DIBR for Immersive Videos: Traditional Depth Estimators versus Learning-Based Depth Estimators

Smitha Lingadahalli Ravi

Orange Labs | INSA Rennes, France

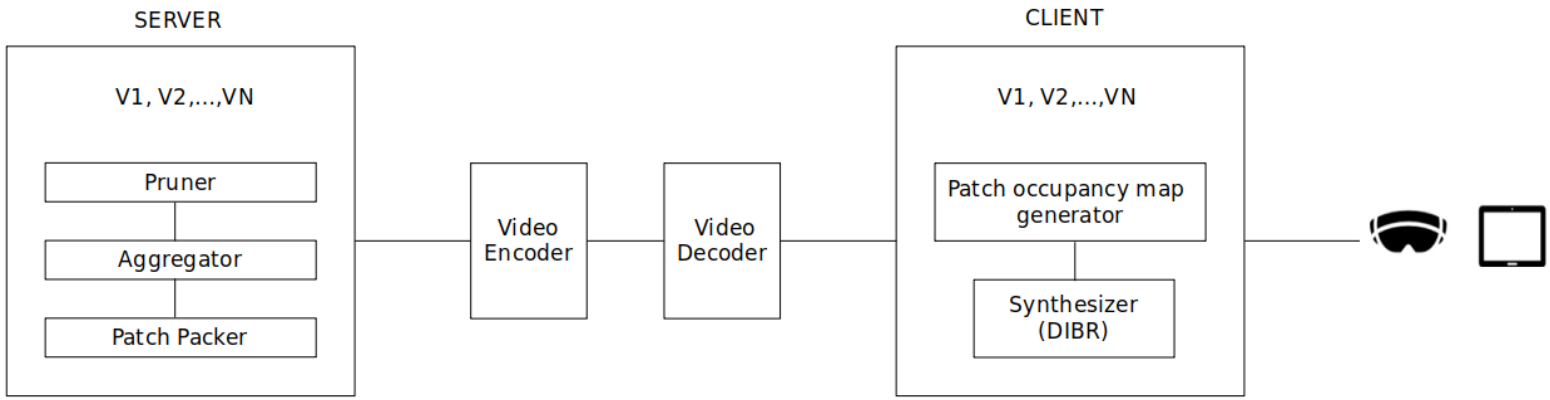
July 4th, 2022

Objective

- The moving picture experts group (MPEG) has been actively developing the MPEG Immersive Video (MIV) standard to efficiently transmit 6 degrees of freedom for realistic and virtual environments.
- The MIV is a concrete step towards a complete chain for immersive video coding, delivery, and rendering.

Objective

- Only studied for their depth accuracy.
- Assumption: Learning-based depth estimation would soon outperform the traditional depth estimators that are currently used.



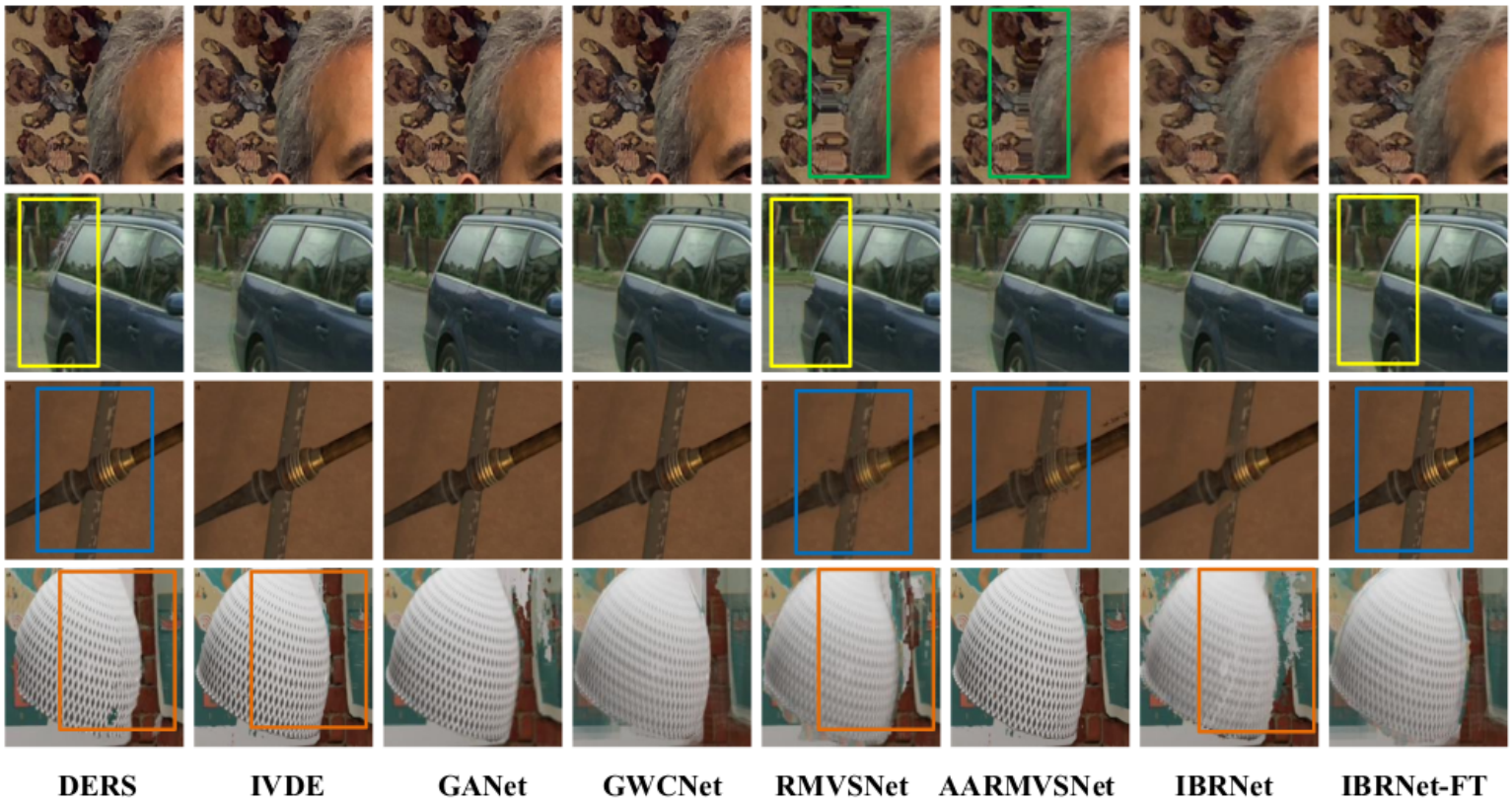
Settings

- **Traditional Methods**: Depth Estimation Reference Software (DERS), Immersive Video Depth Estimation Software (IVDE).
- **Learning-Based Methods**: GA-Net (Stereo), GWC-Net (Stereo) and R-MVSNet (Multi-View), AA-RMVSNet (Multi-View), IBRNet (Multi-View).
- **Synthesis**: Test Model for Immersive Video (TMIV)
- **Dataset**: MPEG-I Sequences.

Results

	IV-PSNR ↑								
Sequences	Original	DERS	IVDE	GANet	GWCNet	RMVSNet	AARMVSNet	IBRNet	IBRNet(FT)
Frog	36.92	34.96	35.13	**35.87**	35.29	30.28	34.48	31.52	32.18
Kitchen	40.81	38.46	38.92	39.27	38.63	35.74	37.61	38.64	**39.42**
Shaman	45.74	41.18	42.06	41.38	41.55	36.79	38.25	40.34	**42.28**
Painter	45.35	39.37	**39.82**	38.91	39.67	37.02	39.29	38.14	38.47
Street	41.59	39.61	38.81	**40.56**	40.08	35.09	39.92	38.94	37.96
Carpark	42.17	39.45	38.29	**40.39**	39.78	34.78	37.16	37.68	37.41
Fan	37.88	35.13	**36.71**	34.84	34.29	33.68	34.48	34.92	35.21
Mirror	41.52	38.56	**40.46**	35.67	35.12	33.46	34.18	36.92	38.18
Average	41.37	38.34	**38.77**	38.36	38.05	34.66	36.92	37.13	37.65
	LPIPS ↓								
Sequences	Original	DERS	IVDE	GANet	GWCNet	RMVSNet	AARMVSNet	IBRNet	IBRNet(FT)
Frog	0.118	0.137	0.149	**0.134**	0.141	0.212	0.178	0.198	0.193
Kitchen	0.112	0.131	**0.126**	0.129	0.129	0.151	0.140	0.129	0.132
Shaman	0.087	0.102	0.112	0.109	0.106	0.184	0.168	0.119	**0.101**
Painter	0.091	0.121	**0.114**	0.132	0.118	0.149	0.125	0.141	0.138
Street	0.109	0.120	0.126	0.125	0.119	0.165	**0.112**	0.153	0.151
Carpark	0.104	0.124	0.129	**0.118**	0.128	0.193	0.152	0.148	0.145
Fan	0.105	0.130	**0.121**	0.135	0.138	0.152	0.141	0.138	0.132
Mirror	0.101	0.118	**0.109**	0.117	0.130	0.148	0.131	0.128	0.121
Average	0.103	0.124	**0.123**	0.125	0.126	0.169	0.143	0.144	0.139
	SSIM ↑								
Sequences	Original	DERS	IVDE	GANet	GWCNet	RMVSNet	AARMVSNet	IBRNet	IBRNet(FT)
Frog	0.864	0.852	0.849	**0.856**	0.852	0.784	0.817	0.795	0.798
Kitchen	0.912	0.881	0.885	0.892	0.882	0.861	0.872	0.879	**0.902**
Shaman	0.926	0.916	0.901	0.906	0.912	0.882	0.892	0.915	**0.919**
Painter	0.920	0.884	**0.896**	0.882	0.887	0.859	0.879	0.872	0.875
Street	0.915	0.887	0.880	**0.894**	0.889	0.865	0.881	0.871	0.874
Carpark	0.919	0.879	0.871	**0.883**	0.880	0.856	0.861	0.868	0.869
Fan	0.904	0.875	**0.889**	0.872	0.868	0.852	0.856	0.864	0.870
Mirror	0.924	0.889	**0.908**	0.892	0.889	0.875	0.882	0.885	0.891
Average	0.910	0.882	**0.884**	**0.884**	0.881	0.854	0.867	0.868	0.874

Results



References

[1] [N18450] E. Juarez, Manual of Depth Estimation Reference Software (DERS 8.0), ISO/IEC JTC1/SC29/WG11 MPEG2019/N18450, March 2019, Geneva, Switzerland.

[2] [N19224] Manual of Immersive Video Depth Estimation, ISO/IEC JTC1/SC29/WG11 MPEG2020/N19224, Online, April 2020.

[3] Feihu Zhang, Victor Prisacariu, Ruigang Yang, and Philip HS Torr. GA-Net: Guided aggregation net for end-to-end stereo matching. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[4] Xiaoyang Guo, Kai Yang, Wukui Yang, Xiaogang Wang, and Hongsheng Li. Group-wise correlation stereo network. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019

[5] Yao, Yao, et al. MVSNet: Depth inference for unstructured multi-view stereo. Proceedings of the European Conference on Computer Vision (ECCV), 2018.

[6] Wei, Zizhuang, et al. AA-RMVSNet: Adaptive aggregation recurrent multi-view stereo network. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021.

[7] Test Model 4 for Immersive Video, ISO/IEC JTC1/SC29/WG11MPEG/N18795, October 2019, Geneva, Switzerland.

[8] Wang, Qianqian, et al. IBRNet: Learning multi-view image-based rendering. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021.

[9] Overview and Efficiency of Decoder-Side Depth Estimation in MPEG Immersive Video / Dawid Mieloch, Patrick Garus, Marta Milovanovic, Joel Jung, Jun Young Jeong, Smitha Lingadahalli Ravi, Basel Salahieh // IEEE Transactions on Circuits and Systems for Video Technology - 2022, p. 1-15

Thanks!