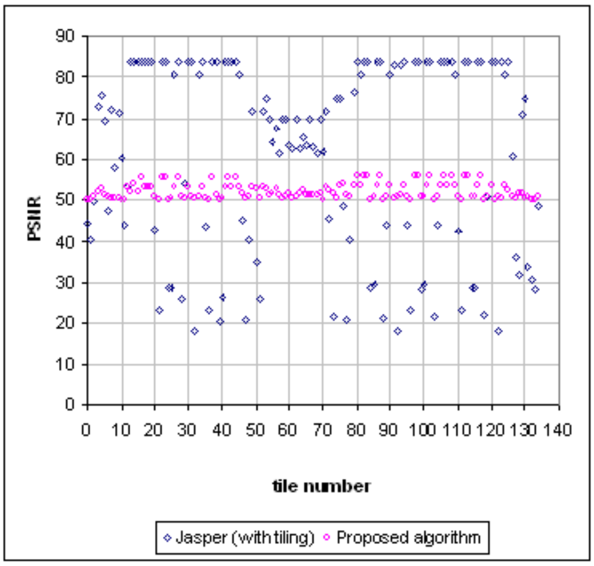

Figure 1. PSNR/tile dependence for the original JPEG2000 algorithm with 1920x8 tiling and the proposed algorithm with 1920x8 tiling and transmitter buffer of size 30kB, rate=0.07

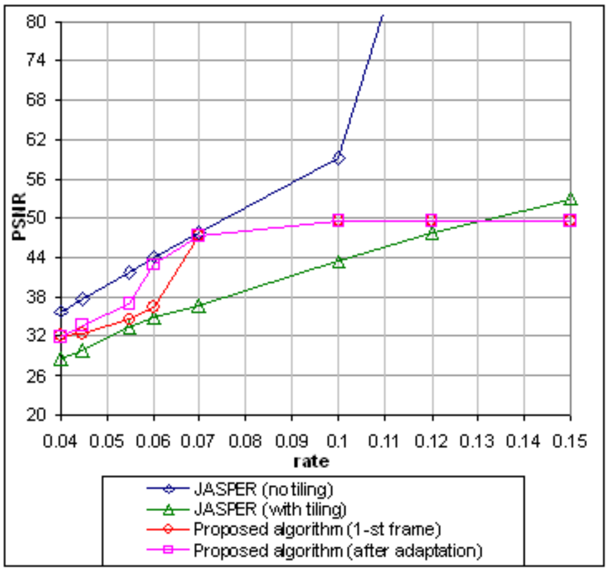

Figure 3. PSNR/rate curves for image with natural and synthetic fragments

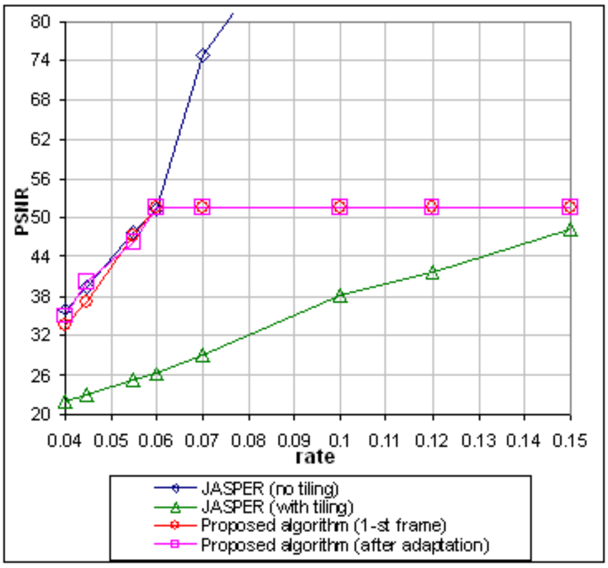

Figure 2. PSNR/rate curves for synthetic image

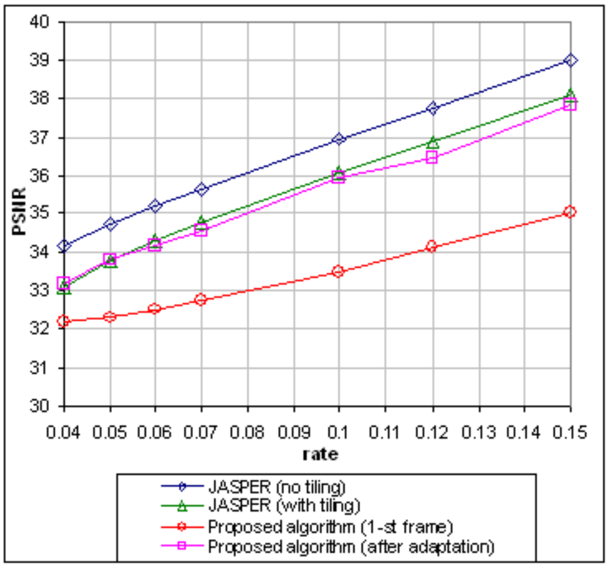

Figure 4. PSNR/rate curves for natural image