MULTI-HYPOTHESIS PREDICTION BASED ON IMPLICIT MOTION VECTOR DERIVATION FOR VIDEO CODING

Zhao Wang[*], *Shiqi Wang*[+], *Xinfeng Zhang*[#], *Shanshe Wang*[*], *Siwei Ma*[*]

*Institute of Digital Media, Peking University, Beijing 100871, China
+Department of Computer Science, City University of Hong Kong, Hong Kong, China
#University of Southern California, Los Angeles, California, USA
{zhaowang, sswang, swma}@pku.edu.cn; {sqwang1986,zhangxinf07}@gmail.com

ABSTRACT

Traditional video coding standards H.264/AVC and HEVC adopt inter-frame prediction with single prediction (P frame) and bi-directional prediction (B frame), in which one and two hypotheses are utilized to generate the final prediction. Though more hypotheses have been proved to significantly improve the prediction accuracy, the additional bits used for signaling the motion information may decrease the overall rate-distortion performance conversely. In this paper, we propose a multi-hypothesis prediction scheme based on implicit derivation of motion vector. In the proposed scheme, the first prediction block is generated identically with the method in HEVC. Subsequently, motion estimation is applied on each reference frame by taking the first prediction block as template, and the search results serve as the additional hypothesis candidates. In this manner, the overheads of motion vector can be reduced. Simulation results show that the proposed algorithm can achieve average 1.1% coding gain compared to HEVC.

Index Terms— Multi-hypothesis prediction, motion compensation, motion vector, video coding

1. INTRODUCTION

Inter prediction plays a crucial role in removing the temporal redundancy based on high similarities among successive frames. By taking the previous decoded frames as the predictive signal, the compression of the current frame can be converted into coding the residuals after prediction [1], and entropy coding is adopted to compactly represent the residual signal. Additionally, the relative position of the prediction block compared to the current block, termed as motion vector (MV), is also required to be transmitted [2].

In the earliest video coding standard H.261, the most recently decoded frame serves as the reference frame and hence only one motion vector is searched for the current block to realize motion compensation (MC). Later, multiple reference frames was adopted in H.263, where the optimal motion vector minimizing the matching error was searched from the candidate frames. Moreover, to further investigate the inter-frame coding efficiency, bi-directional prediction, including one motion vector from the forward frame and the other motion vector from the backward frame, has been introduced [3]. The difference between the single MV (one-hypothesis) prediction and the bi-directional (two-hypothesis) prediction is depicted in Fig. 1. The one-hypothesis prediction directly approximates the original block c with p. In contrast, the prediction of two-hypothesis is obtained by combining the two components, such that the condition that each individual component should be similar to the original block is not demanded.

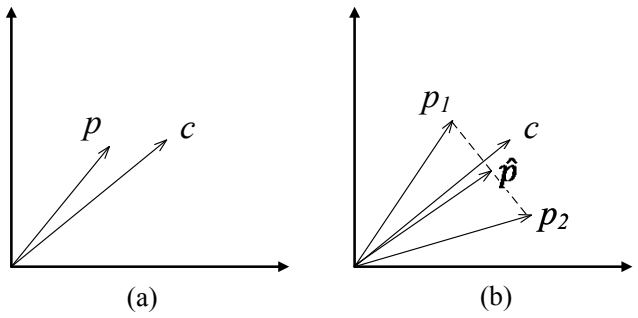

Fig. 1. Illustration of the hypothesis prediction (a) one hypothesis prediction; (b) two hypothesis prediction.

Originated from the two-hypothesis prediction, multi-hypothesis prediction (MHP) has been investigated from spatial and temporal perspectives. For the spatial MHP, overlapped block motion compensation (OBMC) was proposed in [4], which is applied by combining the current motion vector and the motion vectors of the neighbouring blocks to generate the final prediction signal. Regarding the temporal MHP, Sullivan *et al.* pointed out the weakness of the single motion vector prediction and proposed the framework of the MHP in [5]. Performance bound of MHP was studied in [6] by designing a simplified signal mode. In [7], a practically designed algorithm that handled optimal hypothesis selection was introdu-

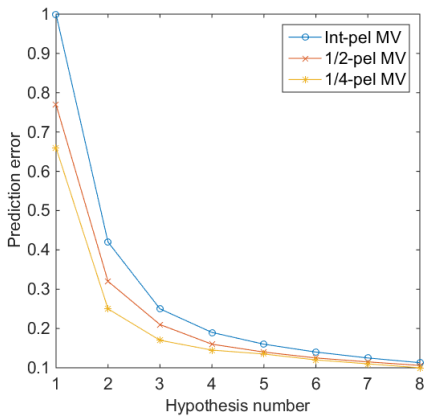

Fig. 2. Prediction error variation in terms of the number of hypotheses.

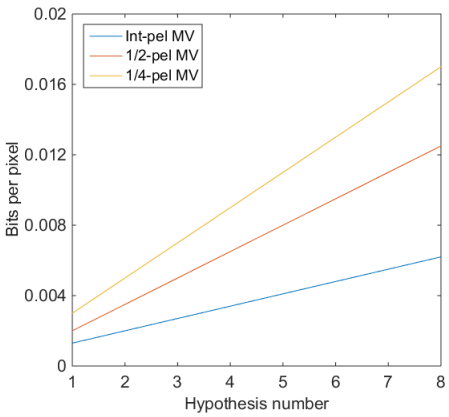

Fig. 3. MV Coding bits variation in terms of the number of hypotheses.

ced. Moreover, in [8]-[9] the rate-distortion performance of MHP and the impact of different hypothesis numbers were further investigated. These works reveal that MHP further investigates the ability of prediction but demands more bits for coding the motion information. To address this issue, we propose an advanced MV derivation method by utilizing the first hypothesis to search for the other hypotheses without introducing the overheads for signaling additional motion information. As such, better prediction accuracy with low signaling overhead can be achieved.

The rest of this paper is organized as follows. In Section 2, we analyzes the variations of the residual energy and bits in terms of the number of hypotheses. Section 3 presents the proposed MHP scheme based on the advanced motion vector derivation method. Simulation results and analyses are presented in Section 4 and Section 5 concludes this paper.

2. RATE-DISTORTION ANALYSIS ON MULTI-HYPOTHESIS PREDICTION

In the inter prediction, motion estimation is first applied on the reference frames to derive the motion vector. Then, the prediction residual can be obtained and processed by transform, quantization and entropy coding. Therefore, the reconstructed distortion of one frame not only affects its own quality, but also has an effect on the future frames. Here, let d denote the reconstructed distortion of the decoded frames and ε_l denote the prediction distortion of the l-th frame. Then, there exists an error propagation relationship that

$$\varepsilon_l = \sum_{i=1}^{n} h_i d_{l-i} \tag{1}$$

where n is the number of hypotheses and h_i represents the weighting factors[10]. This formula indicates that the prediction performance is largely influenced by the number of prediction blocks and the corresponding weighting factors. Here, we focus on studying the influence of hypothesis number and the weighting factors are set to $1/n$ for simplification.

To investigate the MHP scheme, the HEVC reference software HM-16.6 has been modified. To ensure enough hypotheses, we extend the group of picture (GOP) to 16 frames and up to 8 reference frames are supported. To study the influence of hypothesis number on the prediction error, Fig. 2 depicts the curve of prediction distortion when the hypothesis number increases from 1 to 8. Considering the MV resolution also has significant influence on the prediction error, different MV resolutions including integer-pel, 1/2-pel and 1/4-pel, are evaluated. Generally speaking, the MV resolution determines the description precision of the motion trend and higher resolution implies smaller divergence between the used MV and the actual motion [11]. From the Fig. 2, it is obvious that more hypotheses can decrease the prediction error obviously. However, the fact that the error converges to a non-zero value reveals that it is difficult to further improve the prediction quality by increasing the hypothesis number. In Fig. 2, another interesting phenomenon is that MV resolution plays a crucial role under small hypothesis number, *i.e.* 1, 2. However, as more hypotheses are used, the divergence of the prediction error for these three MV resolutions is dramatically suppressed. Moreover, relatively better prediction performance can be achieved by four-hypotheses with just int-pel MV compared to two-hypotheses with 1/4-pel MV resolution. Based on this observation, it is inferred that a certain MV error is tolerable if more hypotheses are used.

Though the increase of the hypothesis number n can reduce the prediction error, more coding bits are desired to represent the additional motion vector. As each hypothesis demands one motion vector, the number of bits required for coding the motion information is linearly proportional to n in general, which is shown in Fig. 3. Generally speaking, the rate of motion information takes up a large proportion of the bit stream, which is only inferior to the rate of residuals. Therefore, the linear growth of motion rate limits the performance of MHP, such that a better scheme that maintains the predict the linear growth of motion rate limits the perfor

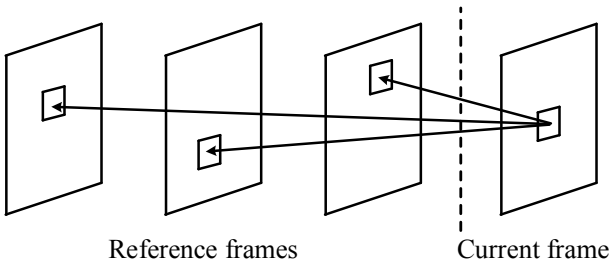

Fig. 4. Illustration of the traditional MHP scheme.

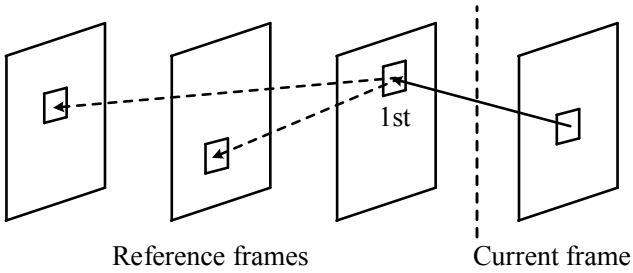

Fig. 5. Illustration of the proposed MHP scheme.

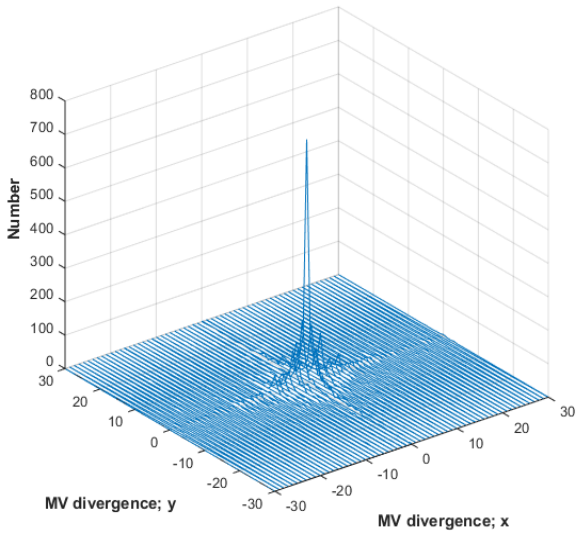

Fig. 6. Distribution of MV divergence between the actual MV and the derived MV.

mance of MHP, such that a better scheme that maintains the prediction accuracy while reducing the signaling overhead is highly desired.

3. PROPOSED MHP SCHEME

According to the analyses in Section 2, the reduction of the rate in conveying motion information is crucial for MHP. In this section, the proposed MHP with implicit motion vector derivation algorithm is detailed.

As shown in Fig. 4, the traditional multi-hypothesis prediction scheme is applied by searching n hypothesis which can be combined into the optimal prediction block, and each hypothesis is indicated by one motion vector. In general, though there exists a certain divergence among successive frames, such as illumination variation, foreground and camera movement, these hypotheses usually share great similarities. By taking one hypothesis as the template, the other hypotheses can be located with motion estimation as the object is usually locally distinct in a frame.

Based on this observation, we propose a MHP framework based on implicit motion vector derivation. As shown in Fig. 5, the first hypothesis is indicated by the explicit MV which is identical with the traditional methods. Subsequently, by taking the first hypothesis as the template, we can apply motion estimation on the other reference frames and the matching results serve as the subsequent hypotheses. In this manner, implicit MV derivation of the additional hypotheses can be realized because the motion estimation can be identically conducted on both the encoder and the decoder. In total, only one MV of first hypothesis and the reference indices of other hypotheses need to be coded.

The proposed scheme is based on the premise that the other hypotheses can be found precisely by the motion estimation with the template of first hypothesis. In other words, the more accurate the derived MV is, the better performance can be obtained. To verify it, we compute the divergence between the derived MV and the actual MV, and the distribution is depicted in Fig. 6. From Fig. 6, it is observed that most MV divergences locate at (0, 0), which reveals that the derived MV is identical with the actual MV. Moreover, almost all MV divergences are smaller than 4 which represents one pixel distance since 1/4-pel MV resolution is used. This observation proves that the first hypothesis could provide sufficiently accurate prediction of the current block in terms of searching for the other hypotheses.

Basically, the encoding computational complexity increases exponentially as the hypothesis number. In view of this, we adopt a two-pass searching method for simplification. In particular, the first pass searches for the first hypothesis and the second pass determines the other hypotheses. More specifically, we conduct motion estimation among all the reference frames to find the optimal hypothesis by minimizing the matching error with the current block. Subsequently, by taking this hypothesis as the template, we apply motion estimation on the other reference frames. If there are n reference frames, we can obtain $n-1$ hypothesis candidates. Finally, we select m $(0 \leq m < n)$ hypotheses to generate the best prediction block. During the motion estimation, we scale the motion vector of the first hypothesis to each reference frame as the searching start point, since linear motion commonly occurs in a short time period.

With respect to the mode signalling, we add the proposed MHP mode to the syntax *inter_pred_idc*. In HEVC, *inter_pred_idc* can be valued as 0, 1 or 2, representing the sin

Table 1. Performance of the proposed algorithm (version 1 with up to two hypotheses).

Sequence	Y	U	V	ET
PeopleOnStreet	-0.2%	-0.4%	-0.3%	100%
Traffic	-0.4%	-0.5%	-0.3%	101%
Kimono	-0.7%	-1.2%	-1.6%	100%
ParkScene	+0.1%	-0.6%	+0.5%	99%
Cactus	-0.4%	-0.2%	-0.3%	102%
BasketballDrill	-0.5%	-0.1%	-0.6%	101%
BQMall	-0.2%	-0.4%	-0.3%	100%
PartyScene	-0.3%	-1.3%	-0.8%	100%
RaceHorsesC	-0.2%	-0.2%	-0.3%	101%
BasketballPass	-0.4%	-0.2%	-0.3%	102%
BQSquare	-0.5%	-1.1%	-2.5%	100%
BlowingBubbles	-0.2%	-0.2%	-0.2%	100%
RaceHorses	-0.3%	-0.6%	-0.4%	99%
FourPeople	+0.2%	-0.3%	-0.7%	100%
Johnny	-0.1%	-0.2%	+0.4%	100%
KristenAndSara	+0.3%	-0.1%	-0.0%	101%
Average	**-0.24%**	**-0.48%**	**-0.43%**	**100%**

Table 2. Performance of the propose algorithms (version 2 with up to four hypotheses).

Sequence	Y	U	V	ET
PeopleOnStreet	-0.6%	-1.9%	-0.5%	106%
Traffic	-1.3%	-1.2%	-2.0%	103%
Kimono	-3.6%	-3.4%	-5.7%	104%
ParkScene	-0.2%	-0.6%	-0.4%	103%
Cactus	-1.3%	-1.7%	-1.4%	107%
BasketballDrill	-1.4%	-0.9%	-1.0%	104%
BQMall	-0.5%	-0.7%	-0.6%	105%
PartyScene	-2.2%	-3.1%	-2.5%	103%
RaceHorsesC	-0.7%	-1.0%	-0.7%	108%
BasketballPass	-1.1%	-0.4%	-0.6%	107%
BQSquare	-2.8%	-3.4%	-4.0%	105%
BlowingBubbles	-0.5%	-0.1%	-0.6%	106%
RaceHorses	-1.1%	-1.1%	-0.8%	104%
FourPeople	-0.1%	-2.3%	-1.8%	106%
Johnny	-0.5%	-0.8%	-1.3%	105%
KristenAndSara	-0.4%	-0.5%	-0.4%	105%
Average	**-1.14%**	**-1.44%**	**-1.52%**	**105%**

gle prediction from List0, single prediction from List1 and bi-prediction, respectively. In our framework, we signal the proposed scheme with *inter_pred_idc* equaling to 3.

4. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed algorithm, the MHP scheme is implemented into HEVC reference software HM-16.6 [12] with Random Access (RA) configuration. The experiments are conducted on the common test sequences under QPs (22, 27, 32, 37) as specified in the HEVC common tests conditions. The coding performance is evaluated in terms of the Bjontegaard-Delta (BD) [13] and the encoding complexity variation is measured by the encoding time (ET).

Here, two versions of the proposed scheme, which employ up to two and four hypotheses respectively, are validated. For version 1, the maximal hypothesis number is limited to two, which implies that only single prediction and bi-prediction are support. Compared to HEVC, the modification of version 1 mainly lies in that the second motion vector is derived implicitly according to the first hypothesis. The performance of version 1 is showed in Table 1, and it is observed that on average 0.24% coding gain can be achieved without encoding complexity increment. This results show that the proposed implicit motion vector derivation method can save the coding bits of MV. Furthermore, we extend the hypothesis number to four in the proposed scheme, and the performance compared to HEVC is shown in Table 2, it is observed that approximately 1.14% coding gain on average can be obtained by the proposed scheme. The performance improvement mainly originates from that more hypotheses provide better prediction signal without the expense of the motion information except for the reference index. In particular, for sequences *Kimono*, *PartyScene* and *BQSquare*, the proposed algorithm can achieve up to 3.6%, 2.2% and 2.8% BD-Rate gain, which reveals that MHP scheme performs better on the textural content.

5. CONCLUSION

We propose a novel MHP scheme to improve the coding performance. The novelty of this paper lies in that the implicit motion vector derivation is adopted to avoid the additional coding bits signaled in indicating the prediction information, such that better prediction accuracy can be obtained without the demand of more coding bits. The impact of hypothesis number on the rate and distortion statistics is first analyzed. The analysis results motivate us to propose the implicit motion vector derivation method, which greatly facilitates the MHP scheme. Extensive experimental results shows that the proposed algorithm can achieve 1.1% BD-Rate gain on average with only marginal complexity increment.

ACKNOWLEDGEMENT

This work was supported in part by National Natural Science Foundation of China (61571017, 61632001), Natural Science Foundation of Guangdong Province, China (2017A030310576)，Top-Notch Young Talents Program of China, which are gratefully acknowledged.

6. REFERENCES

[1] Sugimoto K, Kobayashi M, Suzuki Y, et al. Inter frame coding with template matching spatio-temporal prediction[C].

IEEE International Conference on Image Processing, 2004, 1: 465-468.
[2] Jain J, Jain A. Displacement measurement and its application in interframe image coding[J]. *IEEE Transactions on communications*, 1981, 29(12): 1799-1808.
[3] Wu S W, Gersho A. Joint estimation of forward and backward motion vectors for interpolative prediction of video[J]. *IEEE Transactions on Image Processing*, 1994, 3(5): 684-687.
[4] Orchard M T, Sullivan G J. Overlapped block motion compensation: An estimation-theoretic approach[J]. *IEEE Transactions on Image Processing*, 1994, 3(5): 693-699.
[5] Sullivan G. Multi-hypothesis motion compensation for low bit-rate video coding[C]. IEEE International Conference on Acoustics, Speech, and Signal Processing, 1993, 5: 437-440.
[6] Girod B. Efficiency analysis of multihypothesis motion-compensated prediction for video coding[J]. *IEEE Transactions on Image Processing*, 2000, 9(2): 173-183.
[7] Flierl M, Wiegand T, Girod B. A locally optimal design algorithm for block-based multi-hypothesis motion-compensated prediction[C]. IEEE Data Compression Conference, 1998, 239-248.
[8] Kung W Y, Kim C S, Kuo C C J. Multi-hypothesis motion compensated prediction (MHMCP) for error-resilient visual communication[C]. IEEE International Symposium on Intelligent Multimedia, Video and Speech Processing, 2004: 45-48.
[9] Kung W Y, Kim C S, Kuo C C J. Analysis of multi-hypothesis motion compensated prediction for robust video transmission[C]. IEEE International Symposium on Circuits and Systems, 2004, 3: III-761.
[10] He J, Yang E H, Yang F, et al. Adaptive Quantization Parameter Selection For H. 265/HEVC by Employing Inter-Frame Dependency[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2017.
[11] Wang Z, Wang S, Zhang J, et al. Adaptive progressive motion vector resolution selection based on rate–distortion optimization[J]. *IEEE Transactions on Image Processing*, 2017, 26(1): 400-413.
[12] JVET software repository. Available online: https://jvet.hhi.fraunhofer.de/svn/svn_HMJEMSoftware/
[13] F. Bossen, Common test conditions and software reference configurations, JCTVC-J1100, ITU-T/ISO/IEC Joint Collaborative Team on Video Coding (JCT-VC), Sweden, 2012.