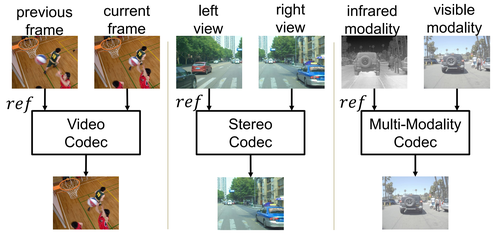
Learning based Multi-modality Image and Video Compression

Guo Lu[1] , Tianxiong Zhong[1] , Jing Geng[1] , Qiang Hu[2] , and Dong Xu[3]
[1]Beijing Institute of Technology [2]ShanghaiTech University [3]University of Sydney

CVPR JUNE 19-24 2022 NEW ORLEANS • LOUISIANA

Motivation

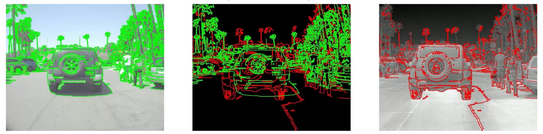
The existing data compression approaches usually adopt individual codecs for each modality without considering the correlation between different modalitie.



Challenge

- Multi-modality image and video frames are **not spatially aligned**.
- Images and videos from different modalities may have **different intensity distribution**.

spatially misaligned

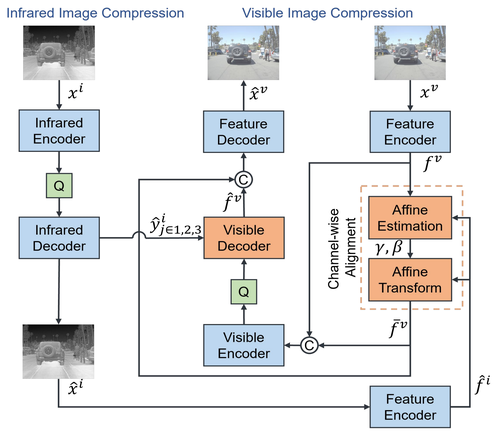


different intensity distribution

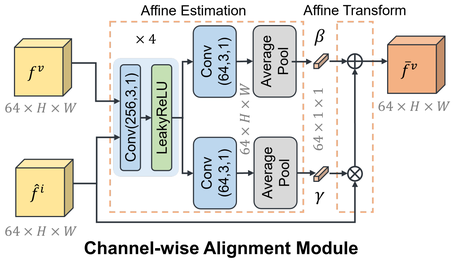
Contribution

- The **first end-to-end optimized framework to compress different modalities**, *e.g.* visible-infrared image pairs, by exploiting the cross-modality redundancy.
- Our framework introduces the **channel-wise and spatialwise alignment modules** to effectively exploit the correlations between different modalities in the feature space.
- The proposed framework is very flexible and can be **extended for multi-modality video compression**.

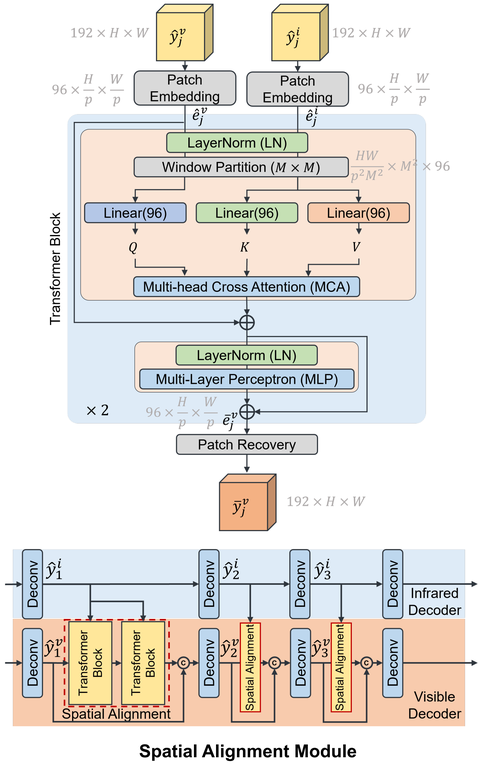
Method



Multi-modality Image Compression Framework



Channel-wise Alignment Module

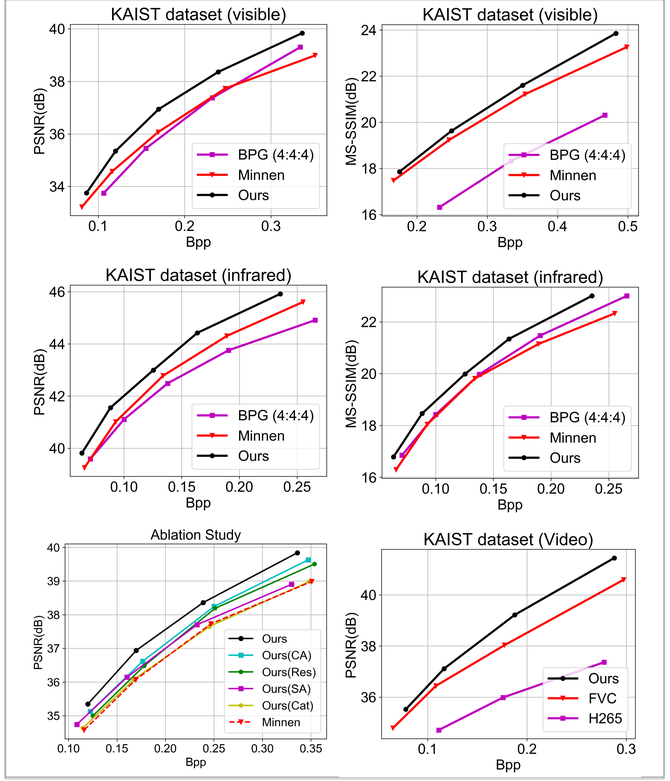


Spatial Alignment Module

Visualization Results

FLIR_09070.jpg (Bpp/MS-SSIM) BPG 4:4:4 (0.1070Bpp/0.9832) Minnen (0.1072Bpp/0.9863) **Ours (0.0985Bpp/0.9880)**

Results



Quantitative Results

BDBR(%) results compared with BPG

Methods	FLIR		KAIST	
	visible	infrared	visible	infrared
Minnen	-22.342	-14.960	-3.624	-8.751
Ours	**-30.226**	**-21.621**	**-18.639**	**-21.289**