Rho-domain based Rate Control Scheme for Spatial, Temporal and Quality Scalable Video Coding

Yohann Pitrey, Marie Babel, Olivier Déforges, Jérôme Viéron

ρ-domain based Rate Control Scheme for Spatial, Temporal and Quality Scalable Video Coding

Yohann Pitrey [a] Marie Babel [a] Olivier Déforges [a] Jérôme Viéron [b]

[a] {yohann.pitrey, marie.babel, olivier.deforges}@insa-rennes.fr
Institute of Electronics and Telecommunication of Rennes (IETR), INSA de Rennes, Rennes, France
[b] jerome.vieron@thomson.net
Video Compression Lab - Thomson R&D France, Cesson-Sévigné, France

ABSTRACT

Rate control is a capital issue in video coding. It allows a regulation of the bitrate out from the encoder, to cope with some network transmission or quality constraints. Scalable Video Coding emerged several years ago as an answer to the growing need of application-adaptable video streams. Although the interest of scalable video coding has been confirmed by recent studies, it can not be used in practical contexts without proper rate control techniques. In this paper we present a new rate control scheme for scalable video, based on a simple yet attractive bitrate modelling framework called ρ-domain. Our scheme performs accurate rate control on spatial, temporal and quality scalabilities, while maintaining a constant PSNR. Inter layer prediction is also handled effectively.

Keywords: video coding and transmission, rate control, MPEG-4 Scalable Video Coding, ρ-domain.

1. INTRODUCTION

Video streaming over communication networks has known a tremendous expansion over the last decade. With numeric television, internet broadcasting and video-capable mobile devices, efficient video coding is becoming a central need. To reduce the amount of data to transmit, video coding techniques such as MPEG-4 AVC/H.264 have been developed. Based on spatial and temporal redundancy removal, together with arithmetic coding, they allow to condense the data in a quite effective way. Nevertheless, today's heterogeneous networks and manifold video-reading devices call on more adaptible video streams. MPEG-4 AVC/H.264 was designed to address any fixed setup and fails to provide flexible video streams. The new MPEG-4 Scalable Video Coding (SVC) standard has been proposed as a response to this need for flexibility. It provides three types of scalability. Spatial scalability acts on the frame resolution, and addresses heterogeneous display devices. Temporal scalability increases the number of frames per second, thus improving the motion smoothness. Quality scalability improves the signal-to-noise ratio of the decoded video stream. A MPEG-4 AVC/H.264-compatible base-quality layer is first encoded, and further enhanced by additional layers. The standard also provides a new tool called inter-layer prediction. This feature allows enhancement layers to use information from a given base layer for prediction, and yields to better coding performances.

As it is based on both spatial and temporal statistical redundancy removal, the performances of video coding can vary, depending on the nature of the sequence to be processed. Thus, the bitrate at the output of the encoder can fluctuate, leading to quality and bandwidth wastage. Rate control is designed to regulate the output bitrate so that it fits to a given constraint. It has been widely studied on MPEG-4 AVC/H.264 and previous video coding standards.[1,2] More recently, few propositions were made for rate control of scalable video.[3,4] In this contribution, we present a new simple yet effective rate control scheme based on the ρ-domain modelling framework.[5] This scheme operates on each type of scalability and is able to deal with inter layer prediction, while keeping a very low computational complexity. The quality of the reconstructed video stream is constant, which is of great importance for human quality perception. The paper first introduces the ρ-domain bitrate representation, then describes the implemented rate control scheme. We further illustrate the performance of our scheme by analyzing some practical results.

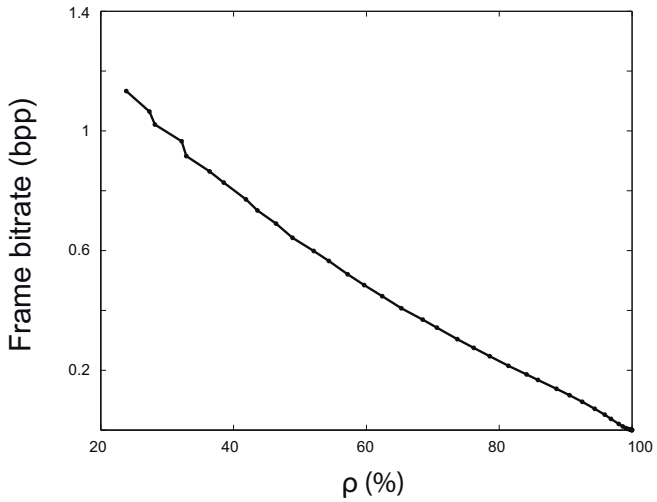

Figure 1. $R(\rho)$ relationship on a hierarchical B frame in the highest enhancement layer of a three-layer spatial-scalable video stream using inter-layer prediction.

2. ρ-DOMAIN BASED RATE MODEL

One of the main issues of rate control is to find an accurate relationship between the bitrate at the ouput of the encoder and the quantization parameter (QP). This way, it is possible to predict the number of bits needed to encode a frame with a given QP, before encoding it. Quantizing a frame introduces a loss from the original data, also known as distorsion. Various rate control approaches try to formulate the relationship between the output bitrate and the distorsion, which are closely related. This relationship can be approximated using a Laplacian or a Gaussian distribution.[6,7] Then, a relationship between the distorsion and the QP is established so that the link between the bitrate and the QP can be found. First introduced in 5, ρ-domain is a slightly different approach. Let ρ be the percentage of zero coefficients in a frame after quantization. The output bitrate linearily decreases when ρ increases (see figure 1).[8] As a relationship can be easily found between ρ and the QP, ρ can be used as an intermediate to get a relationship between the bitrate and the QP. Given the transform coefficients of a frame after the DCT, it is straightforward to determine how many will be coded as zeros after quantization. Let c_{ij}^m be the coefficient at position (i,j) in the macroblock m. This coefficient is coded as a zero if its value is below a dead zone threshold. Due to the adaptive quantization used in H.264, this threshold depends on the position of the coefficient in the transformed macroblock and the value of the QP.[9] The relationship between ρ and the QP can be written as:

$$\rho(q) \;=\; \frac{1}{M}\sum_{m,i,j} z(c_{ij}^m, i, j, q), \tag{1}$$

where M is the total number of coefficients in the frame, and $z(c_{ij}^m, i, j, q) \in \{0, 1\}$ is a function that indicates if a coefficient is under the dead zone threshold (details on the dead zone threshold can be found in 9). Under small QP variations (*i.e.:* below 6), the relationship between ρ and the bitrate can be approximated quite accurately by a linear model. This assumption has been validated for MPEG-2 video streams[5] and for MPEG-4 AVC/H.264.[9] In our previous work,[10] we also validated it for MPEG-4 SVC streams, as presented on figure 1. In 9, the relationship between ρ and the bitrate is established as follows:

$$\rho(R) = \frac{R_0 - R \times (1 - \rho_0)}{R_0}, \tag{2}$$

where R_0 and ρ_0 are initial values which are determined by pre-encoding the frame. Using (1) and (2), we can establish a relationship between the bitrate and the QP which allows us to predict the output bitrate before encoding a frame.

3. RATE CONTROL SCHEME FOR MPEG-4 SVC

A rate control scheme classically contains two modules. First, a budget allocation policy is used to dispatch the available bitrate according to the constraint to fit with. Then, respecting the target budget is handled by a QP selection module. Controlling the bitrate at the output of the encoder can be quite difficult, due to activity variations along the sequence to be processed. When it comes to scalable video, we have to deal with several video layers, interacting with each other using inter-layer prediction. The relationships beween the bitrates from the different layers can be quite complex and make rate control harder to achieve. In this paper we focus on the design of a rate control scheme that is able to control each layer separately. This section describes the rate control scheme we developed for MPEG-4 SVC, based on the model described in section 2.

3.1 Budget allocation

Our goal is to respect a given bitrate-per-second constraint, while aiming for a constant quality in the reconstructed video stream. Quality fluctuation leads to bad user experience and is to avoid. We specify a bitrate constraint for each layer, and layers are handled separately. Our budget allocation policy operates both at group-of-pictures (GOP) and at frame levels.

3.1.1 GOP-level budget allocation

First, we allocate a target budget to each GOP, by simply converting the bitrate-per-second constraint C_l for layer l to a bitrate-per-GOP constraint G_l:

$$G_l = S_l \times \frac{C_l}{F_l} + \delta, \tag{3}$$

where S_l is the size of a GOP in layer l, F_l is the number of frames per second in layer l and δ is a small feedback term to compensate allocation errors over previous GOPs.

3.1.2 Relative frame complexity

Once a target budget is assigned to a GOP, we need to dispatch it among its frames. Each type of frame (*i.e.:* I, P and B) uses different coding tools, and has different coding efficiency. I frames use only intra-frame prediction and are the most reliable. However, their coding efficiency is not as good as P frames which allow inter-frame prediction. B frames allow bidirectional prediction and have even better coding efficiency. To aim for constant quality throughout the GOP, we need to allocate more budget to key-frames (*i.e. :* I, P) than to B frames. We define a complexity measure K_{T_l}, to represent the relative coding efficiency of a given frame type T in a given layer l. To understand the construction of this measure, we need to look at the quantization scheme of H.264, which is the same as in MPEG-4 SVC.[11] After finding a prediction, each residual macroblock is transformed using a DCT. The transformed coefficients are then quantized by a scalar quantizer. A quantized coefficient Z_{ij} can be expressed as :

$$Z_{ij} = round(W_{ij} \times \frac{MF}{2^{qbits}}), \tag{4}$$

where W_{ij} is the corresponding transformed coefficient, MF is a factor given in 11 and

$$qbits = 15 + floor(q/6). \tag{5}$$

As we look for a relative complexity measure, the constant terms in equations (4) and (5) can be discarded, as well as the rounding operations. Then, equation (4) can be reformulated as:

$$2^{q/6} \times Z_{ij} = W_{ij}. \tag{6}$$

The transformed coefficients W_{ij} are a good indicator of the coding complexity of a macroblock. If a good prediction was found, the amount of residual information is small and the transform coefficients contain quite small values. However, if the prediction is bad, the transform coefficients contain large values. Therefore, we use

$2^{q/6} \times Z_{ij}$ as a macroblock complexity measure. At frame-level, the complexity measure is defined as the sum of macroblocks complexities:

$$K_{T_l} = 2^{q/6} \times \sum_{i,j} Z_{ij}. \tag{7}$$

Once quantized, the coefficients are sent to the entropy coder, which is the last step in the encoding process. The number of bits needed to code a frame and the values of the quantized coefficients are thus closely related. As a result, in equation (7) it is possible to replace the quantized coefficients Z_{ij} with the number of bits needed to code it, say b. Thus, the final frame complexity measure can be expressed as follows:

$$K_{T_l} = 2^{q/6} \times b. \tag{8}$$

This complexity measure was designed so that I frames have a higher complexity than P frames, and that P frames have a higher complexity than B frames. For each type of frame in each layer, we handle a separate complexity K_{I_l}, K_{P_l} and K_{B_l}. To take advantage of the hierarchical B frames GOP structure used in MPEG-4 SVC, B frames complexities also depend on their temporal level t, so K_{B_l} is actually $K_{B_{l,t}}$ (but is noted as K_{B_l} for simplicity).

3.1.3 Frame-level budget allocation

The GOP budget is dispatched according to the frames complexities. The target budget for a frame is processed as follows:

$$R_{target} = \frac{K_{T_l}}{K_{TOTAL_l}} \times G_l + \epsilon, \tag{9}$$

where $T \in \{I, P, B\}$ is the type of the frame, K_{TOTAL_l} is the sum of complexities of all frames in the current GOP and ϵ is a small feedback term. As it is not possible to know the complexities of all frames in the GOP in advance, we use weighted mean values as estimations for frames that have not been encoded yet. These mean values are updated as follows after encoding each frame:

$$K_{T_l} = \frac{A \times \bar{K}_{T_l} + B \times \hat{K}_{T_l}}{A + B}, \tag{10}$$

where $\bar{K}_{T_l}$ is the previous mean complexity value, $\hat{K}_{T_l}$ is the complexity of the current frame and A and B are two weighting constants *(for our tests we used $A = 5$ and $B = 3$)*.

3.2 QP processing

Once the budget allocation is performed, the optimal value of QP is computed for each frame. We proceed in two steps. First using (2), we process the target value of ρ, say ρ_{target}, corresponding to the target bitrate R_{target}. Then we search for the QP that generates the closest bitrate to ρ_{target} using (1). This search is performed by computing equation (1) for each value of QP between 0 and 51. To evaluate R_0 and ρ_0 from equation (2), we pre-encode the frame with an initial QP.

4. EXPERIMENTAL RESULTS

We will now analyze the performance of our rate control scheme on MPEG-4 SVC. For each type of scalability, three layers were encoded. Table 1 sums up each tested set. Dyadic scalability was used in the spatial scenario (increase of the frame dimensions by two from one layer to another). For quality scalability, coarse-grain scalability (CGS) was used, and each layer used the lower layer as a reference ID. The GOPs structure was 'PBBB', except for temporal scalability, for which the base layer only contains P frames and the middle layer has a 'PB' structure. Inter layer prediction was activated, allowing a layer to search for motion and texture information in the lower layer. The constraint bitrates for each layer were inspired by the SVC Verification Test Plan,[12] but were not designed to reach equal quality on each layer.

		frame size	frames per second	rate (kbps)	GOP size
SPATIAL	base layer	QCIF	30	160	4
	enh. layer 1	CIF	30	480	4
	enh. layer 2	4CIF	30	1400	4
QUALITY	base layer	CIF	30	300	4
	enh. layer 1	CIF	30	600	4
	enh. layer 2	CIF	30	900	4
TEMPORAL	base layer	CIF	15	240	1
	enh. layer 1	CIF	30	480	2
	enh. layer 2	CIF	60	960	4

Table 1. Test scenarii for each type of scalability.

To evaluate the effectiveness of the rate control, we measure both the allocation error and the signal-to-noise ratio (PSNR) in the reconstructed stream. For this, we report the absolute allocation error on each frame :

$$\Delta = \frac{|R_{target} - R_{real}|}{R_{target}}, \tag{11}$$

where R_{real} is the actual number of bits generated after encoding the frame. Table 2 reports the mean value and standard deviation of Δ and PSNR for each type of scalability.

		HARBOUR				**SOCCER**			
		$\mu\Delta$	$\sigma\Delta$	$\mu PSNR$	$\sigma PSNR$	$\mu\Delta$	$\sigma\Delta$	$\mu PSNR$	$\sigma PSNR$
SPATIAL	base layer	0.62%	13.31	36.83	0.80	1.21%	6.76	39.40	1.98
	enh. layer 1	1.32%	20.66	30.76	0.47	3.18%	8.14	36.05	2.22
	enh. layer 2	4.16%	17.30	26.07	0.32	4.71%	10.31	32.67	2.30
QUALITY	base layer	5.58%	15.51	29.44	0.45	2.76%	11.75	33.41	1.87
	enh. layer 1	2.18%	15.99	32.35	0.42	4.45%	12.84	36.93	2.12
	enh. layer 2	0.73%	11.24	35.08	0.40	2.54%	10.95	39.79	2.32
TEMPORAL	base layer	0.30%	3.78	29.24	0.31	1.83%	18.96	33.94	1.70
	enh. layer 1	0.19%	8.49	27.45	1.41	1.36%	8.16	32.99	1.69
	enh. layer 2	2.67%	9.39	26.98	2.66	1.14%	11.19	31.66	2.27

Table 2. Allocation error and PSNR for each type of scalability.

Table 2 shows that our rate control scheme reaches the constraint bitrate very accurately. The allocation error is below 3% for most encoded layers, and below 5% in worst cases. Inter layer prediction is handled correctly and the method even achieves good performances on SOCCER, which is quite difficult to encode because of the high motion and texture activity observed along the sequence. The PSNR remains quite constant, showing only small variations. Visually, the quality looks constant with no unpleasant effect. On every tested configuration, we observed that the difference in terms of PSNR measured between P frames and B frames inside a GOP is below 1 dB. In our budget allocation scheme, each frame is allocated a number of bits that depends on its relative complexity in front of the other frames in the GOP. The ability of our scheme to reach a constant quality lies on the complexity measure we defined in section 3.1.2, thus proving to be quite accurate. This is confirmed by the graphics in figure 2, which present the allocated and achieved bitrates at frame level for each type of scalability. The saw-toothed shape shows the behaviour of our budget allocation scheme. More bits are granted to P frames than to B frames. Moreover, the highest temporal level B frames are allocated less bits than the lowest ones (the bitrates are displayed in encoding order, so the lowest temporal level B frames appear first). Figure 2 also shows the accurate bitrate matching achieved by our scheme. The final number of bits is very close to the target bitrate for most frames. For a few frames, we can see a mismatch, which is due to abrupt activity changes requiring

large QP variations. They cause important changes in the prediction modes, and make it difficult to predict the output bitrate at frame level.[10] Figure 3 presents the achieved bitrates per second and frame PSNR for each scalability. Once again, the bitrate matches the constraint very closely. The allocation errors we noticed at frame level are taken into account and well compensated by modifying the budget of the following GOPs. The global PSNR variations are due to activity changes in the input sequence. The curve remains quite smooth and shows that locally, frames have a comparable PSNR, regardless to their type.

5. CONCLUSION

In this paper we presented a new rate control scheme for MPEG-4 SVC, based on the ρ-domain framework. This very simple yet efficient rate control scheme manages to respect the specified layer constraints with very small allocation error, and keeps a constant PSNR over the reconstructed sequence. A budget dispatching method is first performed at GOP level, then at frame level. The relative frame types complexities are used to maintain a constant quality throughout the entire GOP. The ρ-rate model is then used to choose a QP value for the whole frame. Tests on spatial, temporal and quality scalabilities show that our scheme is capable of regulating the output bitrate with very little error. Inter layer prediction is handled correctly, which is very promising for scalable video streams. Moreover, the very low computational complexity of our scheme make it almost insignificant in front of the rest of the encoding process. Our scheme can still be enhanced by including some buffer control. Furthermore, choosing a different QP for each macroblock could improve the whole bitrate regulation, at the expense of a dramatic computational complexity increasement. In our future work, we will focus on cross-layer rate control, fully exploiting the inter layer prediction tool. Pre-encoding pass removal and human quality perception shall also be areas of interest.

REFERENCES

[1] ISO/IEC JTC1/SC29 WG11/93-400, "MPEG-2 test model 5," tech. rep., MPEG Commitee (1993).

[2] Li, Z., Pan, F., Lim, K., Feng, G., and Lin, X., "Adaptive basic unit layer rate control for JVT," tech. rep., Joint Video Team (2003).

[3] Xu, L., Ma, S., Zhao, D., and Gao, W., "Rate control for scalable video model," in [*Visual Communications and Image Processing.*], **5960**, 525–534 (2005).

[4] Liu, Y., Li, Z. G., and Soh, Y. C., "Rate control of h.264/AVC scalable extension," *Circuits and Systems for Video Tech., IEEE Trans. on* **18**(1), 116–121 (2008).

[5] He, Z. and Mitra, S., "ρ-domain bit allocation and rate control for real time video coding," *Image Processing, International Conf. on* **3**, 546–549 (2001).

[6] jin Lin, L. and Ortega, A., "Bit-rate control using piecewise approximated rate-distortion characteristics," *IEEE Trans. Circuits Syst. Video Technol* **8**, 446–459 (1998).

[7] Ribas-Corbera, J. and Lei, S., "Rate control in DCT video coding for low-delay communications," *Circuits and Systems for Video Technology, IEEE Trans. on* **9**(1), 172–185 (Feb 1999).

[8] He, Z. and Chen, T., "Linear rate control for JVT video coding," *Information Technology: Research and Education, International Conf. on* , 65–68 (2003).

[9] Shin, I., Lee, Y., and Park, H., "Rate control using linear rate-ρ model for h.264," *Signal Processing - Image Communication* **4**, 341–352 (2004).

[10] Pitrey, Y., Serrand, Y., Babel, M., and Déforges, O., " ρ-domain for low-complexity rate control on MPEG-4 Scalable Video Coding," in [*IEEE International Symposium on Multimedia'08.*], (2008). To be published at the time of submission.

[11] [*H.264 and Mpeg-4 Video Compression: Video Coding for Next-Generation Multimedia*], John Wiley and Sons (2003).

[12] ISO/IEC JTC1/SC29/WG11 MPEG2007/N9189, "Svc verification test plan, version 1," tech. rep., Joint Video Team (2007).

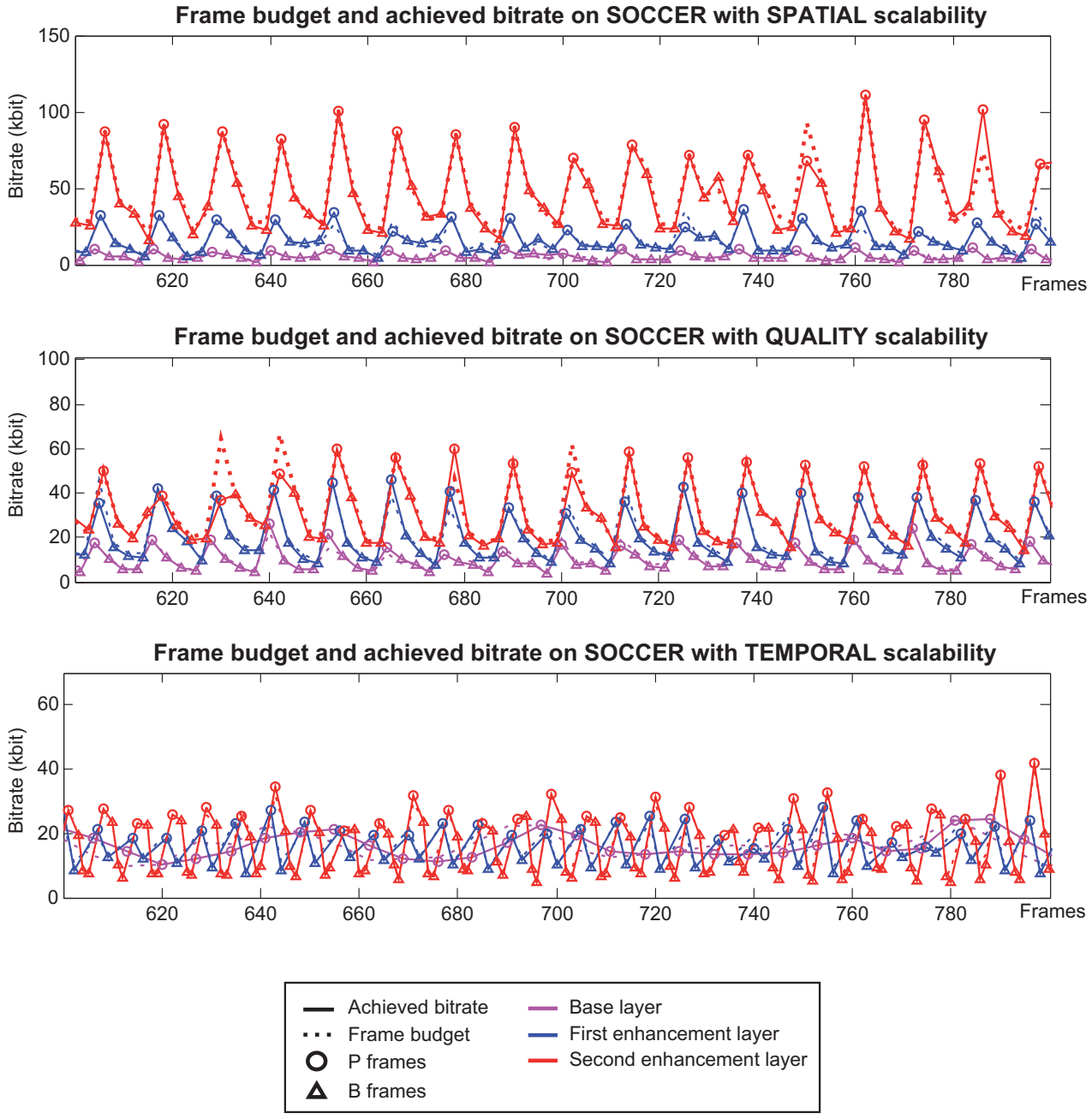

Figure 2. Allocated and achieved bitrates at frame level for each type of scalability.

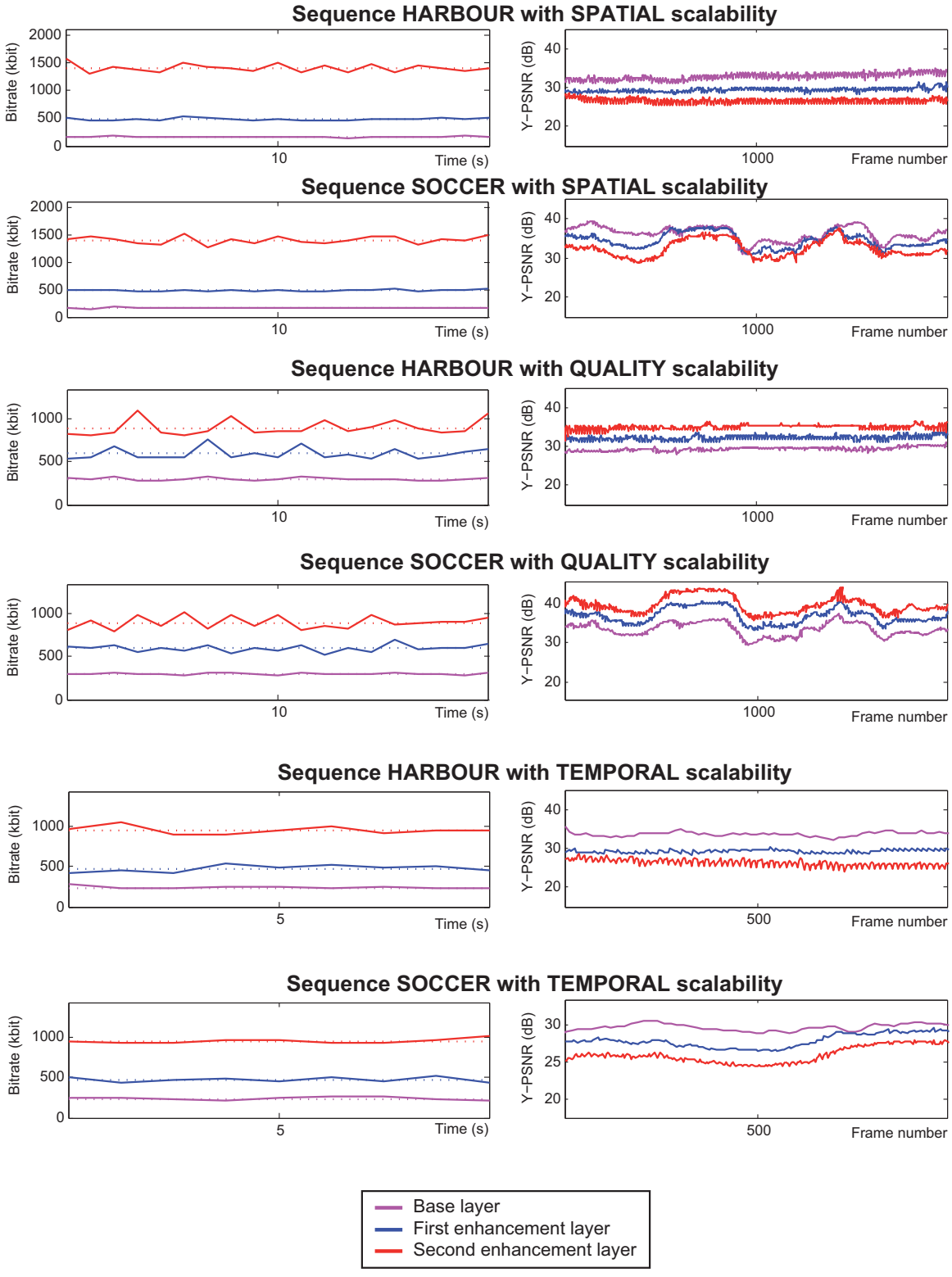

Figure 3. Achieved bitrates per second and PSNR for each type of scalability.