Motion vector coding with selection of an optimal predictive motion vector

Jungyoup Yang, Kwanghyun Won, and Byeungwoo Jeon, MEMBER SPIE
Sungkyunkwan University, Department of Electronic and Electrical Engineering, 300 Chunchun-dong, Jangan-gu, Suwon, Korea

Abstract. A new motion vector coding method with optimal predictive motion vector selection is proposed. To improve compression performance, the proposed encoder selects an optimal predictive motion vector that produces minimum bits for motion vector coding. The proposed decoder estimates the optimal predictive motion vector without additional information for indicating which predictor is to be used at the encoder side. Experimental results show that compared to the H.264/AVC standard, the proposed scheme improves coding efficiency for various video sequences. © *2009 Society of Photo-Optical Instrumentation Engineers.*
[DOI: 10.1117/1.3070632]

1 Introduction

In general inter-predicted video coding, the motion vector (MV) provides a spatial offset of a block in the current picture to a block in the reference picture. Therefore, more MV information is needed to improve the accuracy of the inter prediction. To minimize the number of bits required to represent MV information, the H.264/AVC standard applies a predictive coding method by using predictive motion vectors (PMVs), which are calculated as the median of three spatially neighboring MVs.[1] The median PMV is effective at reducing the required number of compressed MV bits, since it is very similar to the MV in most cases. However, the median PMV is not always optimal for minimizing the number of MV bits. If a more precise PMV exists than the median PMV, there is a chance that even more bits can be saved.

To overcome this problem, several approaches have been taken.[2–4] In Chen and Willson's work,[2] MVs that are located in spatially and temporally neighboring blocks are additionally considered to select a more precise PMV. A spatial or temporal PMV is selected according to the value of the predictors. However, this method cannot ensure selection of the optimal PMV. In Kim and Ra's work,[3] an optimal PMV is selected by using a distance measure function between the MV and PMV. In this method, however, additional information is required to decide which candidate PMV is an optimal one. Finally, in Laroche et al.'s work,[4] several candidate PMVs are generated by a combination of spatial and temporal neighboring MVs. A more precise PMV is then selected by a rate distortion (RD) competing scheme. As in Kim and Ra's work,[3] in some cases this method requires additional information to determine which predictor is to be used. This means that the benefit of using the optimal PMV may not be fully realized, because the choice of the optimal PMV by the encoder must be signaled to the decoder.

Therefore, in this paper, we propose a new motion vector coding method with optimal PMV selection (MVOP) to use the optimal PMV without the need for additional signaling information. First, the encoder defines the set of possible candidate PMVs by using neighboring MVs. To minimize bits of the MV information, an optimal PMV is selected among the candidate PMV set. If the decoder can estimate the optimal PMV by using decoder-side estimation, the encoder selects it as an optimal PMV. Otherwise, the encoder selects the median PMV in the same manner as the H.264/AVC standard. In the worst case of the proposed method, only 1 bit of additional information is required to signal whether the decoder can estimate the optimal PMV or not. Simulation results show that the proposed method reduces the average Bjontegaard delta bit rate (BDBR) by about 2.97% and increases the average Bjontegaard delta peak signal-to-noise ratio (BDPSNR) by about 0.14 dB compared with the H.264/AVC standard.

2 Proposed Method

2.1 PMV Candidate Set

As shown in Fig. 1, the candidate set (CS) is defined to select a more precise PMV than the median PMV. The CS, which is a group of possible and distinct candidate PMVs for the current block, is composed of a combination of horizontal and vertical components of spatial neighboring MVs. The CS is defined by

$$CS = \text{combination of}\{\mathbf{mv}^L, \mathbf{mv}^U, \mathbf{mv}^R\}$$
$$= \{(\text{mv}_x^L, \text{mv}_y^L), (\text{mv}_x^L, \text{mv}_y^U), \ldots, (\text{mv}_x^R, \text{mv}_y^U), (\text{mv}_x^R, \text{mv}_y^R)\}, \quad (1)$$

where mv_x^L and mv_y^L are the horizontal and vertical components of $\mathbf{mv}^L$, respectively. In this paper, spatially neighboring MVs are considered in the CS in the same manner as in the H.264/AVC standard.

2.2 Optimal PMV Selection at the Encoder Side

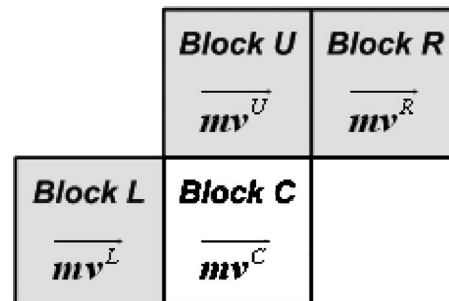To select an optimal PMV among the CS, we define an optimal PMV selection function $f(\cdot)$, which is given by
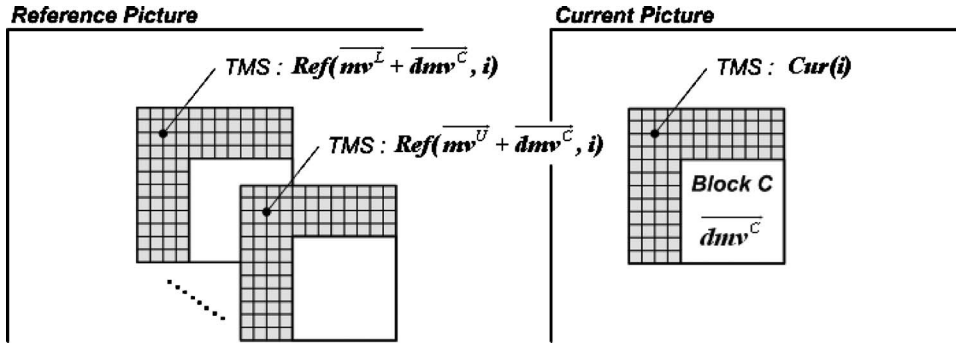

Fig. 1 Candidate PMV set.

Fig. 2 Decoder-side PMV estimation using template matching.

$$f(\mathbf{pmvc}^C|\mathbf{mv}^C) = r(\mathbf{dmv}^C)$$
$$= r(mv_x^C - pmvc_x^C, mv_y^C - pmvc_y^C), \qquad (2)$$

where $r(\cdot)$ is a measure function of the number of bits consumed to encode the differential motion vector (DMV) $\mathbf{dmv}^C$, and $\mathbf{pmvc}^C$ is a possible PMV candidate for encoding $\mathbf{mv}^C$. Consequently, the optimal PMV $\mathbf{pmv}^{C(\mathrm{opt})}$ with minimal bits is given by

$$\mathbf{pmv}^{C(\mathrm{opt})} = \arg\min_{\mathbf{pmvc}^C \in CS} f(\mathbf{pmvc}^C|\mathbf{mv}^C). \qquad (3)$$

If the optimal PMV $\mathbf{pmv}^{C(\mathrm{opt})}$ is more precise than the median PMV $\mathbf{pmv}^{C(\mathrm{med})}$, the coding performance can be improved by saving bits for encoding MV information.

2.3 *Optimal PMV Estimation at the Decoder Side*

To estimate an optimal PMV at the decoder with known information of the DMV $\mathbf{dmv}^C$, template matching[5] is applied with a matching criterion function $g(\cdot)$:

$$g(\mathbf{pmvc}^C|\mathbf{dmv}^C) = \sum_{i \in \mathrm{TMS}} [\mathrm{Ref}(\mathbf{pmvc}^C + \mathbf{dmv}^C, i)$$
$$- \mathrm{Cur}(i)]^2, \qquad (4)$$

where Ref refers to the reference picture and Ref ($\mathbf{pmvc}^C$ $+\mathbf{dmv}^C, i$) denotes a pixel indexed by i with respect to location $\mathbf{pmvc}^C + \mathbf{dmv}^C$ at the reference picture. The index i indicates a pixel location at the template matching set (TMS), and Cur (i) denotes a value at the location indexed by i in the current picture. The TMS is a set of spatially adjacent upper, diagonal, and left regions around the given block as shown in Fig. 2.

In the decoding process, all possible PMV candidates in the CS are tested by template matching those pixels indicated by the TMS to find the optimal PMV having the minimum matching error as

$$\mathbf{pmv}^{C(\mathrm{dec})} = \arg\min_{\mathbf{pmvc}^C \in CS} g(\mathbf{pmvc}^C|\mathbf{dmv}^C). \qquad (5)$$

Note that this derivation can also be made in the encoder by the decoding loop. If the following condition is satisfied, the decoder can estimate the optimal PMV autonomously at the decoder side:

$$\mathbf{pmv}^{C(\mathrm{dec})} = \mathbf{pmv}^{C(\mathrm{opt})}. \qquad (6)$$

In this case, it is sufficient to signal the decoder to find the optimal PMV autonomously to reconstruct the MV out of the DMV, which avoids extra bit transmission to signal a particular PMV, in contrast to Laroche et al.'s work.[4]

2.4 *Encoding Mode Decision for Motion Vector Coding*

Three different modes are considered as follows. Firstly, when neighboring MVs are all unavailable ($|CS|=0$) or identical ($|CS|=1$), there is only one choice for PMV selection. Therefore, in this case, called the *exceptional mode*, the encoder has to use the available PMV. Because the decoder can recognize this situation for itself, no other information is sent. Secondly, the case of the *fallback mode*, in which the optimal PMV is the same as the median PMV, can also be autonomously recognized by the decoder; thus no extra signaling for this mode to the decoder is needed either. The decoder uses the median as the predictor to reconstruct the motion vector. Finally, if a block does not belong to either of those two modes, the decoder recognizes it as belonging to the *competing mode*, which requires the decoder to be informed whether it should use the estimated optimal PMV or not. Thus, 1 bit of additional information, called mvop_flag, is needed. If mvop_flag is 1, the DMV is decoded using an optimal PMV obtained by the decoder using the template matching. If mvop_flag is 0, the decoder uses the median PMV in decoding the DMV. In the *competing mode*, finer macroblock partition for better motion compensation could be implemented. For an encoder to make an RD-optimized macroblock partition decision, a slightly modified RD measure function J is used.

3 Experimental Results

To evaluate the performance of the proposed method, we modified the reference software of the H.264/AVC standard. Joint model (JM) version 12.2 reference software was used for modification and comparison. All sequences ["Coastguard," "Foreman," "Carphone," and "TableTennis" (QCIF, 15 Hz) and "Coastguard," "Foreman," "Paris," and "TableTennis" (CIF, 30 Hz)] have their first 300 frames encoded with four quantization parameters (QPs) of 28, 32, 38, and 40. To obtain a more precise MV, no fast motion estimation process was used. The performance of the proposed method was evaluated in terms of the BDBR and

Table 1 Coding performance of the proposed method.

Format	Sequence	BDPSNR (dB)	BDBR (%)
QCIF	"Coastguard"	0.096	−2.606
	"Foreman"	0.191	−3.463
	"Carphone"	0.133	−2.719
	"TableTennis"	0.176	−3.394
CIF	"Coastguard"	0.086	−2.411
	"Foreman"	0.123	−2.799
	"Paris"	0.152	−2.875
	"TableTennis"	0.150	−3.494
	Average	0.138	−2.970

BDPSNR.[6] Those quantities give the average bit rate and PSNR difference of the proposed method compared to the H.264/AVC standard, which always uses median PMV to encode MVs.

Table 1 shows the BDBR and BDPSNR of the proposed method compared to the H.264/AVC standard. As described in Table 1, the experimental results show that the proposed method decreases the number of bits compared with the H.264/AVC standard by about 2.97% on average. This is because the proposed method selects a more precise PMV than the H.264/AVC standard. In particular, some sequences with fast, nonlinear motion such as "Foreman" and "TableTennis" show better performance when a precise PMV is used. With higher QP values, the proposed method works better. This is because motion vector takes a larger bit portion at the lower bit rate, and thus there is more space to improve the coding efficiency.

4 Conclusions

In this paper, we have proposed a new motion vector coding method using optimal PMV selection. By selecting the optimal PMV, which requires minimal bits to encode MV information, the proposed method can decrease the number of compressed MV bits compared with H.264/AVC. In particular, the proposed method is effective on sequences with fast and nonlinear motion activities. If more candidate PMVs are used, the proposed method can be even more effective without requiring additional signaling information.

References

1. JVT of ISO/IEC MPEG and 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* (2003).
2. M. C. Chen and A. N. Willson, "A spatial and temporal motion vector coding algorithm for low-bit-rate video coding," in *Proc. IEEE Int. Conf. Image Processing*, Vol. **2**, pp. 791–794 (1997).
3. S. D. Kim and J. B. Ra, "An efficient motion vector coding scheme based on minimum bitrate prediction," *IEEE Trans. Image Process.* **8**(8), pp. 1117–1120 (1999).
4. G. Laroche, J. Jung, and B. Pesquet-Popescu, "A spatio-temporal competing scheme for the rate-distortion optimized selection and coding of motion vectors," in *Proc. 14th Eur. Signal Processing Conf.* (2006).
5. Y. Suzuki, C. S. Boon, and T. K. Tan, "Interframe coding with template matching averaging," in *Proc. IEEE Int. Conf. Image Processing*, Vol. **3**, pp. 409–412 (2007).
6. G. Bjonteggard, "Calculation of average PSNR differences between RD-curves," *ITU-T SG16, VCEG-M13* (2001).