

FAST SPATIAL AND TEMPORAL CORRELATION-BASED REFERENCE PICTURE SELECTION

¹YONGTAE KIM, ²JAE-GON KIM, and ³HAECHUL CHOI

^{1,3} Hanbat National University, Department of Multimedia Engineering

² Korea Aerospace University, Department of Computer Science

E-mail: ¹ceo@toprnd.com, ²jgkim@kau.ac.kr, and ³choihc@hanbat.ac.kr

ABSTRACT

Multiple reference pictures and variable prediction block sizes in motion estimation/compensation (ME/MC) adopted in video coding standards such as H.264/AVC and H.265/HEVC achieve high coding efficiency, but these tools require heavy encoding complexity. This paper introduces a reference picture selection method based on the spatial correlation between neighboring coded blocks and the temporal correlation between the reference pictures of a reference picture list. This method can reduce the number of reference picture to be searched in ME process. This reduction provides competitive performance with reduced computational complexity. Experimental results show that the proposed method reduces encoding run-time by 47%, with a negligible degradation of coding efficiency.

Keywords: *H.264/AVC, H.265/HEVC, Multiple Reference Pictures, Motion Estimation*

1. INTRODUCTION

A large quantity of video material is already being distributed digitally over broadcast channels and wire/wireless networks, and in packaged media. More and more of this material will be distributed with increased resolution and quality [1]. Two important requirements of video encoders in the areas of mobile communication, high-definition broadcasting, and digital cinema are low computational complexity and high coding efficiency.

H.264/Advanced Video Coding (AVC) [2]-[4] has adopted includes variable prediction block sizes, quarter-pixel accurate motion vectors (MVs), and multiple reference pictures for the motion estimation/compensation (ME/MC) process. H.265/High Efficiency Video Coding (HEVC) [5]-[8] was developed by the ITU-T Video Coding Experts Group (VCEG) and ISO/IEC Moving Picture Experts Group (MPEG) standardization organizations, working together in a partnership known as the Joint Collaborative Team on Video Coding (JCT-VC). HEVC, the most recent international video coding standard, enables significantly improved coding efficiency compared to H.264/AVC in the range of 50% bit rate reduction for equal perceptual video quality [9]. However, HEVC encoders are also expected to have higher

complexity [10]. Therefore, fast encoding algorithms have been broadly researched to decrease the encoding complexity of HEVC for real-time video encoding systems or power-constrained mobile devices.

The scheme of multiple reference pictures that has been adopted in H.264/AVC and H.265/HEVC can provide more accurate prediction with multiple decoded pictures, which contributes significantly to high coding efficiency. This scheme is more efficient when predicting uncovered background or regions with changes in illumination. However, this scheme requires the selection of an optimal reference picture among multiple decoded pictures, which significantly increases the computational burden in proportion to the number of the searched reference pictures

To solve this complexity problem, various reference picture selection methods have been studied since H.264/AVC standard has been released. Li et al. [11] used the MV linearity of previously coded pictures to select an appropriate reference picture. In [11], MVs for reference pictures with a reference picture index of 0 through 1 are obtained first. When the difference between the two MVs is smaller than a certain threshold, and the ME cost of the reference picture index 0 is smaller than that of the reference picture index 1, the remaining reference pictures having an index number larger

than 1 are skipped for the ME process. H.264/AVC and H.265/HEVC encoders generally allow four reference pictures, of which two that are temporally far from a current picture could be adaptively excluded in the ME process. Su and Sun [12] introduced a fast multiple reference frame motion estimation algorithm using a weighted average for motion compensation to eliminate unnecessary motion estimations. In studies by Wang et al. [13] and H.-J. Li et al. [14], the ME process was performed only for the reference pictures used by neighboring coded blocks. Kuo and Lu [15] selected suitable reference pictures according to the initial motion estimation result for an 8×8 size block, and then only the selected pictures were searched further with variable block sizes. H. Lee et al. [16] used the optimal reference frame information in both the 16×16 inter-prediction mode and the adjacent blocks to decrease any unnecessary searching process in the inter-prediction modes except for the 16×16 inter-prediction mode. M. Xiao and Y. Cheng [17] proposed a fast multi-reference frame selection algorithm based on the correlation between the multi-reference frame selection and the size of the motion vector and the monotonicity of the different reference frame's rate-distortion. C. Jeong and M. Hong [18] estimated the motion estimation of small block size from the motion estimation of the larger block size.

Most of the conventional methods have generally focused on either a spatial correlation or a temporal correlation, but not both correlations together. For example, the studies in [11], [17], and [18] focused on a temporal correlation, but the studies in [13] and [14] focused on a spatial correlation. In this paper, spatial correlation refers to the correlation between the current block and its neighboring blocks, whereas temporal correlation refers to the correlation between the current picture and reference picture candidates from a reference-picture list.

To reduce the computational complexity while keeping the coding efficiency competitive, we propose a fast reference picture selection scheme using both the spatial and temporal correlations. Neighboring blocks are likely to be highly correlated in terms of texture, illumination, motion information, etc. Thus, in order to take into account the spatial correlation, the proposed scheme utilizes the reference pictures used in neighboring coded blocks, which reduces the number of reference pictures to be searched in the ME process. In addition, for large blocks such as blocks larger than 16×16 , only the temporal correlation is used without considering the

spatial correlation. In particular, the reference picture search is terminated early when the ME cost increases dramatically over reference pictures in a reference picture list. Our approach avoids the excessively tight limitation on the number of reference pictures that may occur when using only the spatial correlation. For example, if a coding block should select only the reference picture of the previous block on the left, all subsequent coding blocks should also use the same reference picture because the selection is continuously propagated.

2. FAST REFERENCE PICTURE SELECTION METHOD

In inter prediction modes, the predicted motion vector (PMV) is obtained by using the three available MVs of neighboring blocks since the current block is likely to have a high correlation with the neighboring blocks. This spatial correlation is also utilized in the proposed method. The number of reference pictures to be searched in the ME process is reduced by using the motion information of neighboring blocks, which can reduce the encoding computational complexity.

To exploit the spatial correlation between blocks, the proposed method considers the reference pictures of neighboring coded blocks as reference pictures. This method may be efficient because there is a high correlation between the reference pictures as described above. However, such a strong restriction on reference pictures may lead to unexpected quality degradation. If the reference pictures of the neighboring blocks are identical, then the number of reference pictures decreases to 1; that is, only one reference picture is available for the current block. Furthermore, for all subsequent blocks in the current slice, only one reference picture is available, regardless of the number of possible reference pictures set by the encoding configuration. Thus, this strong restriction can cause unexpected considerable coding loss. To prevent the propagation of this strong restriction, the proposed method selects reference pictures in different ways, depending on whether the inter-prediction block size is larger than $N \times N$ or not. For the small inter-prediction block sizes, reference pictures used by neighboring blocks are searched. In contrast, for the large block sizes, only the temporal correlation between reference pictures of the reference picture list is considered. In the following subsections, we will describe the proposed methods for the large block sizes and for the small block sizes.

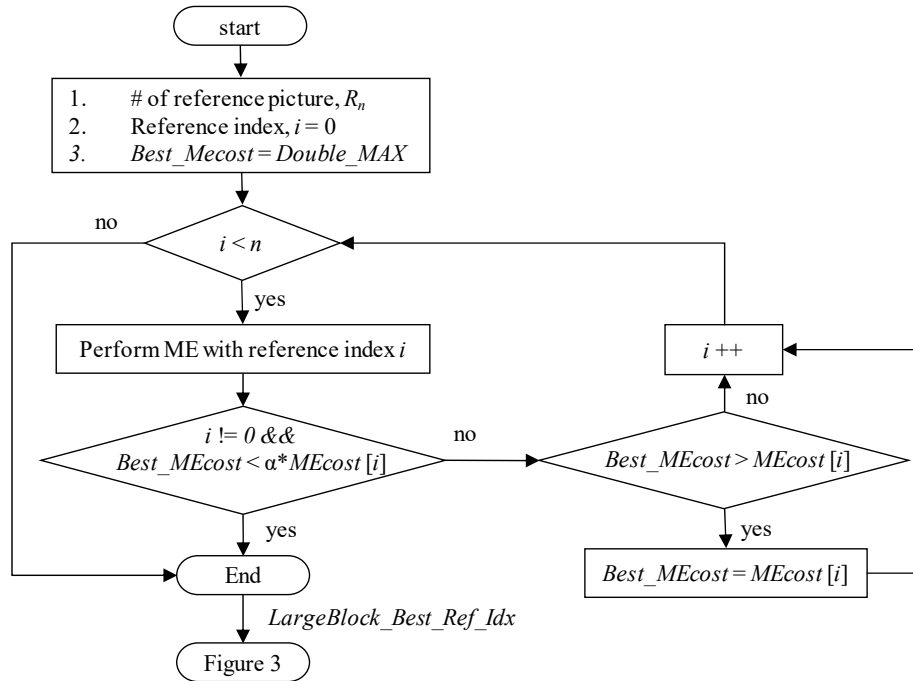


Figure 1: Flow Chart of The Proposed Motion Estimation Process For The Blocks Larger Than $N \times N$.

2.1 Inter Prediction Process for Large Blocks

First, for the blocks larger than $N \times N$, the proposed reference picture selection method is illustrated in Figure 1. In this figure, $MEcost[i]$ is the ME cost for the i -th reference picture in a reference picture list, $Best_MEcost$ is the minimum ME cost among the ME costs calculated so far, and n denotes the total number of reference pictures in the reference picture list. For the reference picture index of 0, the value of $MEcost[0]$ is obtained, and then $MEcost[0]$ is assigned to $Best_MEcost$. Then, $MEcost[i]$ is calculated for the reference picture index i , ($0 < i < R_n$, the number of reference pictures).

Whenever calculating $MEcost[i]$, it is compared with $Best_MEcost$, as shown in the following equation:

$$Best_MEcost < \alpha \times MEcost [i], \quad (1)$$

where α is a weighting factor to account for the temporal correlation between reference pictures. The weighting factor is generally defined as being in the 0 to 1 range. Satisfying (1) means that the temporal correlation is relatively low. Thus, in the case where the ME cost for the i -th reference picture is significantly larger than $Best_MEcost$, the remaining reference pictures, for which the index is larger than i , are not searched in the ME process. In contrast, if

(1) is not satisfied, $Best_MEcost$ is replaced with $MEcost[i]$, and then the $(i+1)$ -th reference picture is continuously searched. This reference picture selection method for the large blocks can reduce computational complexity, especially when the temporal correlation between reference pictures decreases dramatically as the temporal distance increases.

Finally, the reference picture index associated with $Best_MEcost$ is noted as $LargeBlock_Best_Ref_Idx$. It indicates the best reference picture for the blocks larger than $N \times N$. In particular, it will be also taken into account to select reference pictures for the small blocks equal to or smaller than $N \times N$. Note that the spatial correlation between neighboring blocks is not considered for the large blocks. Moreover, all R_n reference pictures can be searched in the case where (1) is not satisfied for all values of i . Thus, the propagation of the strong restriction on the number of reference pictures can be avoided by the reference picture selection of the blocks. The amount of complexity reduction depends on the value of α . In Section 3, the performance of the proposed method will be evaluated for various values of α .

2.2 Inter Prediction Process for Small Blocks

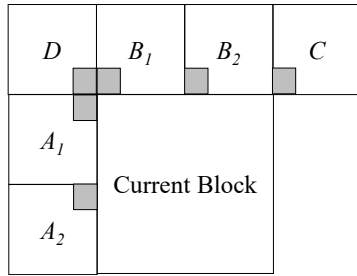


Figure 2: Six Neighboring Blocks To Be Considered in The Reference Picture Selection for The Small Blocks Equal To Or Smaller Than $N \times N$.

The following discussion describes the proposed algorithm to select reference pictures for the blocks equal to or smaller than $N \times N$. Unlike the PMV derivation process, which depends on three neighboring blocks in H.264/AVC, the proposed method uses six neighboring blocks, as illustrated in Figure 2. Considering a larger number of neighboring blocks is helpful to more properly select reference pictures.

Neighboring blocks can be coded as various prediction modes and prediction block sizes. Reference picture index of the j -th neighboring blocks, $spatial_refidx_j$ ($0 \leq j \leq 5$), is derived as follows: When $spatial_refidx_j$ is not available

because the neighboring block is coded as intra-prediction mode or does not exist, $spatial_refidx_j$ is set to 0. When the neighboring block A_1 block has more than one coded blocks, the reference picture index of the top right block within the block A_1 represents that of the block A_1 . In the same way, the bottom right block and bottom left block represent the neighboring block D and the neighboring blocks B_1 , B_2 , and C , respectively.

Figure 3 shows the reference picture selection method for the small blocks equal to or smaller than $N \times N$. Prior to the ME process, the reference picture indices of the six neighboring blocks are examined to find the largest index value, i_{max} . Then, i_{max} is compared with the reference picture index obtained by the reference picture selection method of the large block as illustrated in Figure 1, $LargeBlock_Best_Ref_Idx$. The final maximum reference picture index, i_{final} , to be searched in the ME process is determined as the larger of i_{max} and $LargeBlock_Best_Ref_Idx$. After obtaining i_{final} , the reference pictures having an index value smaller than or equal to i_{final} are searched in the ME process. In other words, reference pictures with an index value larger than i_{final} are not searched, which can reduce the computational complexity in motion estimation process.

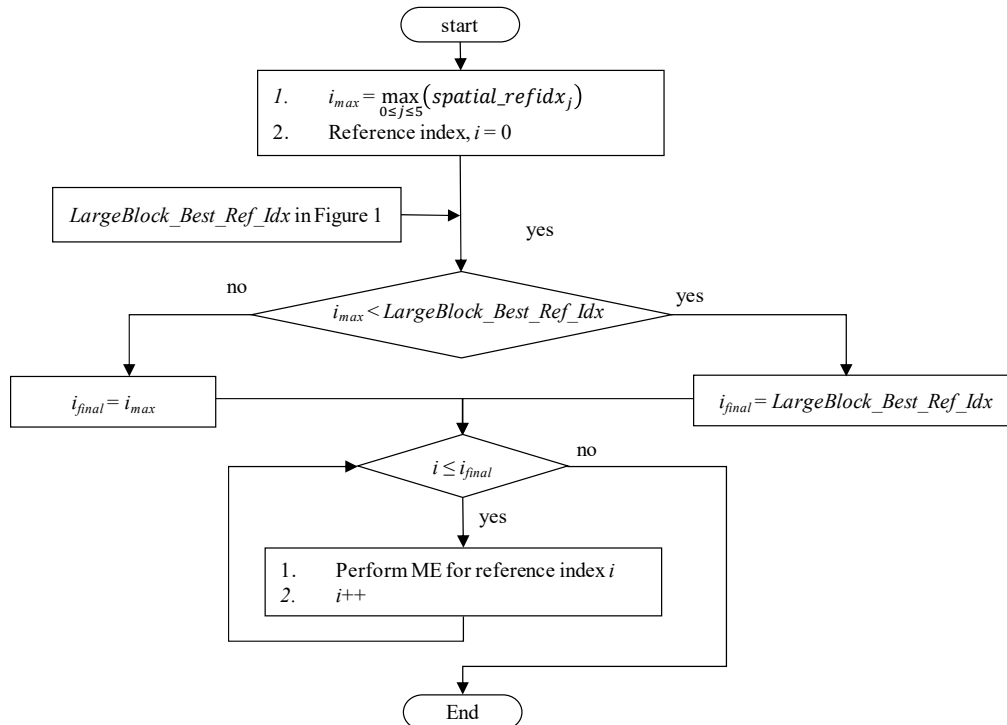


Figure 3: Flow Chart of The Proposed Motion Estimation Process for The Blocks Equal To or Smaller Than $N \times N$.

As described above, the reference picture selection scheme of the large blocks does not consider neighboring blocks. Moreover, the reference picture selection scheme for the small blocks somewhat reduces the dependency on the neighboring coded blocks, because the final maximum reference picture index is determined by using not only reference picture indices of neighboring blocks but also the *Large_Block_Best_Ref_Idx*. This scheme prevents the excessively tight limitation on the number of reference pictures.

In terms of coding structure, most of the conventional reference picture selection methods including the methods described in Section I, have been proposed for the IPPP prediction structure. On the other hand, our proposed method can be simply extended to B (bi-predictive) slices. The proposed reference picture selection scheme can be applied independently to each reference picture list. For instance, when selecting a proper reference picture for the reference picture list 0, only reference pictures of the list 0 used by neighboring blocks are considered. In the same way, the proposed method can be performed in selecting a reference picture of the list 1. The bi-predictive ME cost is obtained by the sum of the uni-predictive ME costs calculated by applying the proposed scheme to each reference picture list. If the *j*-th neighboring block is coded as the uni-prediction mode, the reference picture index of the other list is not available. In this case, *spatial_refidx_j* of the other list is set to 0 and then the proposed ME process for the small blocks is performed.

Table 1: Test Sequences and Encoding Configuration

Test sequences (name, spatial resolution, total encoded frames)	Container, QCIF, 100
	Foreman, QCIF, 100
	News, QCIF, 100
	Silent, QCIF, 150
	Paris, CIF, 150
	Mobile, CIF, 300
	Tempete, CIF, 260
Prediction structure	IPPP
Quantization parameters	32, 36, 40, 44 for I-slices
Search mode	Fast full search
Search range	±16
Number of reference pictures	5

3. EXPERIMENTAL RESULTS

The proposed method is implemented on the top of Joint Model (JM) 18.4 [19], the reference software of the H.264/AVC. As described in Section I, the H.265/HEVC also adopts the multiple reference picture scheme. Therefore, the implementation of the proposed method on JM is easily applied to the HEVC Test Model (HM) [20]. Since H.264/AVC is still being more popular than H.265/HEVC, the experiments are conducted on JM.

In all the following experiments, the threshold, *N*, to categorize blocks into the large or the small is set to 16. The test conditions are summarized in Table 1. Four quantization parameters of 32, 36, 40, and 44 are adopted to cover the testing of various bit rates. Test sequences are a set of public sequences

Table 2: Coding Efficiency of the Proposed Method at IPPP Structure. The ‘spatial correlation’ is experimental result when only motion information of neighboring blocks is used. The ‘spatial and temporal correlations’ is experimental result when the proposed methods for the large blocks and for the small blocks are integrated where the threshold, *N*, to categorize into the large blocks or the small blocks is set to 16. The α is the weighting factor to account for the temporal correlation between reference pictures from (1)

Sequences	Spatial correlation		Spatial and temporal correlations ($\alpha=0.7$)		Spatial and temporal correlations ($\alpha=0.9$)	
	BD-PSNR (dB)	BD-Rate (%)	BD-PSNR (dB)	BD-Rate (%)	BD-PSNR (dB)	BD-Rate (%)
Container (QCIF)	-0.18	2.6	-0.04	0.4	-0.04	0.5
Foreman (QCIF)	-0.05	0.9	-0.01	0.1	0.01	-0.1
News (QCIF)	-0.03	0.6	-0.02	0.3	-0.02	0.4
Silent (QCIF)	-0.07	1.4	-0.01	0.1	0.00	0.1
Paris (CIF)	-0.09	2.0	-0.04	0.8	-0.04	0.8
Mobile (CIF)	-0.38	9.4	-0.08	1.8	-0.10	2.4
Tempete (CIF)	-0.14	2.8	-0.02	0.5	-0.02	0.5
Average	-0.13	2.83	-0.03	0.58	-0.03	0.67

Table 3: Complexity Reduction of the Proposed Method at IPPP Structure. The 'spatial correlation' is experimental result when only motion information of neighboring blocks is used. The 'spatial and temporal correlations' is experimental result when the proposed methods for the large blocks and the small blocks are integrated where the threshold, N , to categorize into the large blocks or the small blocks is set to 16. The α is the weighting factor to account for the temporal correlation between reference pictures from (1)

Sequences	Spatial correlation		Spatial and temporal correlations ($\alpha=0.7$)		Spatial and temporal correlations ($\alpha=0.9$)	
	TET (%)	MET (%)	TET (%)	MET (%)	TET (%)	MET (%)
Container (QCIF)	70	80	41	45	47	52
Foreman (QCIF)	67	77	40	44	48	53
News (QCIF)	68	79	40	44	46	51
Silent (QCIF)	68	79	42	46	46	52
Paris (CIF)	69	80	37	40	47	52
Mobile (CIF)	65	77	40	44	45	51
Tempete (CIF)	73	82	44	48	49	54
Average	70	80	41	45	47	52

that have a variety of characteristics such as motion activities, spatial resolutions, and frame rates. Coding efficiency is measured by the Bjöntegeard delta bit rate (BD-Rate) and Bjöntegeard delta PSNR (BD-PSNR) [21], which provide the relative gain by calculating the average difference between two rate-distortion curves. To compare the encoding computational complexity, the calculation in (2) that represents the encoding run-time saving is also measured:

$$\Delta T(\%) = \frac{T_{JM} - T_{Prop}}{T_{JM}} \times 100 \quad (2)$$

where T_{JM} and T_{Prop} represent the encoding run-time of JM 18.4 and the proposed method, respectively. The ΔT is calculated for the total encoding run-time (TET) and motion estimation encoding run-time (MET), respectively. In all experiments, the anchor is the bit-stream generated using JM 18.4 under the conditions given in Table 1.

The proposed method selects the reference picture differently according to the prediction block sizes, 16×16 and smaller than 16×16 . First, the individual performance using only the spatial correlation is evaluated. In this experiment, only motion information of neighboring blocs is used and the number of reference pictures is restricted by the largest reference picture index of the six neighboring blocks illustrated in Figure 2. This restriction is applied to all inter-prediction processes for the large blocks and the small blocks. The first column of Tables 2 and 3 shows the coding efficiency and computational complexity, respectively, for this experiment. As shown in Tables 2 and 3, the total

encoding time saving is an average of 70% and the BD-Rate increase is 2.83% relative to the anchor. This simple scheme using only the spatial correlation provides a significant reduction in computational complexity, whereas the scheme has a relatively high coding loss.

Next, the proposed methods for the 16×16 inter-prediction mode and for the smaller-size inter-prediction modes are integrated and the integrated method is tested. Experimental results are shown in the second and third columns of Tables 2 and 3. These results are obtained in the cases where α value from (1) is set to 0.7 and 0.9, respectively. When α is set to 0.7, the total encoding time savings is an average of 41% with a negligible coding loss of 0.58% BD-Rate. In addition, when α is set to 0.9, the encoding time saving is an average of 47% with a BD-Rate increase of 0.67%. The experimental results show that computational complexity is further reduced as the value of α is increased. The reason for the higher complexity reduction is that higher value of α increases the threshold from (1) to early terminate the search of the other reference pictures. Even though α is close to 1, the coding loss is still negligible. Therefore, it can be addressed that α is not sensitive to coding loss in the interval where α is between 0.7 and 0.9. Using a much higher or lower value of α , the tradeoff between computational complexity and coding efficiency can be controlled.

In Tables 2 and 3, compared with the experimental results of the first column, the computational complexity reduction of the integrated method is slightly decreased by 0.1% BD-

Table 4: Coding Efficiency and Complexity Reduction of Kuo and Lu's method [15] at IPPP Structure.

Sequences	BD-PSNR (dB)	BD-Rate (%)	TET (%)	MET (%)
Container (QCIF)	-0.01	0.4	35	47
Foreman (QCIF)	-0.04	0.9	21	28
News (QCIF)	-0.03	0.5	35	47
Silent (QCIF)	-0.14	1.9	34	46
Paris (CIF)	-0.04	0.9	33	46
Mobile (CIF)	-0.04	2.2	17	24
Tempete (CIF)	-0.04	1.3	21	28
Average	-0.05	1.14	35	47

Table 5: Coding efficiency and Complexity Reduction for Hierarchical B Picture Structure when α in (1) is equal to 0.7. The 'spatial and temporal correlations' is experimental result when the proposed methods for the large blocks and the small blocks are integrated where the threshold, N , to categorize into the large blocks or the small blocks is set to 16. The α is the weighting factor to account for the temporal correlation between reference pictures from (1)

Method	Spatial and temporal correlations ($\alpha=0.7$)			
Sequences	BD-PSNR (dB)	BD-Rate (%)	TET (%)	MET (%)
Container (QCIF)	-0.023	0.3	42	43
Foreman (QCIF)	-0.019	0.3	42	43
News (QCIF)	0.008	-0.1	40	42
Silent (QCIF)	-0.004	0.1	39	40
Paris (CIF)	-0.018	0.4	41	42
Mobile (CIF)	-0.017	0.4	43	44
Tempete (CIF)	-0.019	0.4	44	46
Average	-0.013	0.26	42	43

Table 6: Coding efficiency and Complexity Reduction for Hierarchical B Picture Structure when α in (1) is equal to 0.9. The 'spatial and temporal correlations' is experimental result when the proposed methods for the large blocks and the small blocks are integrated where the threshold, N , to categorize into the large blocks or the small blocks is set to 16. The α is the weighting factor to account for the temporal correlation between reference pictures from (1)

Method	Spatial and temporal correlations ($\alpha=0.9$)			
Sequences	BD-PSNR (dB)	BD-Rate (%)	TET (%)	MET (%)
Container (QCIF)	0.002	0.1	43	44
Foreman (QCIF)	-0.051	1.0	46	47
News (QCIF)	0.011	-0.2	42	43
Silent (QCIF)	0.002	0.0	41	42
Paris (CIF)	-0.028	0.6	42	44
Mobile (CIF)	-0.016	0.4	46	48
Tempete (CIF)	-0.041	0.8	45	46
Average	-0.017	0.39	44	45

Rate, but the coding loss is significantly reduced by more than 23% in TET. When considering coding efficiency and computational complexity together, it can be suggested that the proposed ME processes for the large and the small blocks perform better than the

ME process that is based on only the spatial correlation.

For the comparison with the proposed method, the coding efficiency and computational complexity of Kuo and Lu's method [15] that was

described in Section 1 are listed in Table 4. Their method reduces the total encoding run-time by 35% with a BD-Rate increase of 1.14% relative to JM 18.4. Comparing to Kuo and Lu's method, our proposed method can provide a greater reduction in computational complexity, as well as a smaller coding loss, in both the two values of α . In particular, when α is equal to 0.9, the proposed method has coding efficiency by -0.47% BD-rate on average and computational reduction by 12% of TET on average relative to Kuo and Lu's method. The reason for the outperformance is that Kuo and Lu's method considers only the initial search result of the 8×8 inter-prediction mode, whereas the proposed method utilizes the spatial correlation as well as the search result of the large blocks such as 16×16 inter-prediction mode.

As described in subsection 2.2, in terms of coding structure, most of the conventional reference picture selection methods have been proposed for the IPPP prediction structure, and have shown experimental results only for that structure. On the other hand, our proposed method can be simply extended to B (bi-predictive) slices. To evaluate performance for the hierarchical B picture structure [22], an additional evaluation is conducted, in which three reference pictures were allowed for each reference picture list. As shown in Tables 5 and 6, the proposed method reduces computational complexity by 42% of TET while keeping the coding efficiency competitive when the weighting factor α is equal to 0.7. When the weighing factor is equal to 0.9, additional computational complexity of 2% of TET is reduced with negligible coding loss. When compared to results for the IPPP prediction structure, those for the hierarchical B picture structure have slightly more coding loss by about 0.1% BD-Rate on average and similar computational complexity reduction. The hierarchical B picture structure has usually better coding efficiency than the IPPP prediction structure because the hierarchical B picture structure utilizes two reference pictures. Therefore, the restriction imposed on the reference picture selection in the proposed method may cause the more coding loss. Nevertheless, the experimental results, the coding loss of 0.13% BD-Rate and the computational complexity reduction of 42% in TET, can reveal that the proposed method also works well for the bi-predictive slices.

4. CONCLUSION

To reduce the encoding complexity required for motion estimation with multiple

reference pictures in H.264/AVC and H.265/HEVC, a reference picture selection method is proposed. Based on the temporal correlation between the reference pictures in a reference picture list, the proposed method terminates the ME process early for the large blocks. Furthermore, by taking into account the reference pictures used in neighboring blocks and the reference selection result of the large blocks together, the proposed method constrains the number of reference pictures for the small blocks. The experimental results demonstrate that the proposed method provides a significant reduction in computational complexity while maintaining a competitive coding efficiency. The proposed method works with any existing motion search algorithm. Thus, if this method is coupled with such an algorithm that is fast and well-developed, the reduction in computational complexity would be enhanced.

ACKNOWLEDGEMENT:

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. NRF-2016R1A2B1012652)

REFERENCES:

- [1] S. Jeong, S.-C Lim, H. Lee, J. Kim, J. S. Choi, and H. Choi, "Highly Efficient Video Codec for Entertainment-Quality," *ETRI Journal*, Vol. 33, No. 2, April 2011, pp. 145-154.
- [2] ITU-T and ISO/IEC JTC 1, Advanced video coding for generic audiovisual services, *ITU-T Recommendation H.264 and ISO/IEC 14496-10 (MPEG4-AVC)*, 4th ed., Sept. 2008.
- [3] T. Wiegand, G. J. Sullivan, G. Bjntegaard, and A. Luthra, "Overview of the H.264/AVC Video Coding Standard," *IEEE Trans. Circuits Syst. Video Technol.*, Vol. 13, No. 7, July 2003, pp. 560-576.
- [4] G. J. Sullivan and T. Wiegand, "Video compression - from concepts to the H.264/AVC standard," *Proc. of the IEEE*, Vol. 93, No. 1, Jan. 2005, pp. 18-31.
- [5] Joint Collaborative Team on Video Coding (JCT-VC), "High Efficiency Video Coding (HEVC) text specification draft 10 (for FDIS & Consent)," *JCTVC-L1003*, Geneva, Jan. 2013.
- [6] G. J. Sullivan and J.-R. Ohm, "Recent developments in standardization of High Efficiency Video Coding (HEVC)," *Proc.*

- 33rd SPIE Appl. Dig. Image Process., Vol. 7798, Aug. 2010, pp. 7798–7830.
- [7] T. Wiegand et al., “Special section on the joint call for proposals on High Efficiency Video Coding (HEVC) standardization,” *IEEE Trans. Circuits Syst. Video Technol.*, Vol. 20, no. 12, Dec. 2010, pp. 1661–1666.
- [8] Gary J. Sullivan et al., “Overview of the High Efficiency Video Coding (HEVC) Standard,” *IEEE Trans. Circuits Syst. Video Technol.*, Vol. 13, No. 12, Sept. 2012, pp. 1649–1668.
- [9] J.R. Ohm, G. J. Sullivan, and H. Schwarz, “Comparison of the Coding Efficiency of Video Coding Standards—Including High Efficiency Video Coding (HEVC),” *IEEE Trans. Circuits Syst. Video Technol.*, Vol. 22, No. 12, Oct. 2012, pp. 1669–1684.
- [10] F. Bossen et al., “HEVC complexity and implementation analysis,” *IEEE Trans. Circuits Syst. Video Technol.*, Vol. 22, No. 12, Dec. 2012, pp. 1685–1696.
- [11] X. Li, E. Q. Li, and Y.-K. Chen, “Fast Multi-Frame Motion Estimation Algorithm with Adaptive Search Strategies in H.264,” *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Vol. 3, May 2004, pp. 369–372.
- [12] Y. Su and M.-T. Sun, “Fast Multiple Reference Frame Motion Estimation for H.264/AVC,” *IEEE Tran. Circuits Syst. Video Technol.*, Vol. 16, No. 3, Mar. 2006, pp. 447–452.
- [13] H.-J. Wang, L. L. Wang, and H. Li, “A Fast Multiple Reference Frame Selection Algorithm Based on H.264/AVC,” *Proc. IEEE Int. Conf. intelligent information hiding and multimedia signal processing*, Vol. 1, Nov. 2007, pp. 525–528.
- [14] H.-J. Li, C.-T. Hsu, and M.-J. Chen, “Fast Multiple Reference Frame Selection Method for Motion Estimation in AVC/H.264,” *Proc. IEEE Asia Pac. Conf. Circuits Syst.*, Dec. 2004, pp. 605–608.
- [15] 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.
- [16] H. Lee, J. Ryu, S. Lee, and J. Jeong, “Effective multi-frame motion estimation method for H.264,” *Proc. Picture Coding Symposium*, May 2009, pp. 321–324.
- [17] M. Xiao and Y. Cheng, “A fast multi-reference frame selection algorithm for H.264/AVC,” *Proc. IEEE International Conference on Computer Science and Automation Engineering*, Vol. 4, Jun. 2011, pp. 615–619.
- [18] C. Jeong and M. Hong, “Fast multiple reference frame selection method using inter-mode correlation,” *Proc. Asia-Pacific Conference on Communications*, Oct. 2008, pp. 1–4.
- [19] H.264/AVC Reference Software Joint Model (JM) version 18.4.
<http://iphome.hhi.de/suehring/tml/>
- [20] High Efficiency Video Coding Test Model Software 16.6,
<https://hevc.hhi.fraunhofer.de/svn/svnHEVCSoftware>.
- [21] G. Bjontegaard, “Improvements of the BD-PSNR model,” *ITU-T SG16/Q6, 35th VCEG Meeting*, Berlin, Germany, Doc. VCEG-AI11, July 2008.
- [22] H. Schwarz, D. Marpe, and T. Wiegand, “Hierarchical B pictures,” *Joint Video Team (JVT) of ISO-IEC MPEG & ITU-T VCEG*, JVT-P014, July 2005.