An Edge Aware Motion Modeling Technique Leveraging on the Discrete Cosine Basis Oriented Motion Model and Frame Super Resolution

A. Ahmmed[1,2], M. Paul[2], M. Pickering[1], A. Lambert[1]

[1]University of New South Wales, Australia.
[2]Charles Sturt University, Australia.

Data Compression Conference, 2021

Problem Statement

– Commonality modeling is a critical component in modern video coding standards.

– To capture motion homogeneity between successive frames, the edge position difference (EPD) measure based motion modeling (EPD-MM)[1] has shown good motion compensation capabilities[2].

– The EPD-MM technique is underpinned by the fact that from one frame to next, edges map to edges and such mapping can be captured by an appropriate motion model.

[1] M. Asikuzzaman, A. Ahmmed, M. Pickering, and T. Sikora, "Edge oriented hierarchical motion estimation for video coding", *IEEE ICIP*, 2020, pp. 1221–1225.

[2] A. Ahmmed, M. Paul, and M. Pickering, "Dynamic point cloud texture video compression using the edge position difference oriented motion model," *DCC*, 2021, pp. 335.

Problem Statement

– For high resolution sequences, the baseline EPD-MM approach may fail to estimate reasonably accurate motion model.

– This fact can be attributed to the significant increase of detailed information in high resolution content.

– The EPD-MM employs a single 6-parameter (affine model) or 8-parameter (discrete cosine basis oriented model) to capture the motion of the moving edges.

– For high resolution content, having to fit in the motion of large number of edge pixels can produce an “average” motion model in terms of model accuracy.

Problem Statement

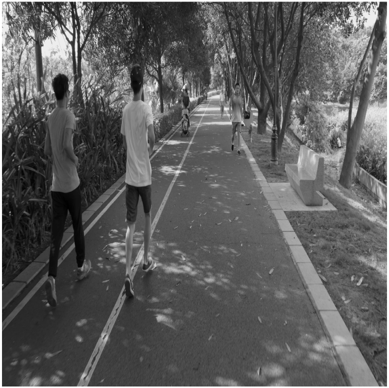

Figure 1: Reference frame, R (POC 23)

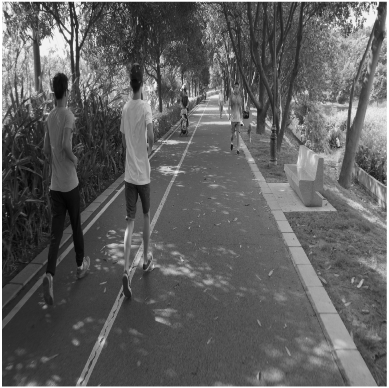

Figure 2: Current frame, C (POC 24)

Frames from the JVET 4K *ParkRunning3* video sequence. The difference frame between R and C has a PSNR of 21.50 dB.

Problem Statement

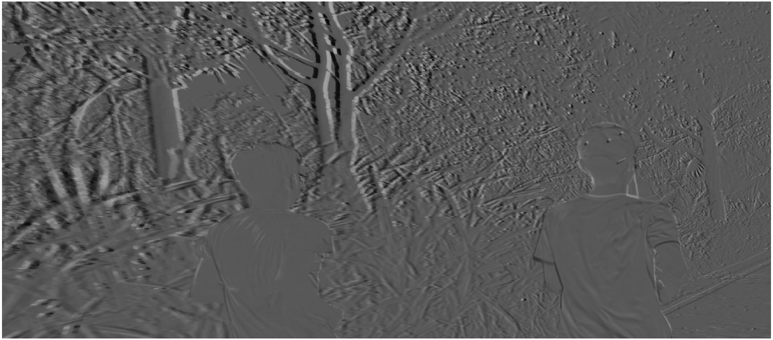

Figure 3: Motion compensated prediction error for the current frame C, $\left(C - \widehat{C}_{dco}^{R\rightarrow C}\right)$: PSNR $=$ 22.86 dB.

The predicted frame $\widehat{C}_{dco}^{R\rightarrow C}$, from the baseline EPD-MM approach[1] fails to compensate motion of some regions/objects in the scene.

[1] A. Ahmmed, M. Paul, and M. Pickering, "Dynamic point cloud texture video compression using the edge position difference oriented motion model," *DCC*, 2021, pp. 335.

Proposed Approach

– noticing that in low resolution version of C, the scene structure, in terms of objects and their relative motion, is present and the motion model estimation process needs to deal with lower number of moving edge pixels, we perform the EPD-MM on lower resolution image.

– for motion modeling, the 8-parameter discrete cosine basis oriented (DCO) motion model[1] is employed.

– the obtained prediction is upsampled back to the original resolution using single image super resolution (SISR) approach.

[1] A. Ahmmed, M. Hannuksela, and M. Gabbouj, "Fisheye video coding using elastic motion compensated reference frames," *IEEE ICIP*, 2016, pp. 2027–2031.

Proposed Approach

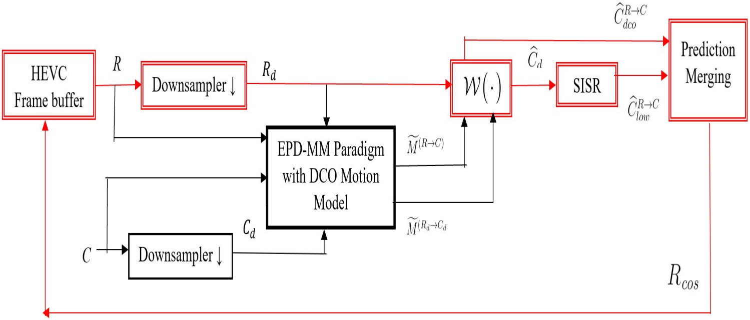
HEVC Frame buffer
R
Downsampler ↓
R_d
$\mathcal{W}(\cdot)$
$\widehat{C}_d$
SISR
$\widehat{C}_{dco}^{R\to C}$
Prediction Merging
$\widehat{C}_{low}^{R\to C}$
EPD-MM Paradigm with DCO Motion Model
$\widetilde{M}^{(R\to C)}$
$\widetilde{M}^{(R_d\to C_d)}$
C_d
C
Downsampler ↓
R_{cos}

Figure 4: A simplified block diagram of the proposed approach[1]. The predicted frame, $\widehat{C}_{dco}^{R\to C}$ is from the baseline EPD-MM approach and the predicted frame $\widehat{C}_{low}^{R\to C}$ is from the proposed approach.

[1] Red colored components are common to both the encoder and decoder; while the black colored components belong to the encoder side only.

EPD-MM in Lower Resolution

– To generate the predicted frame, $\widehat{C}_{low}^{R \to C}$, at first the current frame C and its reference frame R are downsampled, with a downsampling factor of $1/2$.

– It produces the frames C_d and R_d respectively.

– After that their edge maps, $\{c\}$ and $\{r\}$, are extracted to be fed into the following optimization problem defined by for estimating the motion model $\widetilde{M}^{(R_d \to C_d)}$.

$$\widetilde{M}^{(R_d \to C_d)} = \underset{M^{(\{r\} \to \{c\})}}{\arg\min} \quad f\Big(\{c\}, \mathcal{W}\big(M^{(\{r\} \to \{c\})}, \{r\}\big)\Big) \tag{1}$$

herein, $f(\cdot)$, is taken to be the Chamfer distance[1] and $\mathcal{W}(\cdot)$ as the motion compensation operator.

[1] G. Borgefors, "Hierarchical chamfer matching: a parametric edge matching algorithm," *IEEE TPAMI*, vol. 10, no. 6, pp. 849–865, Nov. 1988.

EPD-MM in Lower Resolution

– Next, the estimated 8-parameter DCO motion model $\widetilde{M}^{(R_d \to C_d)}$ is employed to generate an EPD-MM based prediction $\widehat{C}_d^{R_d \to C_d}$ of the downsampled frame C_d.

$$\widehat{C}_d^{R_d \to C_d} = \mathcal{W}\left(\widetilde{M}^{(R_d \to C_d)}, R_d\right) \tag{2}$$

– The obtained frame $\widehat{C}_d^{R_d \to C_d}$ is upsampled back to the original resolution of C using SISR technique and thereby a predicted frame $\widehat{C}_{low}^{R \to C}$ for C is obtained.

– In this regard, we compared the performance (reconstructed frame's PSNR wise and computational time wise) between the bicubic interpolation method[1] and the pre-trained model from the convolutional neural networks (CNN) based SISR approach[2].

[1] W. Siu and K. Hung, "Review of image interpolation and super-resolution," *APSIPA*, 2012, pp. 1–10.

[2] C. Dong, C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE TPAMI*, vol. 38, no. 2, pp. 295–307, 2016.

Results for the Single Image Super Resolution Technique

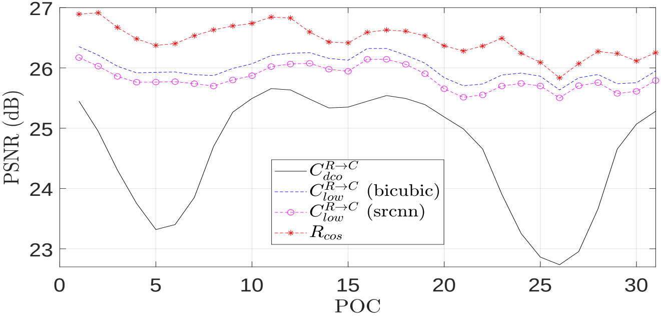

Figure 5: $\widehat{C}_{low}^{R\to C}$ frame wise prediction PSNR values of the *ParkRunning3* sequence from different approaches. Results are shown over the first 32 frames only.

For this work, the bicubic interpolation based upsampling method is adopted as it managed to outperform the pre-trained CNN-based model by 0.17 dB, on average, and has lower computational complexity[1].

[1] The employed system configuration is: Intel Core i7-8650U CPU@1.90GHZ, 32.0 GB RAM.

EPD-MM in Lower Resolution

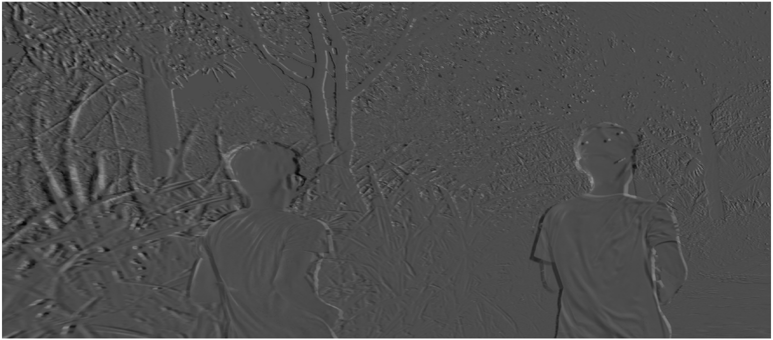

Figure 6: Motion compensated prediction error for the current frame C, $\left(C - \widehat{C}_{low}^{R \to C}\right)$: PSNR $=$ 25.86 dB.

The PSNR of the predicted frame $\widehat{C}_{low}^{R \to C}$ (from the proposed approach) is 3 dB superior compared to that of the predicted frame $\widehat{C}_{dco}^{R \to C}$ (from the existing EPD-MM approach, shown in Fig. 3).

Generation of the Reference Frame R_{cos}

– Fig. 5 points out that the prediction performance of the proposed approach could further be improved by blending the predictions $\widehat{C}_{dco}^{R\to C}$ and $\widehat{C}_{low}^{R\to C}$ together.

– For example, over the foreground objects, $\widehat{C}_{dco}^{R\to C}$ frame has comparatively lower residual energy and in the background region the prediction $\widehat{C}_{low}^{R\to C}$ seems to perform better (Figs. 3 and 6).

– In this regard, the current frame C is partitioned into fixed size blocks of 240×240 pixels and for each such block a co-located block is selected, either from the frame $\widehat{C}_{low}^{R\to C}$ or $\widehat{C}_{dco}^{R\to C}$, as a prediction by maximizing the prediction PSNR.

– Once this process is completed, the resultant frame, denoted herein by R_{cos}, is obtained.

Features of the Reference Frame R_{cos}

$\widehat{C}_{dco}^{R\rightarrow C}$	$\widehat{C}_{low}^{R\rightarrow C}$	R_{cos}
22.86 dB	25.86 dB	26.09 dB
144 bits	132 bits	420 bits

Table 1: PSNR and bit requirements for different predicted frames for the running example current frame, C.

– On average, the prediction PSNR from the proposed approach (R_{cos} frames' PSNR) is 1.85 dB superior compared to that of the baseline EPD-MM approach ($\widehat{C}_{dco}^{R\rightarrow C}$ frames' PSNR).

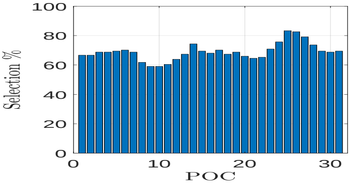

Figure 7: Around 83% of the blocks are selected from the $\widehat{C}_{low}^{R\rightarrow C}$ frame for this particular R_{cos} frame (POC 25).

Experimental Analysis

– To utilize the motion compensation feature of R_{cos} frames, a hybrid coding strategy is adopted.

– For each P-frame, its corresponding R_{cos} frame is generated using the HEVC (HM 16.20) anchor coded reference frame.

– The frame R_{cos} is then employed as an additional reference frame to encode C. That means in the proposed approach, for encoding the frame C, the reference picture list LIST0 contains the frames $\{R, R_{cos}\}$.

– Using this proposed approach, to encode the current frame C, at QP value 23 it costs 1862960 bits with prediction PSNR of 43.36 dB compared to the anchor codec requirements of 1919168 bits and PSNR of 43.19 dB.

– This process is replicated for all subsequent P-frames.

Experimental Analysis

– Low delay-P GOP structure was employed as per the common test conditions.

– Four different QP values were employed: $\{22, 27, 32, 37\}$.

Sequence	Delta rate	Delta PSNR
ParkRunning3	-5.48%	$+0.24$ dB
FoodMarket4	-7.90%	$+0.14$ dB
Tango2	-6.40%	$+0.07$ dB
DaylightRoad2	-3.04%	$+0.04$ dB
Kimono1@1080p	-5.45%	$+0.20$ dB
ParkScene@1080p	-2.79%	$+0.09$ dB

Table 2: The Bjøntegaard delta gains obtained for the test sequences over HEVC when the reference R_{cos} is employed.

Experimental Analysis

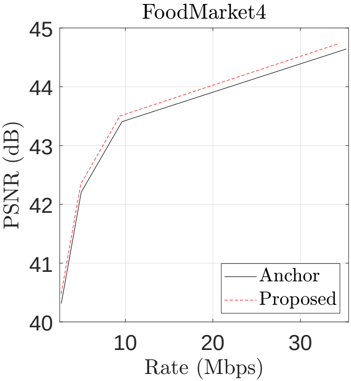

Figure 8: Rate-distortion (RD) curve for the 4K *FoodMarket4* sequence. Delta rate $= -7.90\%$.

Conclusions

- An approach is presented that attempts to model the edge aware motion in lower resolution version of the current frame.
- Can generate predictions of superior PSNR and lower complexity than the baseline approach.
- However, increased codec computational complexity due to an additional reference frame and its generation process.