Low-Latency Neural Stereo Streaming

Qiqi Hou, Farzad Farhadzadeh, Amir Said, Guillaume Sautiere, and Hoang Le†

Qualcomm AI Research*

* Qualcomm AI Research is an initiative of Qualcomm Technologies, Inc.

† Corresponding author

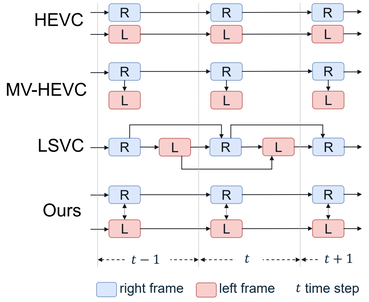
CVPR SEATTLE, WA JUNE 17-21, 2024

Overview

For both Automotive Vehicle (AV) and Virtual Reality (VR) applications, it is critical that the codec efficiently encodes stereo video while minimizing latency.

VR Headsets

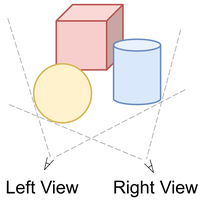
Autonomous Vehicles

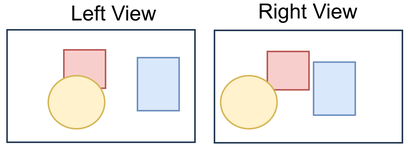


Contributions:

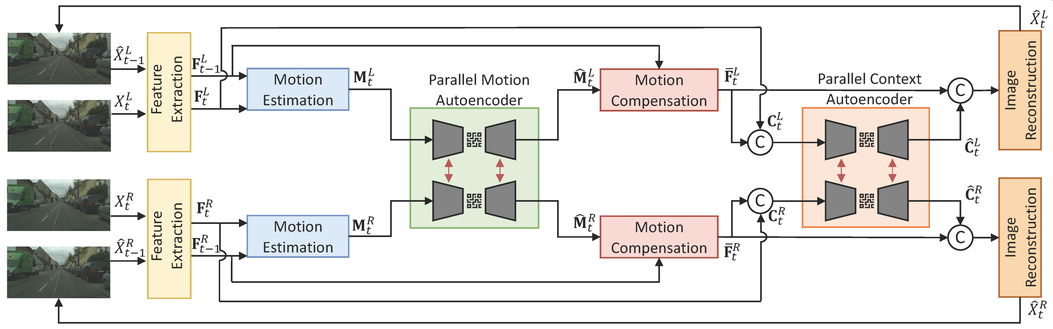
1) A novel low-latency neural stereo video codec architecture that replaces **sequential** inter-view compensation with an efficient and **parallelizable** learned module to connect parallel autoencoders
2) A **bidirectional shift module** that effectively captures and exhibits redundancy between inter-view features
3) Our method improves **50.6%** BD-rate savings compared to MV-HEVC on the cityscape dataset, and **2.8×** less complexity compared to LSVC.

Motivation

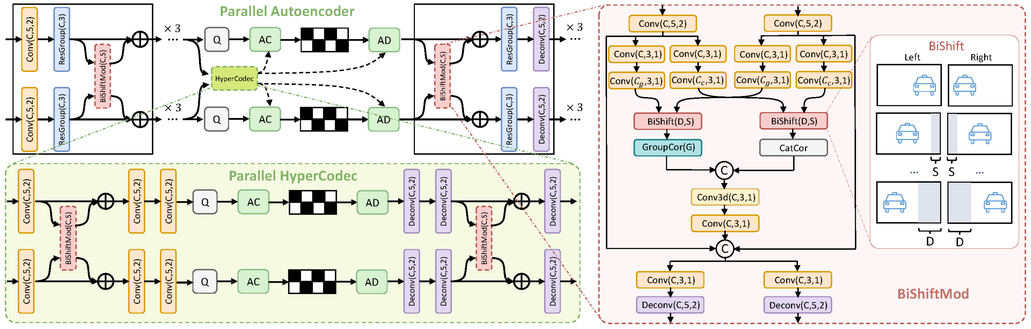




Disparity compensation can be simplified using horizontal shifts



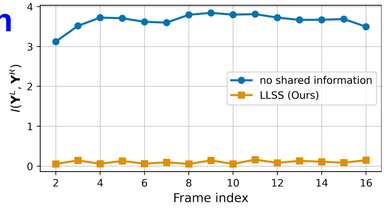
Our network incorporates a parallel motion autoencoder and a parallel context autoencoder to reduce the redundant motion and context information across views, respectively.

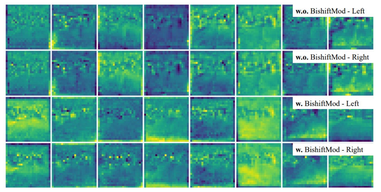


The Bidirectional Shift Module (BiShiftMod) shifts the left and right features bidirectionally.

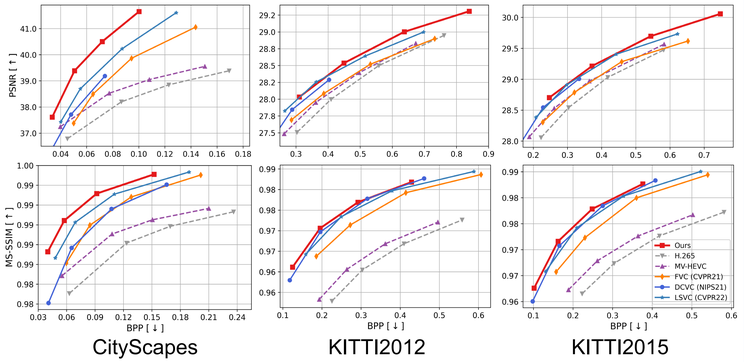
Mutual information

Mutual information between cross-view latent. With BishiftMod, the latent features between left and right views become less alike.



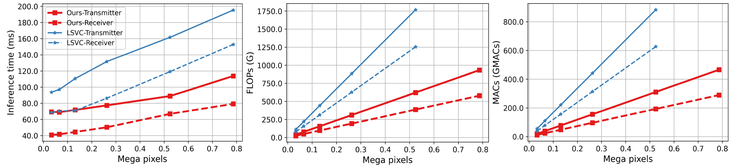


Experiments



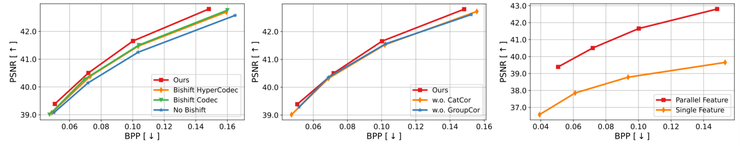
CityScapes KITTI2012 KITTI2015

Rate-distortion curves. On the CityScapes dataset, our method achieves **50.6%** BD-rate savings compared to MV-HEVC. On the KITTI 2012 and 2015 datasets, our method attains **18.2%** and **15.8%** BD-rate savings, respectively.



Inference time FLOPs MACs

Computational complexity. Compared to LSVC, our method is 1.9×, 3.2×, 3.3×, faster in terms of the inference time, FLOPs, and MACs, respectively.



Ablation study. The effectiveness of the BiShiftMod, the components in BishiftMod, and the parallel feature