장기전역움직임보상을 위한 움직임정보 오버헤드감소방법

Thuc Nguyen Huu, 전병우
성균관대학교 전자전기컴퓨터공학과
thuckechsu@skku.edu, bjeon@skku.edu

Reducing Motion Coding Overhead for Long-term Global Motion Compensation

Thuc Nguyen Huu and Byeungwoo Jeon
Department of Electrical and Computer Engineering
Sungkyunkwan University

Abstract

Long-term global motion compensation (LT-GMC) was designed to compensate camera motion effectively. The LT-GMC warps a reference picture according to an estimated affine/homography model and stores it in its decoded picture buffer for long-term reference. Most previous works on LT-GMC have focused on improving quality of the warped picture, however, there has been only little consideration on the overhead of its motion coding. In this paper, we address this problem and propose a method, namely Scaling Predictor, to reduce the motion coding overhead for LT-GMC. Our experiment has shown BD-Rate reduction of 1.40% over conventional LT-GMC scheme by applying the proposed method.

1. Introduction

Motion compensation plays a key role in reducing the temporal redundancy in video sequence. To carry out the compensation, current-state-of-art video coding standards such as H.264 or HEVC [1] [2] adopted the block-based translation model in which a predicted block for a current block is copied directly from the decoded picture buffer. Despite of many sophisticated motion models such as rotation, shear, zoom, perspective change, most current video coding technologies have relied on the translation model only mainly for simplicity. In many cases, the motions in the video actually come from the camera movement. For example, a camera moving forward or backward (which result in zooming in/out) creates motion in a picture showing gradual enlargement or shrinkage of image, and a camera altering its optical axis causes perspective motion. In order to enhance the motion compensation efficiency beyond the translation model, a technique for long-term global motion compensation (LT-GMC) [3] was developed.

The idea of LT-GMC is to have, in the decoded picture buffer, a long-term reference picture which is obtained by warping a picture according to an affine/homography motion model [4]. An encoder can optimize its rate-distortion performance by adaptively selecting either the long-term (LT) reference or the short-term (ST) reference picture [1]. Although the LT-GMC scheme is not adopted in the major current video coding standards such as HEVC yet, in some specific situations where camera motion occurs frequently, the scheme may be very effective in improving coding performance. As a purpose of its usability test, we implement LT-GMC on top of HEVC and observed significant coding gain (TABLE I). The results in TABLE I suggests that LT-GMC is effective at least in some particular applications.

The performance of LT-GMC highly relies on the accuracy of homography model. Some studies on LT-GMC, [5] [6] [7], have tried to improve the affine/homography model estimation for achieving better warped reference picture. Recently, as pointed out in [8] and [9], single homography model is not sufficient to model motion of complicated scene. This leads to using multiple affine/homography models to generate more reference pictures. In this paper, instead of improving homography model or using multiple models, we investigate improving LT-GMC in the aspect of more efficient signaling of translational motion between LT reference (warped picture) and ST reference.

The paper is organized as follows. First, we explain the problem of LT-GMC in Section 2. Following, we propose a solution, namely, Scaling Predictor, in Section 3. Experimental results are shown in Section 4, and finally conclusion is given in Section 5.

2. Problem analysis

In LT-GMC, the decoded picture buffer has two types of reference pictures: 1) ST reference and 2) LT reference pictures. The LT reference is created by warping ST reference according to affine/homography model. An encoder can select either ST or LT reference through rate-distortion optimization. To reduce the overhead of motion signaling, the encoder selects the best motion vector predictor (MVP) among the motion vectors of its neighbors. After selecting a MVP, the motion vector difference, MVD=MV-MVP is encoded [1] [2].

Let's denote the cases of referencing ST reference and LT reference, respectively, by a short-term motion (STM) and long-term motion (LTM). Using an encoder integrating LT-GMC, we investigate availability of the MVP by considering motion types (STM or LTM) of current block and its neighboring blocks. TABLE II shows that MVP is not available in case of different types of motion. Note that the bit overhead for MVD signaling increases in cases 3 and 4. Moreover, we further observe that such two cases happen about 10%. It motivated us to design an efficient method to deal with the unavailable MVP cases. The proposed method is named as scaling predictor.

TABLE I. PERFORMANCE OF LT-GMC WITH RESPECT TO CAMERA MOTION.

Sequence	Camera motion	BD-Rate reduction
Station2	Zoom	-10.25%
LakeWalking	Massive shaking	-19.99%
Blue_sky	Rotation	-15.68%

TABLE II. CASE STUDY OF MVP AVAILABILITY

Case no.	Current	Neighbor	MVP
1	STM	STM	✓
2	LTM	LTM	✓
3	STM	LTM	×
4	LTM	STM	×

3. Proposed Scaling Predictor

Without loss of generality, we only consider the homography model (note that the affine model is a special case of the homography model) which is also known as 8-parameters model [4]. A location (x, y) in the current picture is transformed into a new position (u, v) at a target picture by:

$$z\begin{bmatrix}u\\v\\1\end{bmatrix}=\begin{bmatrix}h1 & h2 & h3\\h4 & h5 & h6\\h7 & h8 & 1\end{bmatrix}\begin{bmatrix}x\\y\\1\end{bmatrix}=H\begin{bmatrix}x\\y\\1\end{bmatrix} \tag{1}$$

where H represents the homography model with eight degrees of freedom (h1, h2, …, h8) , and z is an arbitrary scaling factor.

The purpose of the proposed scaling predictor is to unify the motion type of current and neighboring blocks, thus, making MVP available as much as possible. Depending on the case we deal with, we can divide our proposed method into two sub-methods: *backward scaling* (solve the case 3) and *forward scaling* (solve the case 4).

3.1. Backward scaling

This method solves the case 3 where the current block has the STM type but the neighboring block has the LTM type. Here, we want to convert neighboring motion from LTM to STM. Let's suppose a neighboring block with its center position at (x, y) and its LTM motion vector (dx, dy). The goal here is to find a matching/corresponding position of the block with respect to the ST reference, denoted as (x', y'). Once (x', y') is determined, its MVP can be computed by:

$$\begin{cases}MVP_{hor}=x'-x\\MVP_{ver}=y'-y\end{cases} \tag{2}$$

Since the neighboring motion with MV (dx, dy) is of LTM, the matching position with respect to LT reference is $(x+dx, y+dy)$. The LT reference is a warped version of ST reference by the homography model H. Because the homography model is invertible [4] (we denote it as H^{-1}), the ST reference can be also inversely warped from LT reference. Therefore, the matching position with respect to ST reference (x', y') can be computed as follows, where z is an arbitrary scaling factor.

$$z\begin{bmatrix}x'\\y'\\1\end{bmatrix}=H^{-1}\begin{bmatrix}x+dx\\y+dy\\1\end{bmatrix} \tag{3}$$

By using Eq. 2 and 3 sequentially, one can compute MVP and then use it for MVD calculation. The term *backward scaling* can be understood as using the inversed homography model H^{-1} to perform the scaling.

3.2. Forward scaling

In contrast to backward scaling, this method solves the scaling by using an exact homography model H. Here, we handle the case 4 where the current block is of LTM, but the neighboring block is of STM. The derivation is almost identical to the backward scaling except we change Eq. 2 to use the homography model H since it is to convert STM to LTM.

$$z\begin{bmatrix}x'\\y'\\1\end{bmatrix}=H\begin{bmatrix}x+dx\\y+dy\\1\end{bmatrix} \tag{4}$$

Similarly, by using Eq. 2 and 4 sequentially, one can also compute MVP and use it to compute MVD.

4. Experimental results

Experiments are conducted with a rich set of sequences which have vast motion characteristic caused by camera motion such as zooming, rotation, and shaking. Those sequences can be found in [10] and the User Generated Content (UGC) provided by Huawei [11]. The HEVC reference software version 16.18 [12] was used to implement the proposed scaling predictor. For performance comparison, the original HEVC is used as an anchor. The coding condition for experiment is summarized in Table III. We also integrate the LT-GMC into HEVC, and it is used as the second anchor to evaluate the performance gain of the proposed scaling predictor. Since the estimation of the homography model itself is not in our scope, we were guided by the LT-GMC implementation in [7] regarding how to estimate the homography model. The proposed scaling predictor is implemented in HEVC by modifying the MVP generation process in the Advanced Motion Vector Prediction [2] (AMVP).

TABLE III. ORIGINAL HEVC CONFIGURATION

Reference software	HM 16.18
Test sequence	Sequences from Huawei proposal in MPEG
QP	22, 27, 32, 37
Intra period	32
Coding scheme	Random access (GOP size = 16)
Number of short-term references	Limit to one closest reference each from the past and the future

Our test results are in TABLE IV which measures the BD-Rate reduction for our proposed method, that is, the combination of backward and forward scaling. Some observations can be made in TABLE IV. Firstly, LT-GMC is very powerful in sequence having much camera motion – it gives 3.36% BD-Rate reduction on average. Secondly, the LT-GMC employing the proposed scaling predictor outperforms the conventional LT-GMC in every test sequence by roughly 1.40% on average, thus proving its efficiency. Lastly, comparing to conventional LT-GMC, the scaling predictor achieves the biggest gain in sequence of DriveRecoder1 (1.94% BD-Rate reduction) while the smallest gain from the sequence ParkWalking (0.88%).

5. Conclusion

In this paper, we have proposed a novel scaling predictor method to deal with the unavilable MVP problem of LT-GMC. Our future work will investigate more to see what kind of block characteristic could benefit from the proposed scaling predictor.

Acknowledgement

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ICT Consilience Creative program(IITP-2019-2015-0-00742) supervised by the IITP(Institute for Information & communications Technology Planning & Evaluation).

TABLE IV. BD-RATE REDUCTION FOR VARIOUS METHODS (%)

Sequence	LT-GMC	Scaling predictor (Backward + Forward Scaling)	Scaling predictor v.s. LT-GMC
ParkWalking	-4.84%	**-5.72%**	-0.88%
ParkSunny	-1.87%	**-3.46%**	-1.59%
DriveRecorder2	-1.94%	**-3.37%**	-1.43%
DriveRecorder1	-2.57%	**-4.51%**	**-1.94%**
CStoreGoods	-5.57%	**-6.75%**	-1.18%
Average	-3.36%	**-4.76%**	-1.40%

Reference

[1] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," in IEEE Trans. on Circuits and Systems for Video Technology, vol. 13, no. 7, pp. 560-576, July 2003.

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

[3] Smolic, Aljoscha, et al. "Improved video coding using long-term global motion compensation," Proc. Visual Communications and Image Processing (VCIP) 2004. Vol. 5308, 2004.

[4] Dubrofsky, Elan "Homography estimation," Diplomová práce. Vancouver: Univerzita Britské Kolumbie, 2009.

[5] Smolić, Aljoscha, Yuriy Vatis, and Thomas Wiegand "Long-term global motion compensation applying super-resolution mosaics," Proc. ISCE. Vol. 2. 2002.

[6] Dufaux, Frederic, and Janusz Konrad "Efficient, robust, and fast global motion estimation for video coding," in IEEE Transactions on Image Processing 9.3, pp.497-501, 2000.

[7] A. Stojanovic and J. Ohm, "Exploiting Long-Term Redundancies in Reconstructed Video," in IEEE Journal of Selected Topics in Signal Processing, vol. 7, no. 6, pp. 1042-1052, Dec. 2013.

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

[9] Jean Bégaint, Franck Galpin, Philippe Guillotel, Christine Guillemot, "Region-based models for motion compensation in video compression," Proc. of Picture Coding Symposium (PCS), San Francisco, United States, pp.154-158, June 2018.

[10] "Xiph.org test media," http://media.xiph.org/video/derf/

[11] X. Ma, H. Zhang, Y. Zhao, M. Sun, M. Sychev, H. Yang, and J. Zhou, "Huawei test sequences of UGC feature for video coding development," JCTVC-v0093, 22th JVT-VC Meeting, Geneva, Switzerland, Oct 2017.

[12] https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/.