

Hybrid Cubemap Projection Format for 360-degree Video Coding

Fanyi Duanmu⁺, Yuwen He^{*}, Xiaoyu Xiu^{*}, Philippe Hanhart^{*}, Yan Ye^{*}, Yao Wang⁺

^{*}InterDigital Communications
9710 Scranton Rd, # 250

San Diego, CA92121, USA

{yuwen.he, xiaoyu.xiu, philippe.hanhart,
yan.ye}@interdigital.com

⁺New York University

2 Metro Tech Center

Brooklyn, NY 11201, USA

{fanyi.duanmu,yaowang}@nyu.edu

360-degree video has become popular in recent years with the advances in virtual reality (VR) and augmented reality (AR) technologies and has been rapidly commercialized. To provide viewers with an immersive experience, 360-degree video requires higher resolution and much higher bandwidth compared with conventional 2D video.

In a typical 360-degree video compression and delivery framework, the stitched input 360-degree videos, represented in a native projection format, e.g., equirectangular (ERP), are converted into another projection format, e.g., cubemap (CMP), octahedron (OHP), etc. and frame packed before being fed into existing video codecs. The intermediate projection format is important and would potentially improve the representation efficiency and coding performance. Among all the projection solutions, CMP is very popular and has been widely used in the computer graphics community. The intrinsic rectilinear properties of the CMP format are advantageous for the translational motion model in the modern codec architecture. However, in the CMP representation, the samples on the sphere are not evenly distributed within the faces, resulting in a higher density near the face boundaries and a lower density near the face center. Such non-uniform sampling scheme penalizes the video representation efficiency and degrades the coding performance. Adjusted cubemap projection (ACP) was proposed to address such non-uniform sampling by introducing transform functions to improve the sampling uniformity. However, the transform function parameters in ACP are fixed regardless of the content inside each cube face.

In this paper, a generalized hybrid cubemap projection (HCP) is proposed to improve the 360-degree video coding efficiency beyond ACP. HCP is defined by a pair of forward transform and inverse transform functions with a pair of horizontal and vertical transform parameters per cube face. The encoder can choose the optimal sampling for each face by adjusting the parameters in the horizontal and vertical directions based on the 360-degree video content characteristics inside each cube face. In order to maintain the boundary continuities between two neighboring faces, in a 3x2 packing layout, vertical parameter constraints are imposed such that faces in each face-row have the same vertical parameters. The HCP parameters are chosen to minimize the end-to-end weighted conversion error and determined using iterative search between the horizontal and the vertical directions. Significant changes in HCP parameter values can cause drastic change in sampling distribution, and may affect the inter-picture coding efficiency. Therefore, an efficient HCP parameter estimation algorithm is proposed to achieve a better trade-off between the temporal sampling adaptation and the inter-picture prediction efficiency by reducing the temporal variation of HCP parameters. The proposed HCP parameter search algorithm reduces the computational complexity by 5x compared to the exhaustive search method. The HCP parameters are selected by the encoder using the first picture of each Intra Random-Access Point (IRAP) and signalled once per IRAP. In SPS, projection format, frame packing parameters including number of faces in horizontal and vertical directions and each face's position and orientation are signalled. In PPS, the horizontal and vertical HCP parameters in 6-bit precision are encapsulated.

The proposed HCP solution is implemented upon JEM-6.0 and 360Lib-3.0 software. Simulation results are reported using the test conditions specified in the JVET Call-for-Evidence (CfE) document. Compared with the CMP and ACP formats, the proposed HCP format demonstrates average 3.0 dB (up to 3.6 dB) and 0.2 dB (up to 0.4 dB) End-to-End WS-PSNR improvement for the luma (Y) component, respectively, and average luma (Y) BD-rate reductions of 11.5% (up to 23.0%) and 0.5% (up to 1.0%), respectively.

Keywords-video coding, 360-degree video, projection, cubemap, HEVC.