METHODS TO IMPROVE CODING EFFICIENCY OF SP FRAMES

Wai-tian Tan, Bo Shen

Mobile and Media Systems Lab, Hewlett Packard Laboratories

ABSTRACT

SP-frame is a new picture type supported by H.264, and supports functions such as rate-switching and random-access. In this paper, we investigate several complementary methods to improve the coding efficiency of SP-frames. We show that by appropriately choosing reference pictures, the size of secondary SP frames can be reduced by up to 40% and 2% for random-access and rate-switching, respectively. We also demonstrate that a simple rule exists that allows the joint selection of the two quantization parameters associated with SP frames to minimize "requantization" error. Results shows 0.1 dB PSNR improvement over comparable choices.

***Index Terms*—** video coding, quantization

1. INTRODUCTION

The SP-frame is a new picture type in H.264. There are two types of SP frames: primary and secondary. An example is shown in Figure 1, where the reconstructed picture of the primary SP frame can be *perfectly* reconstructed by a secondary SP frame *SSP*, even when *SSP* has a different reference frame. Readers interested in learning how such properties are achieved are encouraged to read [1].

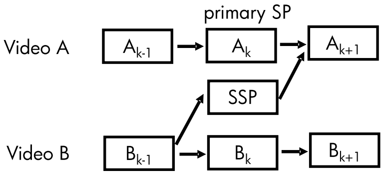


Fig. 1. A secondary SP frame (SSP) can perfectly reconstruct $\hat{A}_k$, the reconstructed frame of A_k, from $\hat{B}_{k-1}$. The frame A_k has to be coded as a primary SP frame.

There are two important applications that benefit from SP-frames, namely:

- *rate-switching*, where a single content is compressed at two-different bit-rates, and SP-frames are used to allow switching from the high-rate to the low-rate streams, and vice versa.
- *random-access*, where SP-frames are used to allow skipping of frames within a single stream.

Traditionally, intra-frames are employed for both applications above. For example, it is typical for DVD to contain 3 intra frames per second to facilitate random-access. Nevertheless, for many lower bit-rate applications, it is impractical to insert large amounts of intra frames. The relatively smaller sizes of SP frames therefore make it a suitable candidate for the applications above.

In this paper, we present methods to improve coding efficiencies of SP frames. To this end, we partition our study into two related parts. The first part is covered in Section 2, and relates to choosing appropriate reference frames for mode decision and motion compensation during encoding of secondary SP frames. The second part is covered in Section 3, and relates to the joint selection of the two quantization parameters, QP and Q_s that are associated with the encoding of SP frames. Specifically, we show that there is a definite advantage in choosing $Q_s = QP - 6$ compared to its immediate neighborhood. Evaluation of our proposed methods is then presented in Section 4, followed by a summary in Section 5.

2. REDUCING SIZE OF SECONDARY SP FRAMES

The basic relationship between primary and secondary SP frames is shown in Figure 1. Generally, there are encoding parameters that jointly affect the compressed sizes of the primary and secondary SP frames. It is outside the scope of this paper to consider the joint encoding of primary and secondary SP frames. We note, nevertheless, that *after* the primary SP frame is produced, a secondary SP frame must be coded to match an exact target frame based on an exact reference frame. Specifically, in the example of Figure 1, *SSP* must perfectly reconstructs $\hat{A}_k$ from $\hat{B}_{k-1}$. There can be no rate-distortion trade-off, and it is desirable to use as few bits in *SSP* as possible. We will next describe a scheme to achieve smaller compressed sizes for secondary SP frames, given primary SP frames. It is based on improved methods to determine the coding mode and motion vectors when coding secondary SP frames.

In the notation of Figure 1, a conventional implementation for computing the coding modes and motion vectors for the secondary SP frame (*SSP*) is based on using $\hat{B}_{k-1}$ and B_k as the reference and target frames, respectively. We write *mode(*$\hat{B}_{k-1}$*,* B_k*)*, *mv(*$\hat{B}_{k-1}$*,* B_k*)*, and we call such strategy *baseline*. One advantage of *baseline* is that the two quantities above are free by-products when compressing frame B_k. However, *baseline* is a compromise since the purpose of *SSP* is to reconstruct $\hat{A}_k$ rather than B_k. The alternative scheme to yield better compression performance, which we call *recon*, avoids the reuse of coding modes and motion vectors. Instead, the actual target $\hat{A}_k$ is used as a target frame for determining coding modes *mode(*$\hat{B}_{k-1}$, $\hat{A}_k$*)* and motion vector *mv(*$\hat{B}_{k-1}$, $\hat{A}_k$*)*, Clearly, the size reduction achieved by *recon* over *baseline* for *SSP* depends on the similarity between $\hat{A}_k$ and B_k. As we shall see later in Section 4, for *rate-switching* application, where A_k and B_k correspond to the same video frame, the difference between $\hat{A}_k$ and B_k is small, and thus the improvement of *recon* over *baseline* is marginal. For *random-access* application however, A_k and B_k correspond to different frames in the same video, and can be significantly different. Under such cases, *recon* can significantly outperform *baseline*.

3. SELECTION OF Q_S

We now investigate the impact of the selection of Q_s on the coding efficiency of the SP frames. H.264 (and many other compression standards) uses uniform scalar quantizer for quantization. A uniform scalar quantizer Q can be modeled as follows. Given input x, quantization output q is produced as

$$q = Q(x) = \begin{cases} sign(x)\left\lfloor \frac{|x|}{s} + \varepsilon \right\rfloor, & \frac{|x|}{s} + \varepsilon > 0 \\ 0, & \text{otherwise} \end{cases}, \quad (1)$$

where s is the quantizer step size, and ε controls the size of the deadzone. The deadzone control factor ε is often within [0,1/2].

The encoding of SP frames involves a $requantization$ process. Specifically, coefficients are first quantized with a first quantizer (generally decided by step size s_1, and deadzone factor ε_1), reconstructed and subsequently quantized by a second quantizer (s_2, ε_2).

For the requantization process, we focus on the more common case when $s_2 > s_1$. That is, a finer quantizing is carried out followed by a coarser quantizing. This case matches the encoding of primary SP frames [1], in which s_1 and s_2 are the quantization step size corresponding to Q_s and QP respectively. With minor loss of generality, we can write $\varepsilon_1 = 1/e_1$, and $\varepsilon_2 = 1/e_2$, where e_1 and e_2 are positive integers.

Clearly, the quantization error through the requantization process is mostly decided by the coarser quantizer. But, the quantization error can be different had the first quantization not happened (*direct quantization*), even though it uses a finer step size. Therefore, we want to evaluate the *requantization error* which is defined as the difference between the results from the direct quantization (by s_2) and the requantization.

Observing that when the boundaries of the quantization bins of the second quantizer perfectly align with that of the first quantizer, the requantization produces identical results as direct quantization. Through careful derivation based on this simple observation, we outline below the conditions that lead to the zero requantization error.

ε_1	r	example r
0	$6k$	6, 12, ...
1/6	$6k+1$	1, 7, 13, ...
1/4	–	–
1/3	$6k+2$	2, 8, 14, ...
1/2	$6k+3$	3, 9, 15, ...

Table 1. Example step size ratio that leads to zero requant error when $\varepsilon_2 = 1/6$.

For a first quantizer with deadzone controlling factor ε_1 and step size s_1 and a second quantizer with deadzone controlling factor ε_2 and step size s_2, there is zero requantization error if and only if both of the following are true:

$$r = \frac{s_2}{s_1} = e_2 k + \frac{e_2}{e_1} \quad (2)$$

for some non-negative integer k, and

$$\frac{e_2}{e_1} \text{ is an integer} \quad (3)$$

For the rest of our discussion, we will assume ε_2 to be 1/6 (e_2 = 6), the suggested value for inter mode in the H.264 encoder implementation [2]. Table 1 lists some example values that achieves zero requantization error. From (2) and (3), we know that r has to be an integer. When $\varepsilon_1 = 1/4$, e_2/e_1 = 6/4 is not an integer, and zero requantization error cannot be achieved. We evaluate the cases when $\varepsilon_1 = 1/3$, in which zero requant error is achieved when $r = 2$, According to the quantization scheme specified in H.264 [3], $r = 2$ indicates the difference between quantization parameters Q_s and QP is 6. We will verify this in the experiment section, where it shows that for these special cases, we can get 0.1 dB gain in PSNR when using Q_s=QP$-$6 than other values around it.

The implication of this discovery is as follows. When we have to choose certain Q_s for encoding SP frames, careful selection of the quantizers' deadzone size and quantization parameters would yield better rate-distortion tradeoff than the points around it. Being aware of this fact allows better design of the overall encoding system.

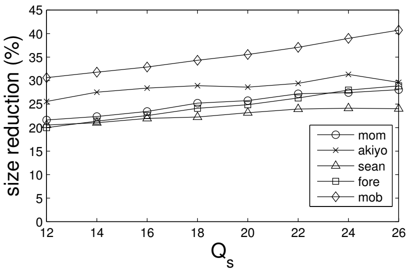
4. EXPERIMENT AND RESULTS

Five QCIF video sequences at 10 frames per second and 10 seconds duration are used for evaluation purposes: Akiyo, Foreman (fore), Mobile Calendar (mob), Mother and Daughter (mom), and Sean. We use the JM-10.1 modified by Eric Setton at Stanford University for production of primary and secondary SP frames.

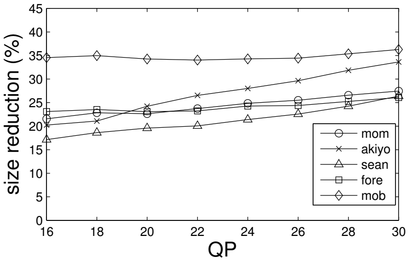
4.1. Selection of reference

Focusing on two main applications of SP frames, namely random seeking and bit rate switching, we investigate the coding efficiency of SSP frames based on our proposed selection of references.

Compare *recon* and *baseline* for *random-access*: For the *random-access* application, we code a primary SP frame every 6 frames, and we produce secondary SP frames that "skips over" 2 frames. In the context of Figure 1, we have $A_i = B_{i-2}$. We perform two experiments. In the first, all P and primary SP frames are encoded using a fixed QP of 25, and we vary the Q_s parameter. In the second, we fixed Q_s to be 18, and vary the QP for all P and primary SP frames. The reduction in size of secondary SP frames achieved by *recon* over *baseline* is shown in Figure 2. We see that size reduction of about 15 to 40% can be achieved. Similar gains are also achieved when we "skip over" 4 frames instead of 2, i.e., $A_i = B_{i-4}$. The large gain highlights the importance of using the proper target frame for mode and motion estimation, when the target frames for the two schemes can be very different.



(a) Fixed QP = 25.

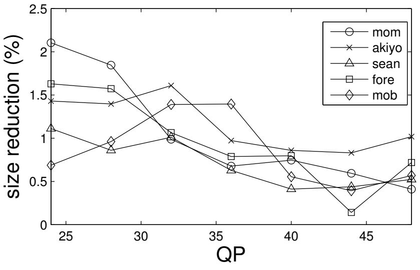


(b) Fixed Q_s = 18

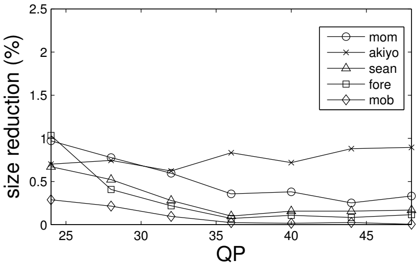
Fig. 2. Size reduction of secondary SP frame achieved by *recon* over *baseline*.

Compare *recon* and *baseline* for *rate-switching*: We next compare the *recon* and *baseline* schemes for the *rate-switching* application, where $A_i = B_i$. Even though based on the same sequence, A and B are coded using QP_1 and QP_2, respectively. We perform two experiments. In the first experiment, we fixed QP_1 and Q_s to be 20 and 17, respectively. We then vary QP_2 from 24 to 48. The results are shown in Figure 3. We see that only a modest size reduction of about 1% is achievable by *recon* over *baseline*. This is due to the similarity of frames B_k and $\hat{A}_k$ in *rate-switching* application.

In the second experiment, we fixed $QP_2 - QP_1$ to be 10, and vary QP_1. The parameter Q_s is set to $QP_1 - 3$. The results are shown in Figure 4. Again, only a gain of about 1% is observed as noted before. Furthermore, we noticed that there is generally higher gain as QP_1 increases. This is due to the larger difference between B_k and $\hat{A}_k$ when QP_1 (and thus QP_2) increases. The gain is modest though, reaching slightly over two percent.

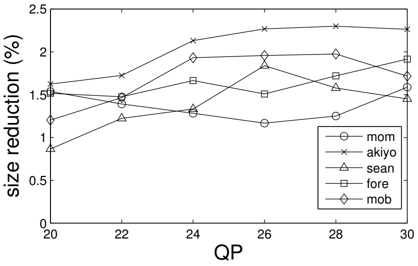


(a) Switching from stream A (high-rate) to stream B.

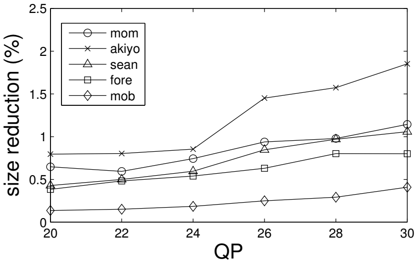


(b) Switching from stream B (low-rate) to stream A.

Fig. 3. Size reduction of secondary SP frame achieved by *recon* over *baseline* for different QP_2. (QP_1=20, Q_s = 17).

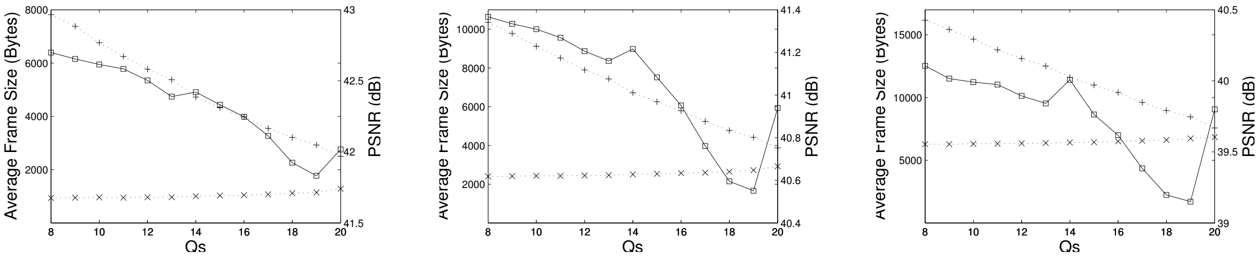


(a) Switching from stream 1 (high-rate) to stream 2.



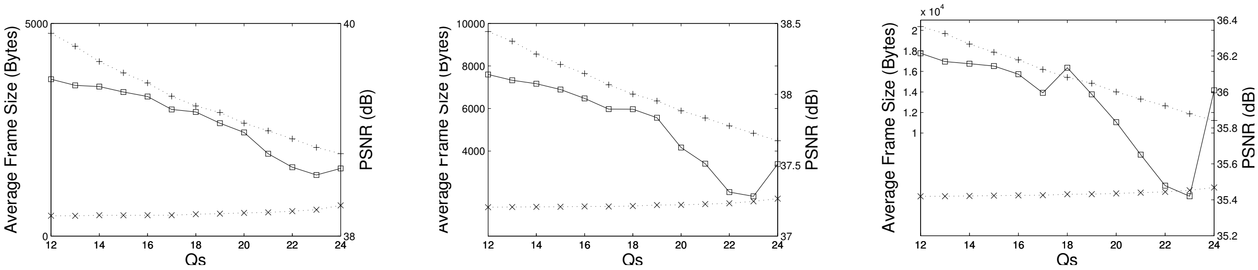
(b) Switching from stream 2 (low-rate) to stream 1.

Fig. 4. Size reduction of secondary SP frame achieved by *recon* over *baseline* for different QP_1. ($QP_2 = 10 + QP_1$, and $Q_s = QP_1 - 3$).



(a) Mom. (b) Foreman (c) Mobile

Fig. 5. PSNR (solid) and the size of primary (x) and secondary (+) SP frames for Rate switching from QP_1=8 to QP_2=20.



(a) Mom. (b) Foreman (c) Mobile

Fig. 6. PSNR (solid) and the size of primary (x) and secondary (+) SP frames for Random seeking for QP=24.

4.2. Selection of Q_s

To evaluate the effect of Q_s selection on the coding efficiency of SP frames, we have modified the JM codec to use $\varepsilon_1 = 1/3$ and carry out the SP frame encoding for two different types of applications.

Bit rate switching: Considering a bit rate switching application from a high rate (QP_1=8) to a lower rate (QP_2=20), we investigate the rate and distortion behavior by selecting Q_s in between QP_1 and QP_2. Results for three test sequences are shown in Figure 5. PSNR is the solid curve and doted curves are the average byte count per primary SP (x) and second SP frame (+).

It is clearly seen that when $Q_s = QP_2 - 6$, that is, when the step size corresponding to QP_2 is twice that of Q_s, we obtain a PSNR higher than its both neighbors. For all three sequences, the PSNR is 0.1-0.2 dB higher.

Note that the plots also reveal that it is not desirable to select Q_s as $QP_2 - 1$ to $QP_2 - 3$. Clearly when $Q_s = QP_2$, there is less requantization error, so $Q_s = QP_2$ is a better choice, just as $Q_s = QP_2 - 6$ is a better choice, than its neighboring values.

Random seeking: Considering a different application in which SP frames are produced for random seeking for a sequence coded with $QP = 24$, we investigate the rate and distortion feature by selecting Q_s from $QP - 12$ to QP. Again, the results for three test sequences are shown in Figure 6. Line style and marker type follow the same as that in Figure 5.

In this random seeking application, we also see the peak PSNR effect at Q_s half the step size of QP case (i.e., $Q_s = QP - 6 = 18$). This effect is more obvious for Mobile sequence. The reason can be that at low bit rate, more complex sequences lead to more non-zero coefficients so that the requantization effect is more pronounced. Again, these plots also confirm that it is better to choose $Q_s = QP$ or $Q_s = QP - 6$ than their corresponding neighboring values.

5. SUMMARY

In this paper, we have presented two methods to improve coding efficiency of SP frames. First, by appropriately choosing reference frames for mode selection and motion compensation, we have shown that the size of secondary SP frames can be reduced by up to 40%. Second, we have shown that it is advisable to always choose $Q_s = QP$ or $QP - 6$ due to smaller requantization errors associated with these choices, leading to 0.1 to 0.2 dB gain over neighboring choices.

6. REFERENCES

[1] M. Karczewicz and R Kurceren, “The SP- and SI-Frames Design for H.264/AVC,” *IEEE Trans. Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 637–44, July 2003.

[2] A. Hallapuro, M. Karczewicz, and H. Malvar, “Low complexity transform and quantization – Part I: basic implementation,” *ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6 Document JVT-B038*, Jan. 2002.

[3] Joint Video Team of ISO/IEC MPEG and ITU-T VCEG, “Joint Final Committee Draft of Joint Video Specification (ITU-T Rec. H.264 — ISO/IEC 14496-10 AVC),” Aug. 2002.