Fast Reference Frame Selection Method Based on Best Reference Frame Index Correlation

Hyungwook Kim, Sojeong Lim, and Sungwook Yu

This letter presents a very simple yet very effective solution for fast reference frame (RF) selection in H.264/AVC. By efficiently making use of the correlation between the best RF indices in various inter modes, the proposed method significantly reduces the number of RFs to be examined at the expense of a very small miss rate. Simulation results show that the proposed method not only improves upon the coding performance of conventional methods but also reduces the encoding time significantly.

Keywords: H.264/AVC, multiple reference frames, best reference frame index.

I. Introduction

The H.264/AVC video coding standard has been widely adopted in many applications [1], [2]. To improve the coding performance, it utilizes multiple reference frames (RFs). The use of multiple RFs becomes especially effective when there are repetitive motions, uncovered backgrounds, shadow and lighting changes, or noises in the source signal [3]. However, motion estimation is the most time consuming process for an H.264 encoder, and the computational complexity increases linearly in proportion to the number of RFs being used.

Thus, many fast RF selection methods that examine only some of the RFs among all RF candidates have been proposed [4]-[9]. Most of these methods make use of spatial and/or temporal correlation to skip the examination of some RFs. For example, Wang and others exploited the motion activity and the best RF (BRF) information of spatially neighboring blocks in RF selection [4]. Shen and others checked if all neighboring blocks had the same BRF to determine whether to skip the examination of the remaining RFs [5]. Wu and Xiao made use of spatial or temporal correlation depending on the partition size of the current inter mode [6]. The RF selection method by Aysu and others is another example that uses both spatial and temporal correlations [7]. On the other hand, some RF selection methods exploit the motion vector information during the RF selection process [8], [9].

Most of these RF selection methods make use of spatial correlation between neighboring macroblocks, temporal correlation between adjacent RFs, or the correlation between motion vectors and RFs. However, this letter presents a new approach that makes use of the correlation between the BRF indices in various inter modes of different block sizes. The proposed method is simple, but it is very effective since it dramatically reduces the number of RFs to be examined at the expense of a very small miss rate.

Manuscript received Aug. 10, 2013; revised Nov. 5, 2013; accepted Nov. 8, 2013.
This research was supported by the Chung-Ang University Research Grants in 2008.
Hyungwook Kim (hyungwook@outlook.com), Sojeong Lim (lim.sojeong@gmail.com), and Sungwook Yu (corresponding author, sungwook@cau.ac.kr) are with the School of Electrical and Electronics Engineering, Chung-Ang University, Seoul, Rep. of Korea.

II. Proposed RF Selection Method

H.264 has the following inter-prediction modes: P16×16, P16×8, P8×16, and P8×8 [1], [10]. In the P8×8 mode, each 8×8 block can be partitioned further into 8×8, 8×4, 4×8, or 4×4 blocks. Thus, the P8×8 mode collectively refers to smaller block size inter modes, such as SMB8×8, SMB8×4, SMB4×8, and SMB4×4 modes [1], [10]. As mentioned, the H.264 standard supports multiple RFs, and the joint model (JM) reference software examines five RFs by default [10]. To select the best motion vector and the BRF, H.264 computes the rate-distortion (RD) cost, which is defined as follows:

$$J_{\text{motion}}(\text{mode, ref}_n, \text{MV}) = \text{SAD}\left(\mathbf{s}, \mathbf{r}\left(\text{mode, ref}_n, \text{MV}\right)\right) + \lambda_{\text{motion}} \cdot R(\text{mode, ref}_n, \text{MV}). \quad (1)$$

Table 1. Probability (%) that ref_n is BRF in each inter mode.

	ref_0	ref_1	ref_2	ref_3	ref_4
P16×16	82.8	8.0	4.8	2.3	2.1
P16×8	82.2	8.3	5.2	2.3	2.1
P8×16	81.9	8.6	5.2	2.3	2.1
P8×8	82.1	8.4	5.4	2.2	1.9
Average	82.3	8.3	5.2	2.3	2.0

Table 2. Probability (%) that ref_n becomes BRF in another mode when ref_n is BRF in P16×16 mode.

	ref_0	ref_1	ref_2	ref_3	ref_4
P16×8	93.2	55.3	53.8	45.4	48.1
P8×16	92.8	54.9	53.7	45.6	46.3
P8×8	90.4	39.0	37.7	27.4	28.5
Average	92.1	49.7	48.4	39.5	40.9

In (1), mode, ref_n, and MV denote the prediction mode, the n-th RF, and the motion vector, respectively. Also, **s** and **r** represent the source and reference blocks, respectively. Finally, SAD, λ_{motion}, and R denote the sum of absolute difference measure, a Lagrange multiplier, and the rate cost, respectively.

As is well known, a recent RF is more likely to be selected as the BRF than an old RF [8]. Table 1 shows the probability that ref_n is the BRF in each inter mode, which is obtained from simulation results with numerous test video sequences. As expected, ref_0 (that is, the most recent RF) has the highest probability, regardless of the inter mode.

In examining the simulation results, we discover that the BRF indices between various inter modes are highly correlated. Table 2 shows the probability that ref_n becomes the BRF in the P16×8, P8×16, and P8×8 modes when ref_n is the BRF in the P16×16 mode. It is clear that the hit rates for ref_0 are quite high for all of these modes. For example, if the BRF of the P16×16 mode is ref_0, the probability that ref_0 is also the BRF in the P16×8 mode is 93.2%. (There are two 16×8 blocks in a 16×16 macroblock, and the BRF of one 16×8 block may be different from that of the other. Thus, the BRF of each 16×8 block is treated separately in the hit rate calculation reflected in Table 2.)

On the other hand, the hit rates for ref_n ($n\geq 1$) are significantly lower than the hit rate for ref_0. For example, if the BRF of the P16×16 mode is ref_4, the probability that ref_4 is also the BRF in the P16×8 mode is only 48.1%. However, it should be emphasized that this value is significantly higher than 2.1%, which is the probability that ref_4 is the BRF in the P16×8 mode, as shown in Table 1. That is, the conditional probability that ref_4 is selected as the BRF when the BRF of the P16×16 mode is ref_4 is considerably higher than the simple probability that ref_4 is selected as the BRF. As shown in Tables 1 and 2, this property applies to all ref_n ($0\leq n\leq 4$). That is, the BRFs of the P16×8, P8×16, and P8×8 modes are highly correlated with the BRF of the P16×16 mode. Thus, it may seem efficient to check only the BRF in the P16×16 mode when the P16×8, P8×16, and P8×8 modes are examined. However, the average hit rates for ref_n are below 50% when $n\geq 1$ (Table 2). This means that it is not sufficient to examine only one RF (that is, the BRF in the P16×16 mode) in the P16×8, P8×16, and P8×8 modes.

Table 3. Probability (%) that one of ref_0, ref_1, …, or ref_n becomes BRF in each mode when ref_n is BRF in P16×16 mode.

	ref_0	ref_1	ref_2	ref_3	ref_4
P16×8	93.2	89.5	93.7	93.8	100.0
P8×16	92.8	89.4	93.6	94.0	100.0
P8×8	90.4	87.3	93.4	94.1	100.0
Average	92.1	88.7	93.6	93.9	100.0

To solve this problem, the proposed method uses the aforementioned observation that a recent RF is more likely to be selected as the BRF. Thus, in the P16×8, P8×16, and P8×8 modes, the proposed method checks not only the BRF in the P16×16 mode but also all of the more recent RFs. For example, if ref_2 is the BRF in the P16×16 mode, the proposed method examines not only ref_2 but also ref_0 and ref_1 in the P16×8, P8×16, and P8×8 modes. Thus, the proposed method checks ($BRFI_{16}$+1) RFs, whose RF indices range from 0 to $BRFI_{16}$, where $BRFI_{16}$ represents the BRF index in the P16×16 mode. Table 3 shows the probability that one of ref_0, ref1, …, or ref_n becomes the BRF in each mode when ref_n is the BRF of the P16×16 mode. For example, if ref_3 is the BRF of the P16×16 mode, the probability that one of ref_0, ref_1, ref_2, or ref_3 becomes the BRF in the P16×8 mode is 93.8%. As shown in Table 3, the hit rates of the proposed method are quite high.

It should be noted that the hit rates (in the P16×8, P8×16, and P8×8 modes) are 100% when $BRFI_{16}$ is 4, as the proposed method examines all five RFs in that case. It should also be noted that all five RFs are examined in the P16×16 mode because this mode serves as the basis for the BRF index correlation. The proposed RF selection method can be summarized as follows.

Step 1. Examine the P16×16 mode for all available RFs.

Step 2. Select the BRF index ($BRFI_{16}$) in the P16×16 mode.

Step 3. Examine each of the remaining modes (the P16×8, P8×16, and P8×8 modes) from ref_0 to ref_{BRFI16}.

Step 4. Choose the best mode (with the BRF) that has the minimum RD cost.

Clearly, the parameter $BRFI_{16}$ plays a very important role in the proposed method, influencing both the performance and computational complexity, since it determines the number of RFs to be examined in the P16×8, P8×16, and P8×8 modes. Table 4 shows the average $BRFI_{16}$ values for various test sequences. The overall average for $BRFI_{16}$ is only 0.329, which means that only 1.329 (=1+0.329) RFs, on average, have to be examined in the P16×8, P8×16, and P8×8 modes. This is a significant reduction, considering the five RFs that are used in the complete search method. It should also be noted that the proposed method checks only one RF in most macroblocks (in the P16×8, P8×16, and P8×8 modes) since $BRFI_{16}$ is 0 in most macroblocks, according to Table 1.

On the other hand, Table 5 shows the hit rate of the proposed method, in which the hit rate refers to the probability that the BRF of the proposed method is the same as the BRF of the complete search method, which examines all five RFs. Notably, the hit rate of the P16×16 mode is 100%, which is because the proposed method examines all five RFs in the P16×16 mode, just as in the complete search method. Also, the hit rates of the P16×8, P8×16, and P8×8 modes are quite high, which is partly because the BRF indices are highly correlated, as explained in Table 2, and partly because the BRF indices are usually very small, as explained in Table 1.

The last column in Table 5 shows the hit rate of the best mode, which represents the mode with the minimum RD cost among the P16×16, P16×8, P8×16, and P8×8 modes. For example, if the P16×8 mode is the best mode for a macroblock, only the hit rate results for the P16×8 mode are considered in the calculation of the hit rate of the best mode. Since the results of the best mode determine the final performance, the results in the last column are more important than the others. As shown in Table 5, the average hit rate of the best mode is 95.9%. Thus, it can be said that the proposed method performs quite well while significantly reducing the complexity of the complete search method.

Table 4. Average BRF index for P16×16 mode.

	Average $BRFI_{16}$		Average $BRFI_{16}$
Bus	0.443	Foreman	0.614
Coastguard	0.274	News	0.092
Football	0.402	Silent	0.150
Average of average $BRFI_{16}$: 0.329			

Table 5. Hit rate (%) of proposed method.

	P16×16	P16×8	P8×16	P8×8	Best mode
Bus	100.0	89.1	88.9	84.6	92.4
Coastguard	100.0	95.6	95.9	93.1	97.1
Football	100.0	90.5	91.0	88.4	94.3
Foreman	100.0	92.2	91.3	88.5	94.7
News	100.0	96.4	95.2	95.5	98.4
Silent	100.0	96.2	95.8	95.1	98.3
Average	100.0	93.3	93.0	90.9	95.9

III. Simulation Results and Conclusion

The proposed RF selection method is implemented with JM 16.0 reference software [10], and various test sequences are simulated using the simulation parameters shown in Table 6. The simulation results are shown in Table 7, which provides a comparison of the proposed method with the conventional methods. For this comparison, the simulation method by Bjøntegaard is used [11]. Table 7 shows only Bjøntegaard delta bitrate (BD-BR) values, because BD-PSNR values and BD-BR values essentially represent the same information [11].

As shown in Table 7, the proposed method reduces the encoding time by 62.8%, on average. As explained in the previous section, this complexity reduction comes from the fact that the proposed method examines a very small number of RFs in the P16×8, P8×16, and P8×8 modes. It should be mentioned that examining the P8×8 mode is far more time consuming than examining other modes since there are four submodes in the P8×8 mode. Thus, it is especially important to reduce the number of RFs to be examined in the P8×8 mode.

Although the method in [5] also reduces the encoding time significantly, the average BD-BR of the method in [5] is relatively large, as evident in Table 7. On the other hand, the average BD-BR of the proposed method is only 2.87%, which is substantially smaller than those of the conventional methods.

Table 6. Simulation parameters.

Parameter	Mode/Value
Coding structure	IPPP…
RD optimization mode	High complexity mode
Maximum number of RFs	5
Motion estimation	Full search
Search range	16
Entropy coding	CABAC
QP	18, 24, 30, 36

Table 7. Comparison of proposed method with other methods.

	[5]		[7]		[9]		Proposed method	
	BD-BR	Δtime	BD-BR	Δtime	BD-BR	Δtime	BD-BR	Δtime
Bus	4.01%	–51.7%	3.74%	–30.9%	6.84%	–47.8%	3.70%	–60.1%
Coastguard	2.91%	–68.2%	2.21%	–31.8%	4.85%	–54.9%	1.43%	–66.9%
Football	4.09%	–62.6%	1.73%	–39.8%	1.76%	–39.3%	2.17%	–62.9%
Foreman	10.89%	–63.1%	7.13%	–40.1%	19.12%	–50.2%	4.51%	–58.6%
News	4.59%	–72.0%	3.61%	–72.9%	2.57%	–63.7%	2.21%	–63.9%
Silent	7.92%	–71.9%	4.71%	–62.5%	2.86%	–59.3%	3.18%	–64.3%
Average	5.74%	–64.9%	3.86%	–46.3%	6.33%	–52.5%	2.87%	–62.8%

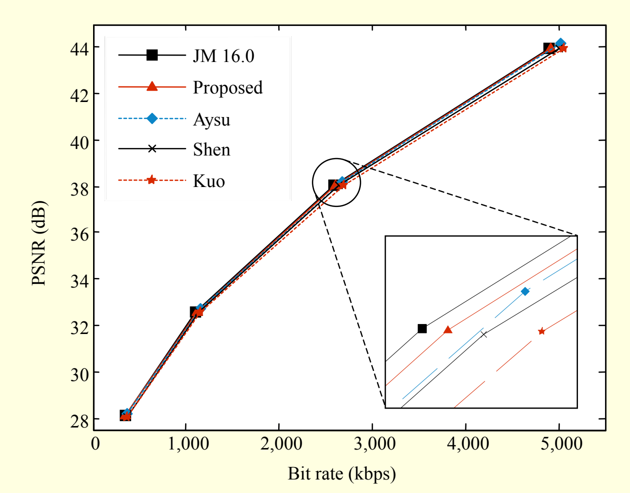

Fig. 1. RD curves for Coastguard video sequence.

As explained in section II, this is mainly because the hit rates of the proposed method are quite high owing to the efficient use of the correlation between the BRF indices of various inter modes. Figure 1 shows the RD curves of the RF selection methods reflected in Table 7 for the Coastguard video sequence. The quantization parameter (QP) values specified in Table 6 are used for the RD curves in Fig. 1. The proposed method shows better performance than the conventional methods throughout the whole QP range. Finally, it should be noted that the maximum number of RFs (in Table 6) can be easily changed in the proposed method, which is another merit of the proposed method.

This letter proposed an RF selection method, which efficiently exploits the correlation between the BRF indices of various inter modes. Although the main idea is very simple, the proposed method effectively reduces the number of RFs at the expense of a very small miss rate. As a result, the proposed method not only reduces the encoding time considerably but also performs well regardless of the QP value.

References

[1] Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, "Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H.264/ISO/IEC 14496-10 AVC)," *JVT-G050*, Mar. 2003.

[2] Y. Choi et al., "Weighted Constrained One-Bit Transform Method for Low-Complexity Block Motion Estimation," *ETRI J.*, vol. 34, no. 5, Oct. 2012, pp. 795-798.

[3] Y. Su and M.T. Sun, "Fast Multiple Reference Frame Motion Estimation for H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 3, Mar. 2006, pp. 447-452.

[4] Z. Wang et al., "An Efficient Algorithm for Motion Estimation with Multiple Reference Frames in H.264/AVC," *IEEE Int. Conf. Image Graph.*, Aug. 2007, pp. 259-262.

[5] L. Shen et al., "An adaptive and Fast Multiframe Selection Algorithm for H.264 Video Coding," *IEEE Signal Process. Lett.*, vol. 14, no. 11, Nov. 2007, pp. 836-839.

[6] P. Wu and C.B. Xiao, "An Adaptive Fast Multiple Reference Frames Selection Algorithm for H.264/AVC," *IEEE Int. Conf. Acoustics, Speech, Signal Process.*, Apr. 2008, pp. 1017-1020.

[7] A. Aysu, G. Sayilar, and I. Hamzaoglu, "A Low Energy Adaptive Hardware for H.264 Multiple Reference Frame Motion Estimation," *IEEE Trans. Consum. Electron.*, vol. 57, no. 3, Aug. 2011, pp. 1377-1383.

[8] Y.W. Huang et al., "Analysis and Complexity Reduction of Multiple Reference Frames Motion Estimation in H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 4, Apr. 2006, pp. 507-522.

[9] T.Y. Kuo and H.J. Lu, "Efficient Reference Frame Selector for H.264," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 3, Mar. 2008, pp. 400-405.

[10] Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, H.264/AVC reference software, version JM 16.0.

[11] G. Bjøntegaard, "Calculation of Average PSNR Differences between RD-Curves," ITU-T 13th VCEG Meeting, Doc. VCEG-M33, Apr. 2001.

http://dx.doi.org/10.4218/etrij.14.0213.0394