Uniformly Accelerated Motion Model for Inter Prediction

Zhuoyuan Li, Yao Li, Chuanbo Tang, Li Li, Dong Liu✉, Feng Wu

University of Science and Technology of China (USTC), Hefei, Anhui 230027, China

{zhuoyuanli, mrliyao, cbtang}@mail.ustc.edu.cn, {lil1, dongeliu, fengwu}@ustc.edu.cn

Abstract—Inter prediction is a key technology to reduce the temporal redundancy in video coding. In natural videos, there are usually multiple moving objects with variable velocity, resulting in complex motion fields that are difficult to represent compactly. In Versatile Video Coding (VVC), existing inter prediction methods usually assume uniform speed motion between consecutive frames and use the linear models for motion estimation (ME) and motion compensation (MC), which may not well handle the complex motion fields in the real world. To address these issues, we introduce a <u>u</u>niformly <u>a</u>ccelerated <u>m</u>otion <u>m</u>odel (<u>UAMM</u>) to exploit motion-related elements (velocity, acceleration) of moving objects between the video frames, and further combine them to assist the inter prediction methods to handle the variable motion in the temporal domain. Specifically, first, the theory of UAMM is mentioned. Second, based on that, we propose the UAMM-based parameter derivation and extrapolation schemes in the coding process. Third, we integrate the UAMM into existing inter prediction modes (Merge, MMVD, CIIP) to achieve higher prediction accuracy. The proposed method is implemented into the VVC reference software, VTM version 12.0. Experimental results show that the proposed method achieves up to 0.38% and on average 0.13% BD-rate reduction compared to the VTM anchor, under the Low-delay P configuration, with a slight increase of time complexity on the encoding/decoding side.

Index Terms—Inter prediction, motion estimation (ME), motion compensation (MC), VVC, uniformly accelerated motion model (UAMM).

I. INTRODUCTION

Block-based hybrid coding framework has been widely adopted in the modern video coding standards [1], [2]. To reduce the temporal redundancy in sequential frames, inter prediction makes a major contribution to these standards and plays a key role in the hybrid video coding scheme [3], [4]. Block-based inter predictive process is usually performed on the assumption that the motion between consecutive frames tends to be uniform, and uses the linear model for motion estimation (ME) and motion compensation (MC) to generate the prediction of coding block. However, in natural videos, there are usually multiple moving objects with variable velocity and irregular deformation, leading to complex motion fields. To tackle the efficient representation of complex motion, many motion modeling methods have been intensively studied in the development of video coding standards. These methods can be classified into two categories: spatial/temporal motion modeling-based methods (SMM/TMM).

Spatial motion modeling. In Versatile Video Coding (VVC), most of the inter prediction modes are designed based on SMM. In the spatial domain, the motion information of previous encoded/decoded blocks located around the current coding block is most relevant to the motion information of the current coding block. Based on this assumption, translational-based [3] and deformable-based [4]–[7] spatial motion modeling methods are widely studied. HEVC considers the translation (uniform speed) motion model in inter prediction, such as advanced motion vector prediction (AMVP) and Merge. Based on these two basic approaches, in VVC, MMVD, CIIP, HMVP [8], GPM [9], etc are gradually adopted as the extension modes, and improve the spatial-motion-information reference ability of coding block. In addition, the deformable motion model [5]–[7] is tentatively studied for higher complex motion modeling (such as rotation and zooming) in HEVC, and widely adopted in VVC and integrated with existing inter modes to improve the flexibility of motion information description effectively. Although the existing SMM-based methods can finely characterize the complex motion, the limitation exists that the SMM only considers the spatial correlation and ignores the variable temporal correlation of motion.

Temporal motion modeling. In the consecutive video frames, the motion propagates with the strong temporal correlation. To model the motion in the temporal domain, in HEVC, *temporal motion vector prediction (TMVP)* [10] has been adopted to derivate the temporal motion vector candidate by directly scaling the motion vector stored at adjacent location (T1, T2 in Fig.1) in the col-located picture to the target reference picture. Furthermore, in VVC, TMVP is integrated into more inter modes, and also extended to *sub-block-based temporal motion vector prediction (SbtMvp)* [4] to improve the accuracy of temporal motion vector prediction at the fine-grained block level. Although these two kinds of TMM-based methods bring the temporal motion vector prediction in inter modes, the performance is lower than the SMM-based method due to their simple temporal modeling strategy.

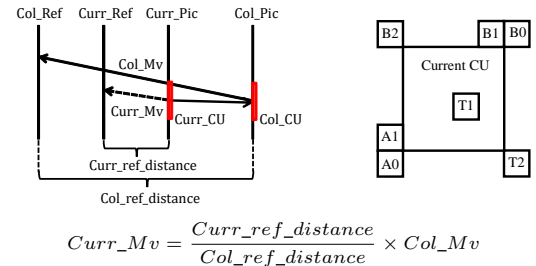$$Curr\_Mv = \frac{Curr\_ref\_distance}{Col\_ref\_distance} \times Col\_Mv$$

Fig. 1. Illustration of temporal motion vector prediction (TMVP) in VVC.
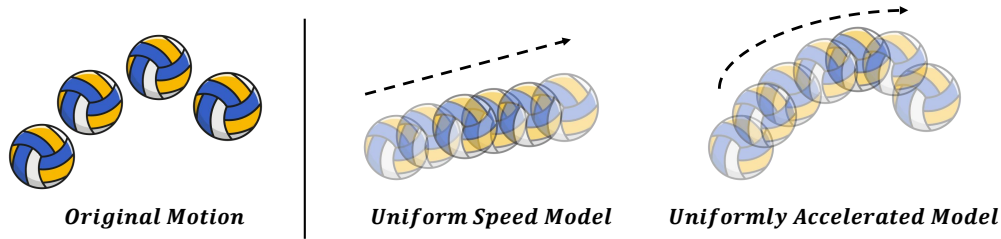
✉: Corresponding author.

Fig. 2. Exploiting the uniformly accelerated model for temporal-domain motion modeling. Here we illustrate the different motion models. For the uniform speed model, the object motion can only be modeled on the linear trajectory (straight line). Compared to the uniform speed model, the uniformly accelerated model is used to model the motion with the velocity and acceleration elements in the x and y directions. The different accelerations of x and y directions enable it to model the motion on the nonlinear trajectory (curve), which makes the uniformly accelerated model enable the stronger representation ability.
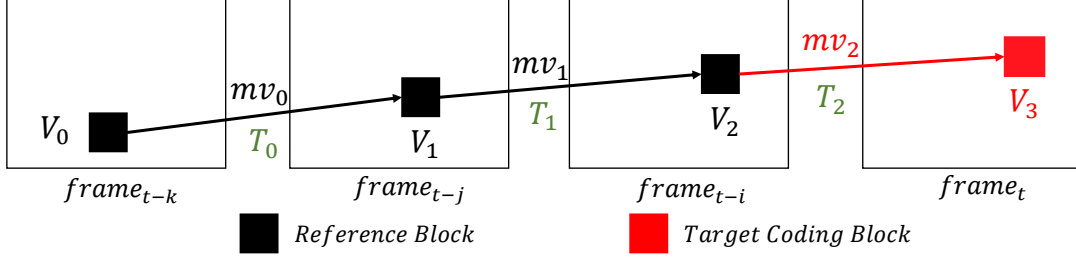

Fig. 3. Illustration of the motion trajectory between the coding (to-be-coded) block and reference (reconstructed) blocks in previous reference frames. The coding block and its motion information are colorized in red, the reference blocks and their motion information are colorized in black.

Based on the limitations of these motion modeling methods, in this paper, we explore the high-order temporal motion model to achieve variable motion-aware temporal motion modeling in video coding scenes. Inspired by [11]–[13], we introduce the uniformly accelerated motion model (Fig. 2) to adapt to more complex motion situations. Rather than the uniform speed (linear) motion model, the uniformly accelerated (high-order) motion model can exploit the motion-related elements (velocity, acceleration) of moving objects, and approximate the complex motion with variable velocity and nonlinear trajectory based on a quadratic function. Specifically, we have made the following contributions that will be detailed in this paper:

- To the best of our knowledge, we are the first to introduce the uniformly accelerated motion model (UAMM) to improve the performance of inter prediction.
- To adapt to the video coding framework, we formulate the usage of UAMM, and propose the UAMM-based parameter derivation and extrapolation schemes in the predictive process.
- We integrate the uniformly accelerated temporal motion model into existing inter modes (Merge, MMVD, CIIP) of VVC to achieve higher prediction accuracy.

II. PROPOSED METHOD

A. Theory of Uniformly Accelerated Motion Model

In the real world, the motion of objects can be formulated as the integration of velocity over time, where the velocity can be expressed as the combination of initial velocity and the integral of acceleration over time:

$$\boldsymbol{x_t} = \int_0^t \boldsymbol{v}\left(t\right) dt, \tag{1}$$

$$\boldsymbol{v}\left(t\right) = \boldsymbol{v}_0 + \int_0^t \boldsymbol{a}\left(t\right) dt, \tag{2}$$

where $\boldsymbol{x_t}$, $\boldsymbol{v}\left(t\right)$, $\boldsymbol{a}\left(t\right)$ denote the object's motion vector (Mv), velocity and acceleration at the t moment, respectively. While

most of the motion variations in the filming scene are not excessively dramatic, the above motion model can be degraded to a uniformly accelerated motion model with the parameter of an initial velocity $\boldsymbol{v}_0$ and a constant acceleration $\boldsymbol{a}$:

$$\boldsymbol{x_t} = \boldsymbol{v}_0 \cdot t + \frac{1}{2} \cdot \boldsymbol{a} \cdot t^2, \tag{3}$$

$$\boldsymbol{v}\left(t\right) = \boldsymbol{v}_0 + \boldsymbol{a} \cdot t, \tag{4}$$

B. Uniformly Accelerated Motion Model-based Parameter Derivation and Extrapolation

Based on the theory of UAMM, here we detail the combination of UAMM and the coding (prediction) process, and propose the derivation scheme of UAMM-related parameters of reconstructed (reference) blocks and the extrapolation scheme of to-be-predicted motion information of the coding block.

1) Derivation of the UAMM-related parameters: To fit the motion trajectory between the coding (to-be-coded) block and reference blocks in previous reference frames, as shown in Fig. 3, the motion information of two reference blocks ($\boldsymbol{mv}_0$, $\boldsymbol{mv}_1$, $\boldsymbol{T}_0$, $\boldsymbol{T}_1$) is used to derive the parameters of the uniformly accelerated motion model in the temporal domain.

Combining the formulas (3), (4), the motion information of the reconstructed block in the third reference frame ($frame_{t-k}$) is served as the initial motion point of motion trajectory, and the correlation of motion information of reconstructed blocks can be constructed as the following equations:

$$\boldsymbol{mv}_0 = \boldsymbol{v}_0 \cdot \boldsymbol{T}_0 + \frac{1}{2} \cdot \boldsymbol{a} \cdot \boldsymbol{T}_0^2, \tag{5}$$

$$\boldsymbol{mv}_1 = \boldsymbol{v}_1 \cdot \boldsymbol{T}_1 + \frac{1}{2} \cdot \boldsymbol{a} \cdot \boldsymbol{T}_1^2, \tag{6}$$

$$\boldsymbol{v}_1 = \boldsymbol{v}_0 + \boldsymbol{a} \cdot \boldsymbol{T}_0, \tag{7}$$

where the $\boldsymbol{mv}_0$ and $\boldsymbol{mv}_1$ represent the motion vector of reconstructed blocks from $frame_{t-k}$ to $frame_{t-j}$ and $frame_{t-j}$ to $frame_{t-i}$, respectively, and $\boldsymbol{T}_0$, $\boldsymbol{T}_1$ denote the time interval
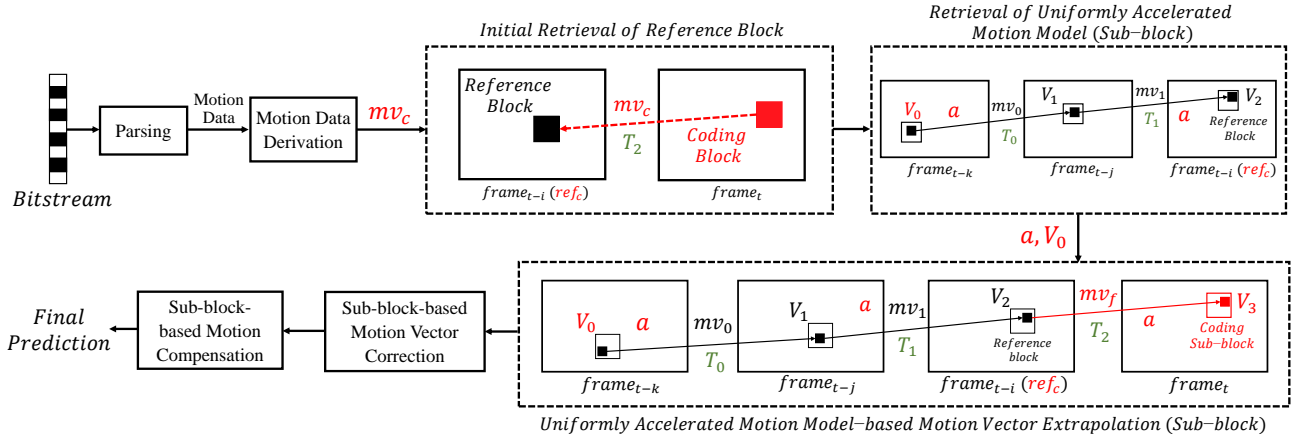
Fig. 4. Illustration of the proposed uniformly accelerated motion model-integrated inter prediction (UAMM) framework from the perspective of decoding. Boxes represent the sub-modules of UAMM, and arrows indicate the direction of the data flow. The coding block and its motion information are colorized in red, the reference (reconstructed) blocks and their motion information are colorized in black.

between these three reference frames, as illustrated in Fig. 3. Note that the time interval between two adjacent frames is a constant in the video, T is directly set as the distance of frame (POC) number. According to the (5), (6), (7), the UAMM-related parameters of reconstructed blocks (initial velocity v_0, acceleration a) can be solved as follows:

$$a = \frac{2\left(mv_1 \cdot T_0 - mv_0 \cdot T_1\right)}{T_0 T_1 \left(T_0 + T_1\right)}, \quad (8)$$

$$v_0 = \frac{mv_0 - \frac{a \cdot T_0^2}{2}}{T_0}. \quad (9)$$

In addition, if the acceleration (a) or initial velocity (v_0) is solved for 0, the UAMM degrades into a linear or constant model.

2) Extrapolation of to-be-predicted motion information: Based on the derivation process, the UAMM-related parameters of each reconstructed (reference) block can be solved and stored for the subsequent prediction process of coding (to-be-coded) block. For the extrapolation of to-be-predicted motion information, we assume that the motion trajectory is obtained with the proper indication, and introduce how to extrapolate the to-be-predicted motion information of the coding block.

To solve the motion information of coding block in UAMM, as shown in Fig. 3, the motion trajectory from the $frame_{t-k}$ to $frame_t$ is taken as a complete motion situation, and the stored acceleration and initial velocity of reconstructed (reference) block are shared to the subsequent coding block. The motion information of the current coding block can be solved as follows:

$$mv_2 = v_2 \cdot T_2 + \frac{1}{2} \cdot a \cdot T_2^2, \quad (10)$$

and further can be written as the follows by decomposing v_2 into v_0, a and time intervals T_0, T_1:

$$mv_2 = v_0 \cdot T_2 + a \cdot T_2 \left(T_0 + T_1\right) + \frac{1}{2} \cdot a \cdot T_2^2. \quad (11)$$

Substituting the solved UAMM-related parameters of the formula (8) and (9) into (11), mv_2 can be extrapolated, which is used to retrieve the reference block of the current coding block. *Note that the parameters are all represented in vector form, and both the horizontal component and the vertical component can be derived simultaneously.*

C. Integration of Uniformly Accelerated Motion Model and Inter Modes in VVC

Based on the above calculation scheme of UAMM-related parameters, here we detail how to integrate the UAMM into the existing inter modes of VVC [3], [4]. The whole prediction process is shown in Fig. 4 from the perspective of decoding.

For the decoding process, first, the motion data of the current coding block is obtained by parsing the bitstream. Concerning the UAMM-related motion data for the coding block, our proposed method is integrated into the existing Merge, MMVD, and CIIP inter modes, and follows the decoding process of these three modes. Second, with the decoded motion information (mv_c, ref_c) obtained by indexing the motion vector candidate, the mv_c and ref_c are used to retrieve the reference block (ending point of motion trajectory) in reference (reconstructed) frames. Furthermore, the stored UAMM-related parameters (v_0, a) of each sub-block of reference block are retrieved and inherited for each sub-block of coding block. Note that the UAMM-related parameters of reference block are derivated in the 4×4 unit granularity and cached into the buffer of motion information, when the coding frame is reconstructed. Third, due to the difference of parameters between the sub-blocks of coding block, the sub-block-level UAMM-based motion vector extrapolation is performed to derive the motion information of each sub-block, respectively. Fourth, with the sub-block-level Mvs obtained, the motion vector correction is used to correct and refine these Mvs in the customized rules, including the constraint between the sub-block Mvs and initial Mv, the consistency of sub-block Mvs. Finally, the sub-block-based motion compensation is performed to generate the final prediction.

For the encoding process, compared to the decoding process, only the process of rate-distortion-optimization (RDO) is increased. For the integration of each mode and UAMM, the UAMM-based motion vector refinement is used as a plug-and-play module after the process of motion vector derivation of each inter mode, without the transmission of any extra signal.

III. EXPERIMENTAL RESULTS

A. Experimental Settings

In our experiment, the VVC reference software VTM-12.0 is used as the baseline. The codec adopts the Low-delay P

Class	Sequence	Low-delay P		
	Name	Y	U	V
ClassB (1920x1080)	MarketPlace	-0.11%	0.43%	0.47%
	RitualDance	0.01%	0.10%	-0.04%
	Cactus	-0.10%	-0.27%	-0.02%
	BasketballDrive	-0.12%	0.05%	-0.03%
	BQTerrace	0.03%	0.01%	-0.04%
ClassC (832x480)	BasketballDrill	-0.16%	0.22%	0.02%
	BQMall	-0.02%	-0.09%	-0.20%
	PartyScene	-0.05%	-0.28%	0.24%
	RaceHorsesC	0.06%	-0.46%	0.70%
ClassD (416x240)	BasketballPass	-0.16%	0.44%	-0.85%
	BQSquare	-0.14%	-0.54%	-0.53%
	BlowingBubbles	-0.15%	-0.82%	-0.12%
	RaceHorses	-0.28%	0.34%	-1.02%
ClassE (1280x720)	FourPeople	-0.35%	-0.49%	-0.43%
	Johnny	-0.38%	0.29%	-1.23%
	KristenAndSara	-0.08%	0.27%	0.14%
Overall		**-0.13%**	**-0.05%**	**-0.17%**
EncT		101%		
DecT		102%		

Sequence	Resolution	LDP (%)
Beauty	1080P	-0.16%
Jockey	1080P	-0.19%
WestLibrary	1080P	-0.15%
WestLabBuilding	1080P	-0.21%
CrossRoad	1080P	-0.14%
CrossRoad2	1080P	-0.12%
Overall	**-0.16%**	
EncT	101%	
DecT	103%	

Sequence	BQMall	BQSquare	Johnny	BlowingBubbles	BasketballPass
BD-rate	-0.19%	-0.35%	-0.11%	-0.11%	-0.15%

(LDP) configuration according to the VVC Common Test Condition (CTC). The test sequences from classes B to E with different resolutions are tested as specified in [14]. For each test sequence, quantization parameter (QP) values are set to 22, 27, 32, 37, and Bjontegaard Delta-rate (BD-rate) [15] is used as an objective metric to evaluate coding performance.

B. Performance

1) Overall Performance Under Common Test Conditions: The R-D performance of the entire UAMM-integrated inter prediction framework on the VVC common test sequences is illustrated in Table I. Y, U, and V represent the R-D performance gain of the three channels of YUV, EncT and DecT represent the encoding/decoding time. We can see that our proposed UAMM can achieve, on average, 0.13%, and achieve up to 0.38% BD-rate reduction (Y component) on VTM-12.0 for all sequences under the LDP configuration, which does not obviously increase the time complexity in the encoding/decoding process. The experimental results show that the proposed framework performs better for sequences with moving objects, such as $BasketballDrill$, $RaceHorses$, $Cactus$, and $Marketplace$.

2) Comparison with the Other Temporal Motion Modeling Methods: In VVC reference software (VTM-12.0), the technology of inter prediction most relevant to temporal motion modeling is the *temporal motion vector prediction (TMVP)* [10] and *sub-block-based temporal motion vector prediction (SbtMvp)* [4]. Both our method and these methods aim to achieve a flexible estimation of the motion field and match the appropriate motion trajectory in the temporal domain. In Table I, *TMVP* and *SbtMvp* are all enabled in the VTM anchor by default, under the LDP configuration. In comparison with anchor, UAMM-integrated inter modes achieve good perfor-

mance with *TMVP* and *SbtMvp* enabled. It can be concluded that UAMM can efficiently deal with some complex motion scenarios that *TMVP* and *SbtMvp* are difficult to address.

3) Performance on the Specific Scenes: According to the performance of UAMM under CTC, the proposed UAMM is beneficial to the scene in that the foreground and background are clearly separated, such as the surveillance-related scene ($Johnny$, $BasketballDrill$, $FourPeople$). To further verify the characteristics and advantages of UAMM, we test UAMM on some selected test sequences of specific scenes from some surveillance-related datasets [16], [17]. Table II shows the performance of these specific scenes. The results demonstrate the stable generalization of UAMM on specific scenes.

4) Bi-directional Prediction Exploration: To further explore the potential of proposed method, we also test our proposed method on the Low-delay B (LDB) configuration, as shown in Table III. The results of some sequences verify the potential of the proposed method in bi-directional setting.

5) Limitation: From the CTC results (Table I), we find that the proposed method is limited to some motion scenes, such as the $BQTerrace$ (tiny objects with camera motion), $RitualDance$ (indistinguishability of foreground and background). Although the proposed acceleration model can handle some complex temporal motion fields to some extent, the characterization of extreme time motion remains a huge challenge in the way of the predefined motion model.

IV. CONCLUSION

In this paper, the uniformly accelerated motion model is proposed to optimize the capacity of temporal motion modeling in VVC. Based on the theory of UAMM, the motion-related parameters (velocity, acceleration) are mined to assist the accurate ME and MC in the temporal domain. The experimental results demonstrate it can achieve good performance with a slight increase of time complexity compared to the VTM-12.0 anchor. For future work, first, we will extend the proposed motion model to improve more inter modes, such as the combination with the more flexible motion estimation schemes [5], [18]–[20], etc. Second, we will further explore the variable accelerated motion model to achieve more flexible motion modeling.

REFERENCES

[1] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, vol. 22, no. 12, pp. 1649–1668, 2012.

[2] B. Bross, Y.-K. Wang, Y. Ye, S. Liu, J. Chen, G. J. Sullivan, and J.-R. Ohm, "Overview of the Versatile Video Coding (VVC) Standard and its Applications," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, vol. 31, no. 10, pp. 3736–3764, 2021.

[3] W.-J. Chien, L. Zhang *et al.*, "Motion Vector Coding and Block Merging in the Versatile Video Coding Standard," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, vol. 31, no. 10, pp. 3848–3861, 2021.

[4] H. Yang *et al.*, "Subblock-based motion derivation and inter prediction refinement in the versatile video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, vol. 31, no. 10, pp. 3862–3877, 2021.

[5] L. Li, H. Li, D. Liu, Z. Li, H. Yang, S. Lin, H. Chen, and F. Wu, "An efficient four-parameter affine motion model for video coding," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, vol. 28, no. 8, pp. 1934–1948, 2017.

[6] K. Zhang, Y.-W. Chen, L. Zhang, W.-J. Chien, and M. Karczewicz, "An improved framework of affine motion compensation in video coding," *IEEE Transactions on Image Processing (TIP)*, vol. 28, no. 3, pp. 1456–1469, 2018.

[7] N. Zhang, X. Fan, D. Zhao, and W. Gao, "Merge mode for deformable block motion information derivation," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, vol. 27, no. 11, pp. 2437–2449, 2016.

[8] J. Li, M. Wang, L. Zhang, K. Zhang, H. Liu, S. Wang, S. Ma, and W. Gao, "History-based motion vector prediction for future video coding," in *2019 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2019, pp. 67–72.

[9] H. Gao, S. Esenlik, E. Alshina, and E. Steinbach, "Geometric partitioning mode in versatile video coding: Algorithm review and analysis," *IEEE transactions on circuits and systems for video technology*, vol. 31, no. 9, pp. 3603–3617, 2020.

[10] K. McCann *et al.*, "High Efficiency Video Coding (HEVC) Test Model 13 (HM 13) Encoder Description," *JCT-VC document, JVET-O1002*, 2018.

[11] X. Xu, L. Siyao, W. Sun, Q. Yin, and M.-H. Yang, "Quadratic video interpolation," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 32, 2019.

[12] Y. Liu, L. Xie, L. Siyao, W. Sun, Y. Qiao, and C. Dong, "Enhanced quadratic video interpolation," in *Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part IV 16*. Springer, 2020, pp. 41–56.

[13] R. Yang, R. Timofte, and L. Van Gool, "Advancing learned video compression with in-loop frame prediction," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, 2022.

[14] K. Suehring and X. Li, "JVET Common Test Conditions and Software Reference Configurations," *JVET document, JVET-G1010*, 2017.

[15] G. Bjontegaard, "Calculation of Average PSNR Differences between RD-curves," *ITU-T VCEG-M33, April, 2001*, 2001.

[16] Xiph Video Test Media, "https://media.xiph.org/video/derf/."

[17] K. Yang, D. Liu, Z. Chen, F. Wu, and W. Li, "Spatiotemporal generative adversarial network-based dynamic texture synthesis for surveillance video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 1, pp. 359–373, 2021.

[18] L. Li, H. Li, Z. Lv, and H. Yang, "An affine motion compensation framework for high efficiency video coding," in *2015 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2015, pp. 525–528.

[19] Y. Li, Z. Li, L. Li, D. Liu, and H. Li, "Global homography motion compensation for versatile video coding," in *2022 IEEE International Conference on Visual Communications and Image Processing (VCIP)*. IEEE, 2022, pp. 1–5.

[20] Z. Li, Z. Yuan, L. Li, D. Liu, X. Tang, and F. Wu, "Object segmentation-assisted inter prediction for versatile video coding," *arXiv preprint arXiv:2403.11694*, 2024.