

# RATE-DISTORTION ANALYSIS OF SP AND SI FRAMES

*Eric Setton, Prashant Ramanathan and Bernd Girod*

Information Systems Laboratory, Department of Electrical Engineering  
 Stanford University, Stanford, CA 94305-9510, USA  
 {esetton, pramanat, bgirod}@stanford.edu

## ABSTRACT

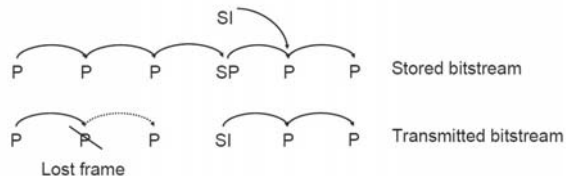
SP and SI frames in the H.264 video coding standard can be used for error resilience, bitstream switching or random access. Despite a widespread interest in these new types of frames, no work so far has investigated, in a systematic way, their rate-distortion efficiency. In this paper, we propose a high-rate model for the rate-distortion performance of SI and SP frames. A comparison to experimental results, obtained with our implementation of an SP encoder, confirms its validity. The model predicts how the relative sizes of SP and SI frames can be traded off. We analyze, both theoretically and experimentally, how this can be used to minimize the transmitted bit-rate when SP frames are used for video streaming with packet losses.

## 1. INTRODUCTION

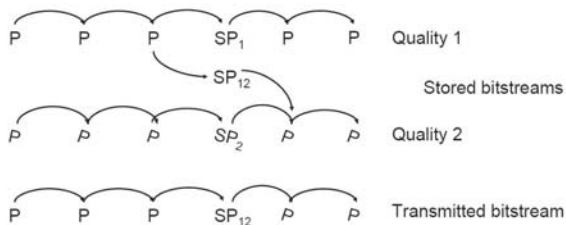
The design of the latest video coding standard, H.264, reflects the increasing need for video streaming solutions which can adapt to varying network conditions. In addition to achieving superior coding efficiency, H.264 uses network-friendly syntax and incorporates several new encoding features which can be taken advantage of when designing flexible and adaptive streaming systems. The new picture types SP and SI are one of these features.

Based on the seminal work by Färber et al. [1], SP and SI frames were proposed in 2001 by Karczewicz and Kurceren, as a solution for error resilience, bitstream switching and random access [2, 3]. They are now part of the extended profile of H.264. The main advantage of this new picture type is that, it can be reconstructed exactly by using different sets of predictors or no predictor at all. This leads to interesting applications, such as refreshing a prediction chain or switching between different streams, as depicted in Fig. 1 and Fig. 2.

Despite a widespread interest in SP and SI frames, no work so far has addressed the question of how efficient SI and SP frames are, and how their relative sizes



**Fig. 1.** SI frames share the instant refresh properties of I frames but are only sent after a frame is lost.



**Fig. 2.** Switching SP frames allow to switch streams using predictive frames only.

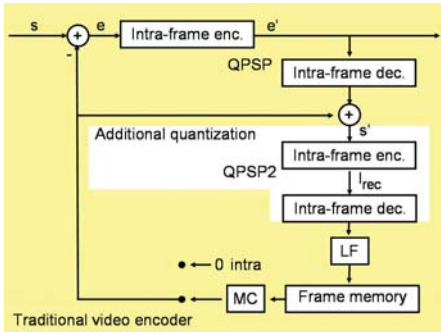
can be traded off? This is, in part, due to the fact that no reference implementation of an SP encoder has been provided to the community. The purpose of this work is to attempt to address these questions by proposing a model for the rate-distortion functions of SP and SI frames. The model is used to analyze the properties of these pictures and derive optimal settings for their encoding.

In the next section, we define switching and non-switching SP frames and describe their encoding. In Section 3, we propose a high-rate model of the rate-distortion performance of SP and SI frames and compare it to experimental results. The model predicts how the relative sizes of SP and SI frames can be traded off. We analyze, in Section 4, both theoretically and experimentally, how this can be used to minimize the transmitted bit-rate when SP frames are used for video streaming with packet losses.

## 2. ENCODING OF SP AND SI FRAMES

Predictively encoded P frames can only be reconstructed exactly when their set of predictors is decoded correctly. To alleviate this requirement, a non-switching (also called primary) SP frame may be inserted in the bitstream as shown at the top of Fig. 1 and 2. Along with this non-switching SP frame, a corresponding SI frame or a switching SP frame may be created. The SI frame can be decoded without any predictor and will correspond exactly to the initial primary SP frame. Likewise, the switching (also called secondary) SP frame, can be decoded with its own set of predictors. Its reconstruction corresponds exactly to the initial primary SP frame.

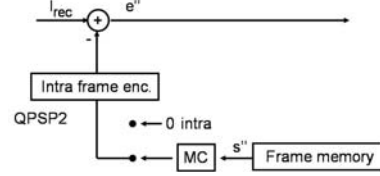
The diagram of a primary SP frame encoder is shown in Fig. 3. It is mainly composed of a traditional video encoder followed by an additional intra-frame encoder<sup>1</sup> which operates on the reconstructed image signal  $s'$ . It is this second quantization stage that allows identical reconstruction from different predictors and provides the switching and restart functionalities of SP frames. This design differs slightly from what was proposed in [2], and is comparable to a later design accepted by JVT [4].



**Fig. 3.** Primary SP frame encoder. Loop-filtering and motion compensation are represented by the symbols LF and MC

The quantized coefficients output by the second intra-frame encoder,  $l_{rec}$ , are subsequently entropy-coded to produce SI frames. For switching SP frames, only the residual of a motion-compensated prediction of  $l_{rec}$  is entropy-coded, as depicted in Fig. 4.

<sup>1</sup>We call intra-frame encoder the combination of a spatial transform followed by quantization. In the figures of the paper, we show next to this block a symbol (either QPSP or QPSP2) representing the value of the quantizer.



**Fig. 4.** Secondary SP frame encoder.

## 3. RD ANALYSIS OF SP AND SI FRAMES

In this section we will explain how the rate-distortion performance of primary and secondary SP frames can be modelled. Our analysis follows the classic model described in [5] for motion-compensated coding.

### 3.1. RD analysis of primary SP frames

We assume that the image signal  $s$  and the prediction error  $e$  are both stationary and jointly Gaussian zero-mean signals. We denote their spatial power spectral density (PSD) respectively by  $S(\Lambda)$  and by  $S_{ee}(\Lambda)$ , where  $\Lambda$  is a vector representing spatial frequency.

As shown in Fig. 3, there is no difference between the encoded residual error signal  $e'$  in a primary SP frame encoder and in a traditional video encoder. Hence, we obtain from [5] the expression for the rate of primary SP frames:

$$R_{SP1} = \frac{1}{8\pi^2} \iint_{\Lambda} \max(0, \log_2(\frac{S_{ee}(\Lambda)}{\theta})) d\Lambda \quad (1)$$

The second intra-frame encoder depicted in Fig. 3 increases the distortion of the reconstructed signal  $s'$ . At high rates, we can assume that the PSD of  $s'$  is close to that of the original signal  $s$ . We further assume that the distortion contributed by the second intra-frame encoder is additive relative to the distortion introduced by the first encoder. Hence, we can express the mean square error distortion of the primary SP frame as a sum of two terms corresponding, respectively, to the distortion contribution of the first and the second intra-frame encoders:

$$D_{SP1} = D_1 + D_2 \quad (2)$$

$$D_1 = \frac{1}{4\pi^2} \iint_{\Lambda} \min(\theta, S_{ee}(\Lambda)) d\Lambda \quad (3)$$

$$D_2 = \frac{1}{4\pi^2} \iint_{\Lambda} \min(\theta_2, S(\Lambda)) d\Lambda \quad (4)$$

In (1)-(4),  $\theta$  and  $\theta_2$  are parameters which take on all positive values to generate the rate-distortion curves.

### 3.2. RD analysis of SI and secondary SP frames

The reconstruction of SI and secondary SP frames corresponds exactly to the primary SP frames they stem from. Therefore, their distortions, denoted  $D_{SI}$  and  $D_{SP_2}$  respectively, are equal to  $D_{SP_1}$ :

$$D_{SP_2} = D_{SI} = D_{SP_1}. \quad (5)$$

As stated in Section 2, the signal  $l_{rec}$  is entropy-coded to produce SI frames. This signal is a compressed version of the signal  $s'$ , encoded by an ideal frame encoder. As the PSD of this signal is assumed to be Gaussian and equal to  $S(\Lambda)$ , the rate function of SI frames is:

$$R_{SI} = \frac{1}{8\pi^2} \iint_{\Lambda} \max(0, \log_2(\frac{S(\Lambda)}{\theta_2})) d\Lambda \text{ bit}. \quad (6)$$

To encode secondary SP frames, the signal  $e''$  is entropy-coded as shown in Fig. 4. This signal is the residual of the motion-compensated prediction of  $l_{rec}$  by a previously compressed frame, denoted  $s''$ , stored in the frame memory. If  $s''$  was compressed at a lower quality than the primary SP frame corresponding to  $l_{rec}$ , the secondary SP frame will serve to switch from a low quality bitstream to a high quality bitstream and vice versa. After motion compensation,  $s''$  is encoded to produce indices which can be subtracted to  $l_{rec}$ , as shown in Fig 4. We will distinguish between two cases depending on whether  $s''$  was quantized more finely or more coarsely than this subsequent encoding.

In the first case, the encoding degrades the signal used for prediction. We will assume that we can neglect the penalty introduced by performing the motion compensation in the transformed and quantized domain. Equivalently, we consider the PSD of the residual signal  $e''$  to be equal to that of the encoded residual of the motion-compensated prediction of  $s'$  by  $s''$ . In this case, the rate is independent of the quality at which  $s''$  was initially compressed and can be expressed as a function of the PSD of the error signal  $e$ :

$$R_{SP_2} = \frac{1}{8\pi^2} \iint_{\Lambda} \max(0, \log_2(\frac{S_{ee}(\Lambda)}{\theta_2})) d\Lambda \text{ bit}. \quad (7)$$

In the second case, the compression of  $s''$  limits the quality of the motion compensated prediction. This increases the variance of the power spectral density of the error signal  $e''$ . We have observed empirically that the size of secondary SP frames, which remains constant in the first case, increases when the compression of  $s''$  is coarser.

### 3.3. Experimental results

Figure 5 shows the rate-distortion performance of SP and SI frames according to (1)-(7). The distortion is represented, in dB, by its SNR. As a reference, the rate-distortion curves of I and P frames, taken from [5], are also represented. All the curves are obtained by letting the parameter  $\theta$  take on all positive values. The settings for  $\theta_2$  are those derived in Section 4. The expressions used for  $S$  and  $S_{ee}$ , are those suggested in [5]. The derivation of  $S_{ee}$  is obtained by assuming an optimal loop filter and a small Gaussian displacement error with variance  $\sigma_{\Delta d}^2 = 0.04 \cdot f_{sx}^{-2}$ , where  $f_{sx}$  is the sampling frequency.

One interesting design parameter is the parameter  $\theta_2$  which controls the trade-off between the rate-distortion efficiency of non-switching SP frames and SI frames. Decreasing  $\theta_2$  leads to smaller primary SP frames but to larger SI and secondary SP frames. The rate-distortion performance of primary SP frames never exceeds that of P frames (with equality when  $\theta_2 = 0$ ). Likewise, the performance of SI frames is limited by that of I frames (with equality when  $\theta = 0$ ).

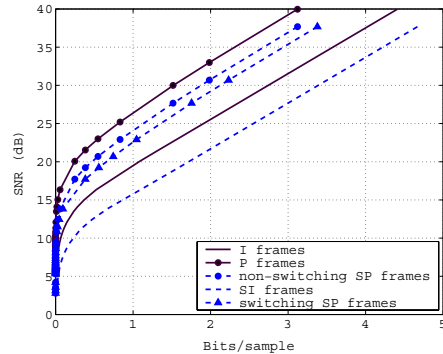


Fig. 5. Theoretical rate-distortion performance.

These theoretical curves correspond to the empirical performance of SP and SI frames shown in Fig. 6. They were obtained by encoding the *Foreman* sequence with our implementation of an SP encoder [6], based on the H.264 codec. In H.264, the counterparts to  $\theta$  and  $\theta_2$  are the two quantization parameters  $QPSP$  and  $QPSP2$ , which control the relative sizes of primary SP frames and SI frames (or secondary SP frames). They are set according to the last column of Table 1.

## 4. OPTIMAL SETTING FOR STREAMING

In this section, the model is used to derive settings of  $QPSP$  and  $QPSP2$  which minimize the expected bit-rate when SP and SI frames are used for streaming with packet losses.

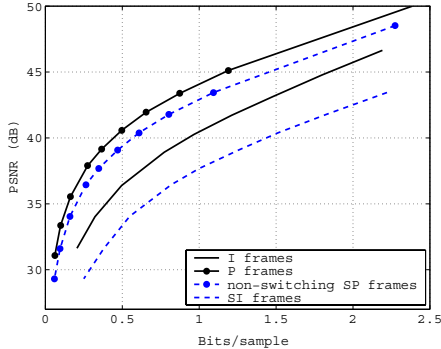


Fig. 6. Experimental rate-distortion performance.

We assume SP frame positions are spaced regularly in the transmitted video stream. At each of these positions, an SI frame can be sent instead of a primary SP frame to stop potential error propagation, as depicted in Fig. 1. One expects this to result in bit-rate savings compared to periodic I frame insertion which occurs regardless of the outcome of previous transmissions. To take full advantage of this effect, we seek an optimal tradeoff between the sizes of SP and SI frames. Depending on the packet error rate and on the spacing of SP frames, different relative proportions of SI and SP frames will be transmitted. We denote  $x$  the probability of transmitting an SI frame at an SP frame position. Minimizing the expected bit-rate is equivalent to minimizing the expected size of a frame sent at an SP position:

$$\mathcal{R} = xR_{SI} + (1-x)R_{SP1}, \quad (8)$$

In our model (1)-(7), for a given quality and a given probability  $x$ , (8) is minimized for one specific value of the parameters  $\theta^*$  and  $\theta_2^*$ . By determining these optimal values for different qualities, we observed a simple linear relation between these two parameters:  $\theta_2^* = \alpha\theta^*$ , where  $\alpha$  is an increasing function of  $x$ .

The counterpart to this relation in H.264 is an equation relating the two quantization parameters  $QPSP$  and  $QPSP2$ . It can be obtained by observing that  $\theta^*$  can be written as a function of the Lagrange multiplier  $\lambda$  which trades off rate and distortion when encoding a P frame, with a quantizer  $QPSP$  in H.264. Likewise,  $\theta_2^*$  can be written as a function of the Lagrange multiplier  $\lambda_2$  which trades off rate and distortion when encoding an I frame, with a quantizer  $QPSP2$  in H.264. As  $\lambda$  and  $\lambda_2$  can be expressed as functions of  $QPSP$  and  $QPSP2$ , we finally get the corresponding relation which is a simple offset:

$$\lambda = 2\log(2) * \theta^*, \quad (9)$$

$$\lambda_2 = 2\log(2) * \theta_2^*, \quad (10)$$

$$\lambda = 0.85 * 2^{\frac{QPSP-12}{3}}, \quad (11)$$

$$\lambda_2 = 0.85 * 2^{\frac{QPSP2-12}{3}}, \quad (12)$$

$$QPSP2 = QPSP + 3\log_2(\alpha). \quad (13)$$

Table 1 indicates the actual settings for  $QPSP$  and  $QPSP2$  as a function of the quantization parameter  $QP$ , used for P frames. Note that  $QPSP$  and  $QPSP2$  need to be integers; this limits the range of their values.

$x$	$\leq 0.1$	$\geq 0.1$ and $\leq 0.19$	$\geq 0.19$
QPSP	$QP - 1$	$QP - 2$	$QP - 3$
QPSP2	$QP - 10$	$QP - 5$	$QP$

Table 1. Optimal settings for QPSP and QPSP2, for different probabilities of transmitting an SI frame

## 5. CONCLUSION

We analyze the rate-distortion performance of SP and SI frames. The encoding process of non-switching SP frames, SI frames and switching SP frames is described, and equations are derived to approximate, at high rates their efficiency. Experimental results, obtained with our implementation of an SP encoder, based on H.264, validate the theoretical results. The model predicts the relative sizes of SP and SI frames can be traded off. We apply the model to determine the optimal tradeoff which minimizes the expected bit-rate for streaming. Finally, we derive corresponding practical settings for the encoding of SP and SI frames.

## 6. REFERENCES

- [1] N. Färber and B. Girod, "Robust H.263 Compatible Video Transmission for Mobile Access to Video Servers," *Proc. ICIP-97, Santa Barbara, CA, USA*, vol. 2, pp. 73–76, Oct. 1997.
- [2] M. Karczewicz and R. Kurceren, "A Proposal for SP-Frames," in *VCEG-L27, ITU-T VCEG Twelfth Meeting: Eibsee, Germany*, Jan. 2001.
- [3] M. Karczewicz and R. Kurceren, "The SP- and SI-frames design for H.264/AVC," *IEEE Trans. CSVT*, vol. 13, no. 7, pp. 637–644, July 2003.
- [4] X. Sun, S. Li, F. Wu, J. Shen, and W. Gao, "The improved SP frame coding technique for the JVT standard," *Proc. ICIP, Barcelona, Spain*, vol. 3, pp. 297–300, Sept. 2003.
- [5] B. Girod, "The Efficiency of Motion-Compensating Prediction for Hybrid Coding of Video Sequences," *IEEE JSAC*, vol. 5, no. 7, pp. 1140–1154, Aug. 1987.
- [6] "H.264 SP frame codec," [http://www.stanford.edu/~esetton/H264\\_2.htm](http://www.stanford.edu/~esetton/H264_2.htm).