Decoding-Assisted Inter Prediction for HEVC

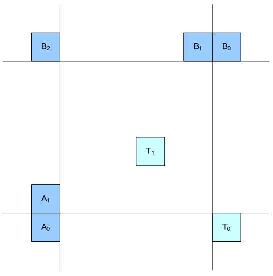
Yi-Sheng Chang and Yinyi Lin
Department of Communication Engineering
National Central University, Taiwan 32054, R.O.C.
Email: yilin@ce.ncu.edu.tw

Abstract — **In this paper, we suggest a decoding-assisted inter prediction algorithm for HEVC to improve its coding performance. In the proposed algorithm, both sum of absolute bi-directional prediction differences (SABPD) and template matching (TM) operations are employed to help merge mode decision and inter mode decision in HEVC. The experimental results reveal that the proposed algorithm outperforms the inter prediction proposed in HEVC, and the results demonstrate that average 2.2% BDBR reduction can be achieved with 11% increase of encoding time, compared to the original HEVC inter prediction.**

I. INTRODUCTION

The newly developed High Efficiency Video Coding (HEVC) achieves much better coding efficiency, compared to the previous coding standard H.264/AVC, due to its complicated coding techniques. For instance, HEVC extended inter/intra prediction with block size up to 64x64 for mode decision instead of 16x16 macroblock and used three hierarchical unit representations (including coding unit (CU), prediction unit (PU) and transform unit (TU)) to obtain the optimal coding efficiency based on the quad-tree structure. HEVC supports four coding units: CU0(64x64), CU1(32x32), CU2(16x16) and CU3(8x8). There are eleven candidate PU modes in each CU level of size 2Nx2N: Merge mode, eight inter modes (2Nx2N, 2NxN, Nx2N, 2NxnU, 2NxnD, nLx2N and nRx2N, and NxN), and two intra modes (2Nx2N and NxN). Both inter NxN and intra NxN are only available in the smallest CU (i.e., CU3).

The HEVC encoder uses five merge candidates obtained from both spatial and temporal blocks for merge mode decision. As illustrated in Fig. 1, among the five candidates, four candidates are selected from spatially adjacent blocks A_1, B_1, B_0 and A_0 (or B_2), and one candidate is selected from temporal blocks T_0 or T_1. If the number of available spatial and temporal merge candidates is less than five, both combined merge candidates and zero-motion vector merge candidates are considered. The five merge candidates are indicated by a merge index that is entropy-coded using a simple prefix code, with associated code words displayed in Fig. 1. The best merge candidate is determined based on rate distortion optimization (RDO) evaluation.



Merge idx	bin
0	0
1	10
2	110
3	1110
4	1111

Fig. 1 Merge candidates used in HEVC merge mode decision

The inter mode decision in HEVC uses three motion vector prediction (MVP) candidates (referred to as advanced MVP, AMVP) for motion estimation (ME). Among them two candidates are chosen from spatial adjacent blocks and one from temporal blocks that are derived from the same positions as shown in Fig. 1, but with different derivation order. The best candidate is similarly determined with RDO technique. The motion vector difference between AMVP candidate and the true MV derived from motion estimation (referred to as MVD) is encoded and transmitted together with other information, e.g., AMVP and reference frame indexes.

For inter coded blocks, a great amount of the bitrates is required for coding MVD information. To improve coding performance of inter prediction, many decoder-side motion derivation schemes have been investigated [1-5]. Some algorithms [1-3] use template matching (TM) at both encoder and decoder sides to derive MVD as well as reference picture information, to reduce encoded bitrates. Others [4-5] use sum of absolute bi-directional prediction differences (SABPD) technique for low bitrates B-frame coding or frame rate up-conversion.

In this paper we propose a decoding-assisted HEVC inter prediction algorithm that uses SABPD and/or TM criteria to both merge mode decision and inter mode decision to improve the coding efficiency. This paper is organized as follows. The proposed merge candidate decision scheme for merge mode decision is introduced in Section 2. Section 3 describes the proposed algorithm. The experimental results are shown and discussed in Section 4 and Section 5 gives conclusion.

II. MERGE CANDIDATE DECISION SCHEME FOR MERGE MODE DECISION

HEVC uses five merge candidates obtained from both spatial and temporal blocks for merge mode decision. Although the spatiotemporal merge mode decision proposed in HEVC is shown to achieve better coding efficiency, compared to temporal direct mode decision or spatial direct mode decision respectively proposed in H.264/AVC, it requires explicit coding the merge index information. In addition, the merge mode decision has high computational load since five RDO operations have to be performed to select the best merge candidate.

In [4], Ates et al. proposed a decoder-derived motion vector estimation scheme in H.264/AVC B-frame prediction to omit explicit coding of MVD information, based on SABPD criterion. To eliminate coding of merge index information we apply the SABPD operation to select the best merge candidate instead of RDO operation. Since SABPD operation can be derived at decoder side, and as a result coding of the merger index is not required. The experimental results showed that the SABPD-based scheme achieves slightly worse coding performance. This is due to that the SABPD technique is a suboptimal one compared to RDO technique, and less than 50% of best candidates are selected based on SABPD. The experimental study also indicates that the SABPD operation is not accurate enough to select the best merge candidate, especially for occlusion areas. In the study we applied the TM operation together with SABPD to select the best candidate and the results showed that its accuracy can be improved.

Another statistical study also indicates that with the RDO technique there is a very high probability of selecting the merge index 0 as best candidate. Based on these observations in this section we suggest a merge candidate decision algorithm for merge mode decision to reduce merge index overhead.

In the proposed algorithm we first apply SABPD/TM operations to select the best merge candidate among all merge indexes except index 0 and compare the best merge candidate with merge candidate index 0 based on the RDO operation. The one with minimum RDO cost is selected. The proposed algorithm only requires a 1-bit flag added in the syntax to signal the best merge candidate, merge index 0 or one from other indexes. If the best merge candidate is from other merge indexes, the decoder then performs SABPD/TM operation at decoder side to find out the best merge candidate. The HEVC encoder uses five merge candidates obtained from both spatial and temporal blocks for merge mode decision. In the proposed algorithm we also consider other merge candidates from temporal blocks in addition to the five candidates used in HEVC. This can enhance its coding efficiency and causes no extra bitrates. Figure 2 shows other merge candidate indexes from temporal blocks in addition to T_0 and T_1, obtained based on their occurrence.

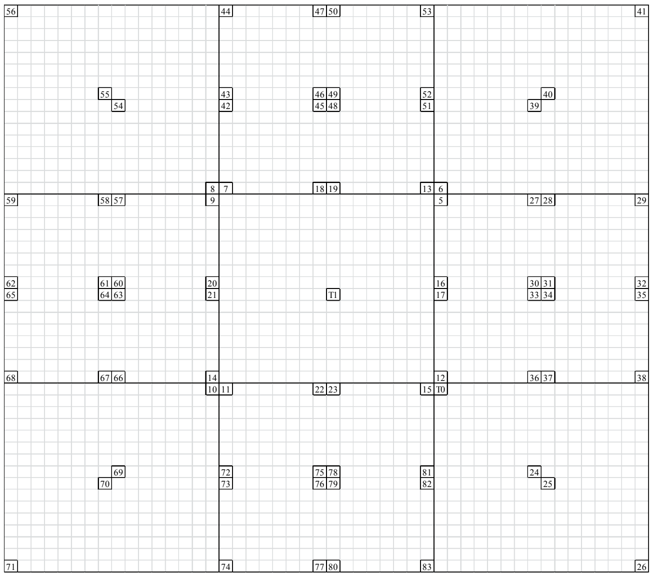


Fig. 2 Extra merge candidate index with N=5, 6,.., 83

In the SABPD/TM operation the best merge candidate among indexes $\{1,2,3 \ldots N\}$ is selected, instead of using RDO technique, by minimizing SABPD/TM between prediction blocks

$$Merge_{best} = \arg \min_{Merge_k} [SABPD_k + TM_{b,k} + TM_{f,k}, k \in \{1,2,3,\cdots N\}]$$

With

$$SABPD_k = \sum_{i,j} \left| B_{t-1,L0}(i + r_k, j + s_k) - B_{t+1,L1}(i - u_k, j - v_k) \right|$$

$$TM_{b,k} = \sum_{x,y} \left| B_t(x,y) - B_{t-1,L0}(x + r_k, y + s_k) \right|$$

$$TM_{f,k} = \sum_{x,y} \left| B_t(x,y) - B_{t+1,L1}(x - u_k, y - v_k) \right|$$

The inverse L-shape template matching region is obtained by extending N/2 pixels from the top and left of the target region for 2Nx2N merge candidate decision, as depicted in Fig. 3 [2].

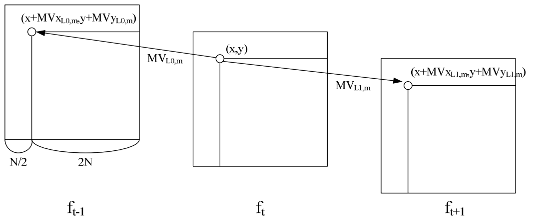


Fig. 3 Template matching

The proposed merge candidate decision scheme for merge mode decision is shown in Fig. 4. The BDBR (BDPSNR) performance is evaluated in the test model HM15.0 with IntelCore i7 3.4GHz and compared for various merge candidate number. The experimental results are displayed in TABLE I. As shown, the proposed scheme can save average 1% of bitrates compared to HEVC with same merge candidates. Furthermore the proposed algorithm achieves average 1.2% bit rate reduction when more candidates are considered. The increases of average encoding time are -1%, 0.3%, 2.3% and 9%. The result reveals that 1% of computation time can be saved in proposed algorithm with same merge candidates. This is because only two (instead of five) RDO operations are performed in the proposed algorithm although some extra SABPD/TM operations are required. There is a competitive computation time for the algorithm with 12 merge candidates, but with 1.2% bitrates saving compared to HEVC. The proposed merge candidate decision scheme with 12 candidates will be used for further study in the proposed inter prediction algorithm.

TABLE I BDBR AND BDPSNR PERFORMANCE

Sequence		BDBR(%)				BDPSNR(dB)			
		5 Cand.	12 Cand.	25 Cand.	85 Cand.	5 Cand.	12 Cand.	25 Cand.	85 Cand.
2560x1600	PeopleOnStreet	-2.078	-2.259	-2.286	-2.270	0.090	0.098	0.099	0.099
	Traffic	-1.085	-1.219	-1.287	-1.271	0.036	0.041	0.043	0.042
1920x1080	BasketballDrive	-0.264	-0.409	-0.398	-0.388	0.005	0.008	0.008	0.008
	Cactus	-1.615	-1.876	-1.890	-1.886	0.035	0.040	0.041	0.040
	BQTerrace	0.045	0.026	0.026	0.011	-0.001	0.000	-0.001	0.000
	Kimono1	-1.521	-1.698	-1.776	-1.773	0.049	0.054	0.056	0.057
	Tennis	0.119	-0.068	-0.025	-0.034	-0.004	0.002	0.001	0.001
	ParkScene	-1.007	-1.114	-1.094	-1.108	0.032	0.036	0.035	0.036
1280x720	SlideEditing	0.411	0.147	0.112	0.134	-0.062	-0.022	-0.016	-0.019
	SlideShow	0.199	0.563	-0.074	-0.235	-0.021	-0.049	0.002	0.015
	vidyo1	-1.803	-1.864	-1.815	-1.937	0.058	0.059	0.058	0.062
	vidyo3	-1.273	-1.373	-1.343	-1.320	0.043	0.046	0.045	0.045
	vidyo4	-1.526	-1.695	-1.713	-1.663	0.044	0.050	0.050	0.048
832x480	BasketballDrill	-0.543	-0.617	-0.560	-0.567	0.022	0.025	0.022	0.023
	BQMall	-0.872	-0.954	-0.992	-0.902	0.035	0.039	0.040	0.037
	Flowervase	-2.180	-2.188	-2.231	-2.301	0.081	0.081	0.082	0.086
	Keiba	-1.044	-1.598	-1.009	-1.192	0.037	0.061	0.038	0.045
	Mobisode2	-1.665	-2.033	-1.691	-1.751	0.038	0.047	0.040	0.040
	PartyScene	-0.485	-0.609	-0.638	-0.634	0.022	0.028	0.029	0.029
	RaceHorses	-0.146	-0.142	-0.148	-0.171	0.005	0.005	0.005	0.006
416x240	BasketballPass	-1.277	-1.383	-1.356	-1.490	0.061	0.066	0.065	0.070
	BlowingBubbles	-1.064	-1.277	-1.283	-1.320	0.042	0.051	0.051	0.053
	BQSquare	-0.616	-0.845	-0.788	-0.795	0.028	0.039	0.036	0.036
	Flowervase	-3.235	-3.489	-3.549	-3.535	0.181	0.195	0.199	0.197
	Mobisode2	-1.485	-1.883	-1.982	-2.015	0.065	0.083	0.087	0.088
	RaceHorses	-0.726	-0.829	-0.850	-0.866	0.033	0.038	0.038	0.039
Average		-1.028	-1.180	-1.178	-1.203	0.037	0.043	0.044	0.045

III. PROPOSED DECODING-ASSISTED HEVC INTER PREDICTION ALGORITHM

In HEVC each CU level uses up to eight inter modes (2Nx2N, 2NxN, Nx2N, 2NxnU, 2NxnD, nLx2N and nRx2N, and NxN) for motion compensated prediction. The MVD between AMVP candidate and the true MV is derived from motion estimation. For inter coded blocks, coding of MVD sometimes requires a significant amount of bitrates. To improve coding performance of inter prediction, the decoder-side motion derivation scheme have been investigated in literatures [1-3].

In this paper we also incorporate both SABPD and TM operations into the inter mode prediction. In the proposed inter mode prediction, in addition to regular motion prediction in HEVC encoder we also perform the both uni-directional prediction and bi-directional prediction to derive the MVD estimates based on TM and SABPD/TM operations respectively. The same search algorithm in regular ME is used. The coded mode is selected through RDO procedure. The 1-bit flag added in the syntax for merge candidate decision can be reused to indicate the MVD in the coded mode is derived from regular ME process or from SABPD/TM process. The block diagram of the proposed HEVC inter prediction is depicted in Fig. 4.

As shown in Fig. 4, the proposed inter prediction algorithm requires extra computations for performing SABPD/TM operations. To reduce its computational complexity, in the proposed algorithm we first perform the regular uni-directional ME process and the one with TM operation, then select the best inter mode. If the best mode is from regular HEVC encoder, further SABPD/TM operations for bi-directional prediction is skipped. The experimental result shows that the early termination scheme does not degrade too much coding performance and average 5% of computation time can be saved.

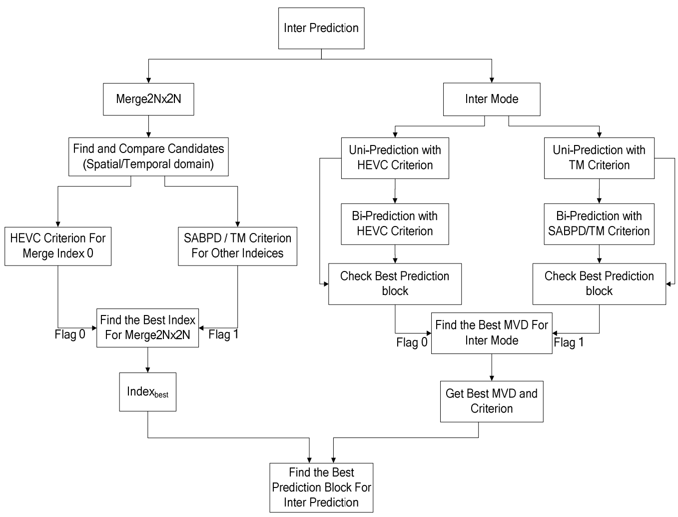


Fig. 4 Proposed HEVC inter prediction algorithm

IV. EXPERIMENTAL RESULTS

In this section, we compare the performance of the proposed inter prediction algorithm with original HEVC encoder. We implement these algorithms into the HM encoder HM15.0 to evaluate their performance. Figures 5 and 6 compare rate-distortion performance for various QPs, tested on *PeopleOnStreet* and *Flowervase* video sequences. Three curves are shown: original HEVC, proposed algorithm with/without early termination scheme. As shown, the proposed algorithm significantly outperforms original HEVC.

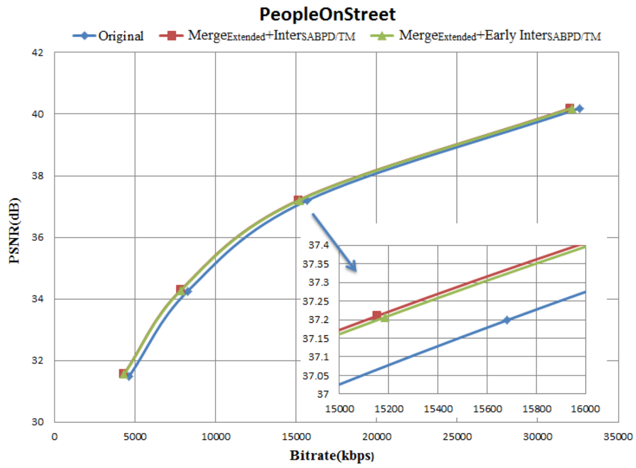


Fig. 5 Rate-distortion performance on *PeopleOnStreet*

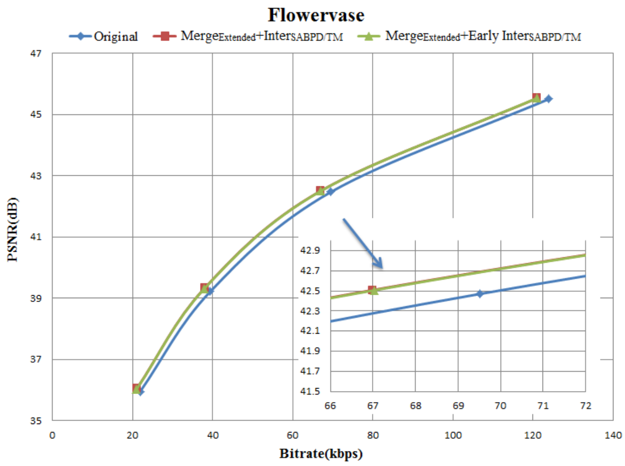


Fig. 6 Rate-distortion performance on *Flowervase*

To obtain further insight, the proposed algorithm was tested on several video sequences of different formats (including classes A, B, C, D and E). We compare their BDBR, BDPSNR performance as well as total encoding time and the results are demonstrated in TABLE II for QP=22, 27, 32 and 37. As shown, the proposed algorithm achieves average 2.3% and 2.2% BDBR reduction (or equivalently 0.1 dB coding gain), compared to the original HEVC inter prediction. The increases of encoding time are 16% and 11% respectively. Average 5% of computation time can be saved with the use of early termination scheme and it does not degrade too much coding performance

V. CONCLUSION

In this paper we suggest a decoding-assisted algorithm for HEVC inter prediction to improve its coding efficiency. The proposed algorithm uses both sum of absolute bi-directional prediction differences (SABPD) and template matching (TM) operations in merge mode decision and inter mode decision in HEVC. The experimental result shows that average 2.2% of bitrates can be reduced compared to the original HEVC, with only 11% increase in encoding time.

TABLE II BDBR, BDPSNR AND COMPUTATION TIME

Sequence		BDBR(%) Merge^{Extended} +Inter$_{SABPD/TM}$	BDPSNR(dB) Merge^{Extended} +Inter$_{SABPD/TM}$	Time(%) Merge^{Extended} +Inter$_{SABPD/TM}$
2560x1600	**PeopleOnStreet**	-4.765	0.211	17.02
	Traffic	-2.863	0.097	18.19
1920x1080	**BasketballDrive**	-1.548	0.032	17.20
	Cactus	-2.913	0.064	12.88
	BQTerrace	-0.465	0.009	17.49
	Kimono1	-2.694	0.087	18.54
	Tennis	-1.496	0.046	19.43
	ParkScene	-2.310	0.075	18.50
1280x720	**SlideEditing**	-0.002	0.006	17.74
	SlideShow	-1.822	0.143	19.07
	vidyo1	-2.894	0.096	18.04
	vidyo3	-2.888	0.100	16.06
	vidyo4	-3.120	0.091	18.13
832x480	**BasketballDrill**	-1.687	0.069	16.24
	BQMall	-2.125	0.088	15.18
	Flowervase	-4.038	0.152	18.56
	Keiba	-2.643	0.103	19.19
	Mobisode2	-2.354	0.056	17.86
	PartyScene	-1.237	0.057	12.37
	RaceHorses	-1.372	0.054	17.69
416x240	**BasketballPass**	-2.425	0.116	12.37
	BlowingBubbles	-2.371	0.095	8.79
	BQSquare	-1.285	0.059	9.97
	Flowervase	-4.763	0.268	14.04
	Mobisode2	-2.664	0.119	13.72
	RaceHorses	-2.126	0.100	11.04
Average		**-2.341**	**0.092**	**15.975**

Sequence		BDBR(%) Merge^{Extended} +Early Inter$^{SABPD/TM}$	BDPSNR(dB) Merge^{Extended} +Early Inter$^{SABPD/TM}$	Time(%) Merge^{Extended} +Early Inter$^{SABPD/TM}$
2560x1600	**PeopleOnStreet**	-4.508	0.200	12.41
	Traffic	-2.712	0.092	10.44
1920x1080	**BasketballDrive**	-1.241	0.027	11.25
	Cactus	-2.762	0.060	9.71
	BQTerrace	-0.449	0.009	11.58
	Kimono1	-2.448	0.079	13.19
	Tennis	-1.366	0.042	14.54
	ParkScene	-2.155	0.070	10.41
1280x720	**SlideEditing**	-0.032	0.010	15.46
	SlideShow	-1.991	0.154	14.54
	vidyo1	-2.759	0.090	12.00
	vidyo3	-2.730	0.095	11.81
	vidyo4	-2.911	0.084	12.41
832x480	**BasketballDrill**	-1.646	0.067	12.17
	BQMall	-2.008	0.083	11.71
	Flowervase	-3.931	0.147	13.68
	Keiba	-2.494	0.096	14.66
	Mobisode2	-2.425	0.057	12.67
	PartyScene	-1.133	0.053	9.87
	RaceHorses	-1.228	0.048	12.20
416x240	**BasketballPass**	-2.190	0.105	8.59
	BlowingBubbles	-2.293	0.092	6.58
	BQSquare	-1.225	0.056	8.07
	Flowervase	-4.680	0.263	11.54
	Mobisode2	-2.226	0.098	9.32
	RaceHorses	-1.996	0.093	6.77
Average		**-2.213**	**0.087**	**11.446**

REFERENCES

[1] K. Sugimoto, M. Kobayashi, Y. Suzuki, S. Kato, and C. S. Boon, “Inter frame coding with template matching spatio-temporal prediction,” in *Proc. IEEE ICIP,* Oct. 2004, pp. 465-468.

[2] S. Kamp and M. Wien “Decoder-Side Motion Vector Derivation for Block-Based Video Coding,” *IEEE Trans. CSVT*, vol. 22, no. 12, pp. 1732-1745, Dec. 2012.

[3] Y. J. Chiu, L. Xu, W. Zhang and H. Jiang, “Decoder-Side Motion Estimation And Wiener Filter For HEVC,” in *Proc. IEEE VCIP*, pp. 1-6, Nov. 2013.

[4] H. F. Ates and B. Cizmeci, “Decoder side true motion estimation for very low bitrate B-frame coding,” in *Proc. IEEE ICIP*, pp. 1673-1676, Sep. 2011.

[5] Y. Ling, J. Wang, Y. Liu and W. Zhang, “Spatial and temporal correlation based frame rate up-conversion,” in *Proc. IEEE ICIP*, pp. 909-912, Oct. 2008.