Global Homography Motion Compensation for Versatile Video Coding

Yao Li, Zhuoyuan Li, Li Li*, Dong Liu, Houqiang Li
CAS Key Laboratory of Technology in Geo-Spatial Information Processing and Application System,
University of Science and Technology of China, Hefei 230027, China
{mrliyao, zhuoyuanli}@mail.ustc.edu.cn, {lil1, dongeliu, lihq}@ustc.edu.cn

2022 IEEE International Conference on Visual Communications and Image Processing (VCIP) | 978-1-6654-7592-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/VCIP56404.2022.10008901

***Abstract*—In Versatile Video Coding (VVC), local affine motion compensation (LAMC) is adopted to handle complex motions, such as rotation and zooming. However, it is inefficient to use LAMC to handle the global motion due to the following two reasons. First, the use of LAMC may lead to some extra bit cost on the affine motion model parameters. Second, the precision of LAMC is restricted by the MV precision of the control points. Therefore, in this paper, we propose a global homography motion compensation (GHMC) framework to better characterize the global motion. For each coding block, an extra mode is added to perform motion compensation based on an 8-parameter global homography motion model. In addition, an extrapolation scheme is designed to derive the parameters from reference frames to save the bit cost for signaling them. The proposed framework is implemented into the VVC reference software VTM-6.0. Experimental results show that, on average, 0.69% and 0.66% BD-rate reduction is achieved under Low Delay P and Low Delay B configurations, respectively, for sequences with rich complex global motions.**

***Index Terms*—Global motion compensation, homography motion model, versatile video coding**

I. Introduction

Motion compensation prediction (MCP) is one of the fundamental techniques in video coding, which aims to remove the temporal redundancy between neighboring frames. In almost all the existing video coding standards, such as H.265/High Efficiency Video Coding (HEVC) [1] and H.266/Versatile Video Coding (VVC) [2], block-based motion estimation (ME) and block-based motion compensation (MC) are widely utilized to perform MCP, in which a motion vector (MV) is employed to obtain the prediction block. However, the underlying motion model of block-based MC, the translational motion model, is too simple to represent complex motions in natural videos, such as rotation and zooming.

To handle this issue, affine motion compensation (AMC) techniques [3] were developed. In particular, local affine motion compensation (LAMC) including both 4-parameter [4] and 6-parameter affine motion models is adopted into VVC to handle complex motions in natural videos. In LAMC, the MC is performed for each 4×4 sub-block, whose MV is derived from control point motion vectors (CPMVs). Though LAMC can describe the local complex motion accurately, it is inefficient to use LAMC to handle the global motion such as camera perspective motions due to the following two reasons. First, the use of LAMC may lead to some extra bit cost on the affine motion model parameters. Second, the precision of LAMC is restricted by the MV precision of the control points.

Over the past decades, there have been many methods proposed to handle the global motion. As early as 1993, Seferidis *et al.* [5] found that high-order motion models were more suitable for addressing complex global motions compared with the translational ones. Later, Dufaux *et al.* [6] proposed a general ME method for various global motion models, from simple translation to 8-parameter perspective models. They applied the global motion estimation (GME) algorithm and global motion compensation (GMC) to MPEG-4. After that, Wiegand *et al.* [7] proposed to use a set of global affine motion models to generate several warped reference frames for a better prediction. In addition, to reduce the overhead bits, Li *et al.* [8] employed a 4-D vector quantizer to code the affine motion parameters more efficiently. Furthermore, Heithausen *et al.* [9] presented a high-order MC extension to HEVC by integrating the motion parameter estimation, interpolation, and parameter coding into the HEVC framework. In summary, many methods have been proposed to conduct GMC based on previous coding standards. However, the performance of GMC is unknown when it is applied to VVC, which has sophisticated inter prediction tools designed especially for complex motions.

In this paper, we propose a global homography motion compensation (GHMC) framework to conduct GMC in VVC, enabling the codec to handle the global motion compensation prediction efficiently. For each coding block, an extra mode is added to perform motion compensation based on GHMC. Moreover, instead of transmitting the motion model parameters explicitly, we propose to extrapolate the parameters from reference frames to save some header bits. We implemented the proposed method into the VVC framework. Experimental results show that our methods can achieve, on average, 0.69% and 0.66% BD-rate reduction under Low Delay P (LDP) and Low Delay B (LDB) configurations, respectively, for sequences with rich complex global motions.

The rest of the paper is organized as follows. In Section II, we introduce the proposed GHMC framework in detail. In Section III, the experimental results are shown and discussed. Section IV concludes this paper.

*Corresponding author

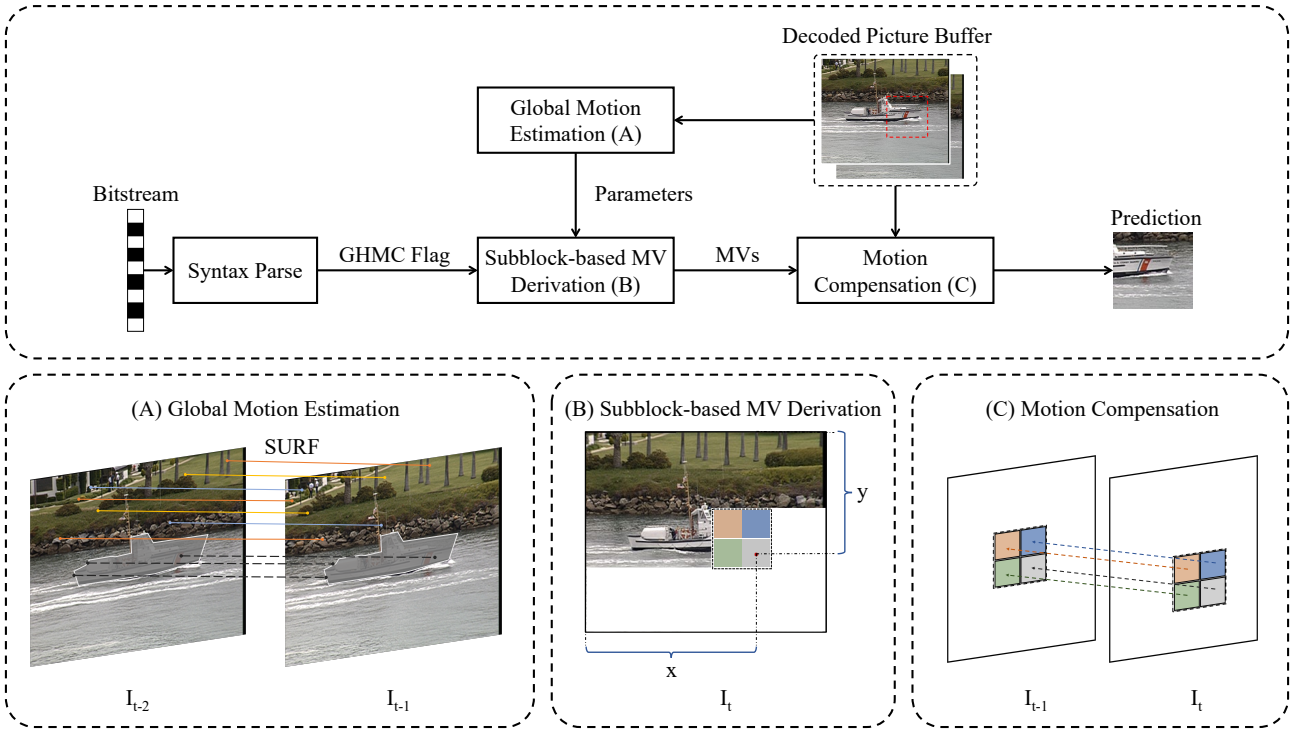

Fig. 1. Structure of the global homography motion compensation framework. In (A), SURF features are utilized to estimate the global homography motion parameters between the last two frames, and moving foreground objects marked in gray are filtered out by RANSAC. In (B), the current CU to be decoded is divided into 4×4 size sub-blocks. The MVs of sub-blocks are derived from global motion parameters and their central samples' coordinates and are further used to conduct subblock-based motion compensation, as shown in (C).

II. Proposed Algorithm

Fig. 1 gives an overall structure of the proposed GHMC framework. In the decoder, when the GHMC Flag is 1, the last two decoded pictures are sent to the Global Motion Estimation module, where Speeded Up Robust Features (SURF) [10] extraction and matching, along with Random Sample Consensus (RANSAC) [11] fitting, are performed to estimate a set of global motion parameters for the current frame. Subsequently, the current CU to be decoded is partitioned into 4×4 sub-blocks, whose MVs are then derived from the obtained parameters and their coordinates. Finally, in the Motion Compensation module, sub-blocks are obtained from the last decoded frame according to their MVs. In the following, we will introduce these three modules in detail.

A. Global Motion Estimation

1) Homography: The global motions in videos are usually caused by camera motions. Under the global motion assumption, the backgrounds in two adjacent frames, which have only global motions, can be regarded as the projections of the same background on two camera image planes in different poses, as shown in Fig. 2. In this work, we use homography to describe the correspondence between pixels that lie on these two image planes.

Homography is an invertible mapping of points on the projective plane, which can be represented by a non-singular matrix $H \in \mathbb{R}^{3\times3}$. When the camera does not have motions along the z-axis or the background scene is on a single plane, the H can perfectly describe the mapping between backgrounds in adjacent frames. For an arbitrary point $\mathbf{x}$ on an image, its corresponding point on another image equals $H\mathbf{x}$,

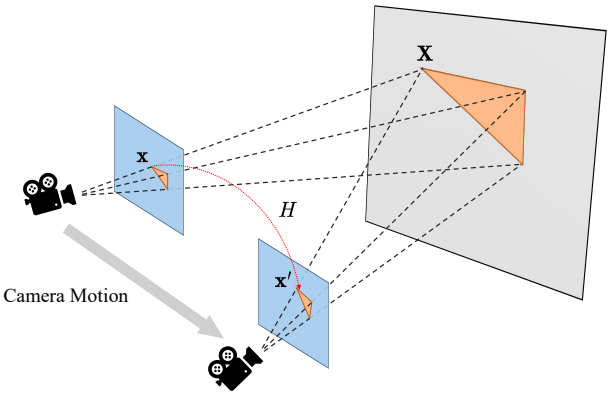

Fig. 2. Illustration of homography. Projections of the same point on camera image planes in different poses can be mapped with H, which can be denoted as $\mathbf{x}' = H\mathbf{x}$

$$s\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{1}$$

where $(x, y, 1)$, $(x', y', 1)$ are the homogeneous coordinates of $\mathbf{x}$, $\mathbf{x}'$, respectively, and s is a non-zero scaling factor. Since this equation is homogeneous, H has only eight degrees of freedom though it contains nine elements.

2) Feature-based Motion Parameter Estimation: After the determination of motion model, the key problem becomes how

to perform GME to determine the model parameters H. Many researches have been conducted to determine H and can be roughly classified into two classes: feature-based ones and image-intensity-based ones. Considering that the former one has less computational complexity, in this paper, we adopt a feature matching scheme followed by a robust estimation method to perform GME.

In this work, SURF [10] extraction and matching are employed to generate the point correspondences. Additionally, point correspondences filtering is also applied to decrease the noise by comparing the Euclidean distance of the feature descriptors of the point pairs with a predefined threshold. Hartley et al. [12] pointed out that one point correspondence could provide two equations for solving the 8 parameters in H. Hence at least four non-collinear point correspondences are required to compute the motion model. Usually, the number of point correspondences obtained by feature matching is far greater than four, and that leads to an overdetermined equation for solving H. To address this problem, RANSAC [11], a widely used robust motion model parameters estimation algorithm is applied subsequently to obtain a group of robust motion parameters.

3) Parameter Extrapolation: At the decoder, the homography motion parameters H are required to perform GHMC. They can be explicitly transmitted in the picture header or slice header. In this way, the model parameters may cost many bits especially for the low-resolution sequence under low bitrate scenarios if we transmit them explicitly. Without loss of precision, the bit cost for each frame is $32 \times 8 = 256$ bits, which is quite expensive for some frames.

To save the bit cost, a motion parameter extrapolation method is proposed in this work. Considering the motions between adjacent frames have high temporal correlations, we propose to estimate the global homography motion parameters H between adjacent frames using the previously coded frames. After the reconstruction of one frame except for the first one, H is calculated between the frame and its nearest reference frame, which will then be applied to the next to-be-encoded frame. We use the nearest reference frame to estimate the model parameters as it has the highest correlation with the current frame and is usually the most widely used one. In this way, the motion model of the current frame can be generated without any bit consumption.

B. Sub-block based MV derivation

To adapt the proposed method to a standard-based framework, we need to use the above method to calculate a MV for each pixel. With the homography motion model in (1), the MV of pixel (x, y) can be calculated by

$$\begin{cases} MV_x = \frac{h_{11}x+h_{12}y+h_{13}}{h_{31}x+h_{32}y+h_{33}} - x \\ MV_y = \frac{h_{21}x+h_{22}y+h_{23}}{h_{31}x+h_{32}y+h_{33}} - y \end{cases} \quad (2)$$

The (x, y) here is the global coordinate in a frame instead of a local coordinate within a coding block as shown in Fig. 1. The derived MV is then rounded to 1/16 and 1/32 precisions for the Luma and Chroma components, respectively.

C. Motion Compensation

Theoretically, we can perform pixel-wise MC to achieve the best performance. However, the encoding complexity can be quite high in this way. In modern video coding standards before VVC, the basic unit to perform MC is a prediction unit. With the use of the affine motion models in VVC, each pixel can have different motions within a coding unit. The 4×4 block is used as the basic unit for MC to obtain a tradeoff between the complexity and the performance. We follow VVC to perform MC for each 4×4 block. The MV of the central pixel of the current 4×4 block is calculated using (2) and used for the MC of the current block. The same set of interpolation filters from AMC is employed to obtain the prediction for each 4×4 block. Note that the MV of the current block is stored for predicting the MVs of the to-be-coded neighboring blocks for both translational and affine motion models.

D. Integration into VTM Reference Software

The proposed GHMC method is integrated into VVC reference software VTM-6.0, by adding an additional inter prediction mode. This mode competes with the affine mode and other translational modes and is selected if it outperforms the other modes when calculating the Rate-Distortion (R-D) cost. Both skip mode without transmitting any residues and non-skip mode transmitting some residues are supported under the proposed scheme. A GHMC flag is signaled inside the merge data syntax to indicate whether the current CU uses the GHMC mode or not.

III. Experimental results

A. Experimental Settings

To evaluate our proposed GHMC method, experiments are conducted using VTM-6.0 as an anchor. We use Low Delay P (LDP) and Low Delay B (LDB) coding configurations as the test conditions and test four common quantization parameters (QPs): 22, 27, 32, and 37, following the JVET Common Test Conditions (CTC) [13]. The Bjontegaard Delta-rate (BD-rate) [14] is adopted as a fair R-D performance comparison metric.

Since the proposed GHMC framework aims to provide more accurate predictions for global motion, we selected some sequences with rich global motions to test the performance of the proposed method, following [4]. The details of these sequences are shown in Table I.

B. Results and Analyses

Table II and Table III summarize the BD-rate reduction results for selected test sequences, under LDP and LDB configurations, respectively. Parameter extrapolation guided GHMC (PE-GHMC), and Simple GHMC with motion parameters transmitted are compared to show the effect of the extra motion parameter bits on coding performance. It can be observed that PE-GHMC achieves a better performance than VTM-6.0 anchor, with an average BD-rate reduction of 0.69% and 0.66% BD-rate in Luma component (Y), under LDP and LDB configurations, respectively. The highest gain for PE-GHMC is achieved on sequence "*BlueSky*" under the

TABLE I
CHARACTERISTICS OF THE SELECTED TEST SEQUENCES

Sequence Name	Picture Order Count	Video Characteristic	Video Resolution	Frame Rate
Tractor	591-690	zooming	1920x1080	25
Shields	415-514	zooming	1920x1080	50
Jets	0-99	zooming	1280x720	25
BlueSky	0-99	rotation	1920x1080	25
Station	0-99	zooming	1920x1080	25

TABLE II
BD-RATE RESULTS OF PE-GHMC AND SIMPLE GHMC FOR SELECTED SEQUENCES UNDER LDP

Sequence	PE-GHMC			Simple GHMC		
	Y	U	V	Y	U	V
Tractor	-0.54%	-0.21%	0.01%	0.06%	1.01%	0.67%
Shields	-0.33%	-0.20%	-0.34%	-1.65%	-0.01%	0.41%
Jets	-0.72%	-0.31%	-0.60%	3.06%	3.15%	3.19%
BlueSky	-1.08%	-0.55%	-0.70%	-2.07%	-0.93%	-0.97%
Station	-0.77%	-0.51%	0.44%	-0.57%	-0.29%	0.35%
Average	-0.69%	-0.36%	-0.24%	-0.23%	0.59%	0.73%
EncT	122%			121%		
DecT	1498%			96%		

LDP configuration, as "*BlueSky*" almost only contains global rotation and hence can make the full use of GHMC. Note that with the use of the parameter extrapolation scheme, we achieve consistent BD-rate savings for all tested sequences.

As for the Simple GHMC, the experimental results show that its average coding performance is worse than the VTM-6.0 anchor. The loss can mainly attribute to the extra bit cost of the motion parameters, which is especially evident for small sequences, such as "*Jets*". Through the comparison of the BD-rates of "*Jets*" between PE-GHMC and Simple GHMC, we can see that the proposed parameter extrapolation mechanism can effectively address the problem of extra overhead bits. In addition, we can see from the tables that the Simple GHMC is sometimes better than the PE-GHMC for the sequences "*BlueSky*" and "*Shields*", which also indicates that our proposed PE-GHMC method cannot always estimate the model parameters accurately. We will further consider a better balance between the header bit cost and the model accuracy in the future.

In terms of the complexity, the encoding time ratio of the PE-GHMC compared with the VTM-6.0 anchor is 112% and 122% in LDP and LDB cases, respectively, which shows that the proposed framework will not increase the encoding complexity significantly. However, the decoding time ratio is 1498% and 1253% accordingly, which is caused by the parameter extrapolation process at the decoder. Note that the Simple GHMC scheme leads to less decoding complexity compared with the VTM-6.0 anchor. We attribute this to more large-block MCs are performed due to the use of the proposed mode.

To better explain the performance brought by the proposed framework, we compare the mode selection results of GHMC and LAMC in Fig. 3. The red and blue blocks indicate the

TABLE III
BD-RATE RESULTS OF PE-GHMC AND SIMPLE GHMC FOR SELECTED SEQUENCES UNDER LDB

Sequence	PE-GHMC			Simple GHMC		
	Y	U	V	Y	U	V
Tractor	-0.57%	0.03%	-0.09%	-0.06%	0.65%	0.70%
Shields	-0.51%	-0.34%	0.51%	-0.86%	0.61%	0.96%
Jets	-0.82%	-1.30%	0.38%	3.89%	3.13%	5.06%
BlueSky	-0.96%	-0.63%	-0.73%	-1.77%	-1.01%	-0.96%
Station	-0.43%	-0.25%	0.54%	-0.08%	-0.14%	0.84%
Average	-0.66%	-0.50%	0.12%	0.23%	0.65%	1.32%
EncT	112%			102%		
DecT	1253%			91%		

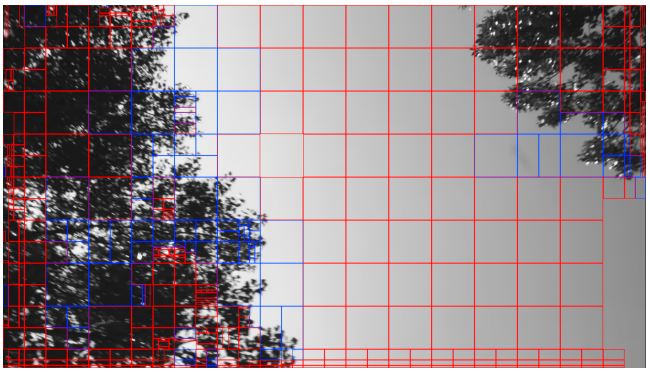

Fig. 3. Mode selection results, BlueSky, LDP, QP32, POC5.

coding units predicted by GHMC and LAMC, respectively. It can be observed that in the sequence with rich global rotation, a large percentage of the blocks choose the GHMC mode. The blocks at the boundary of the leaves and the sky choose GHMC mode less since GHMC can not characterize the local motions of the leaves well.

IV. CONCLUSION

This paper presents a global homography motion compensation (GHMC) framework for VVC to better handle the global motion compensation prediction. For each coding block, an extra mode is added to perform motion compensation based on an 8-parameter global homography motion model. In addition, a motion model parameter extrapolation method is proposed to save the header bits. The experimental results show that our proposed GHMC framework can achieve, on average, 0.69% and 0.66% BD-rate reduction compared with the VVC anchor for sequences with rich global motions, under LDP and LDB configurations, respectively. The experimental results demonstrate the effectiveness of the proposed framework. In the future, we plan to develop an adaptive parameter transmission method to better balance the global motion compensation accuracy and the bit cost of the parameters.

ACKNOWLEDGEMENT

This work was supported by the Natural Science Foundation of China under Grant 62171429 and 62021001, USTC Research Funds of the Double First-Class Initiative Grant YD3490002001, and Fundamental Research Funds for the Central Universities under Grant WK3490000006.

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*, 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*, vol. 31, no. 10, pp. 3736–3764, 2021.

[3] 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.

[4] 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*, vol. 28, no. 8, pp. 1934–1948, 2017.

[5] V. E. Seferidis and M. Ghanbari, "General approach to block-matching motion estimation," *Optical Engineering*, vol. 32, no. 7, pp. 1464–1474, 1993.

[6] F. Dufaux and J. Konrad, "Efficient, robust, and fast global motion estimation for video coding," *IEEE transactions on image processing*, vol. 9, no. 3, pp. 497–501, 2000.

[7] T. Wiegand, E. Steinbach, and B. Girod, "Affine multipicture motion-compensated prediction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 2, pp. 197–209, 2005.

[8] X. Li, J. R. Jackson, A. K. Katsaggelos, and R. M. Merserau, "Multiple global affine motion model for H.264 video coding with low bit rate," in *Image and Video Communications and Processing 2005*, vol. 5685. SPIE, 2005, pp. 185–194.

[9] C. Heithausen and J. H. Vorwerk, "Motion compensation with higher order motion models for HEVC," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2015, pp. 1438–1442.

[10] H. Bay, T. Tuytelaars, and L. V. Gool, "SURF: Speeded up robust features," in *European conference on computer vision*. Springer, 2006, pp. 404–417.

[11] M. A. Fischler and R. C. Bolles, "Random Sample Consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[12] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, 2004.

[13] F. Bossen, J. Boyce, X. Li, V. Seregin, and K. Sühring, "JVET common test conditions and software reference configurations for SDR video," *Joint Video Experts Team (JVET) of ITU-T SG*, vol. 16, pp. 19–27, 2019.

[14] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," *VCEG-M33*, 2001.