Scene reconstruction and depth map / wireframe from arbitrary stereo/pan-video

Computer Vision (CS6350)
TPA - 5

1 Problem Statement

The purpose of this project is to develop algorithms capable of three-dimensional Scene Reconstruction from a pan-video or Stereo Images. The basic steps in the reconstruction process are : predicting the depth map (disparity map), estimating depth of (visually) salient landmarks, tessellation to create a wireframe representation and finally rendering (preferably use OpenGL) with pseudo-color or pixels from an image. Depending on the model used, alternative methods can be adopted.

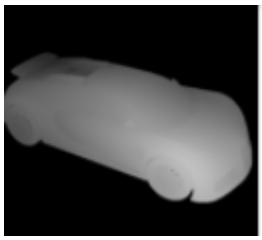
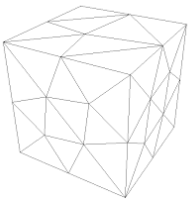
2 Input

- **A pair of stereo Images / Sequences of Images from Pan Video**



3 Output

- **Depth Map, Wireframe, Rendered 3D scene with novel views**


Depth -Map

Wireframe

Rendered 3-d object

4 Datasets

- KITTI stereo Dataset

  link -http://www.cvlibs.net/datasets/kitti/eval_object.php?obj_benchmark=3d

- Pan-videos downloaded from the web
  For e.g. https://www.youtube.com/watch?v=eBL6vu9NQtw

- Any other standard stereo dataset.

Caution/Warning: Reconstruction from pan-video may be considered more challenging than from arbitrary (not perfect) stereo; the former may get you more marks**.**

References

[1] C. B. Choy, D. Xu, J. Gwak, K. Chen, and S. Savarese, "3d-r2n2: A unified approach for single and multi-view 3d object reconstruction," in *European conference on computer vision*. Springer, 2016, pp. 628–644.

[2] U. Gu¨du¨kbay and F. Durupınar, "Three-dimensional scene representations: Modeling, animation, and rendering techniques," in *Three-Dimensional Television*. Springer, 2008, pp. 165–200.

[3] X. Han, H. Laga, and M. Bennamoun, "Image-based 3d object reconstruction: State-of-theart and trends in the deep learning era," *IEEE transactions on pattern analysis and machine intelligence*, 2019.

[4] C. Russell, R. Yu, and L. Agapito, "Video pop-up: Monocular 3d reconstruction of dynamic scenes," in *European conference on computer vision*. Springer, 2014, pp. 583–598.

[5] D. Shin, Z. Ren, E. B. Sudderth, and C. C. Fowlkes, "3d scene reconstruction with multilayer depth and epipolar transformers," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 2172–2182.

[6] C. Wang, S. Lucey, F. Perazzi, and O. Wang, "Web stereo video supervision for depth prediction from dynamic scenes," in *2019 International Conference on 3D Vision (3DV)*. IEEE, 2019, pp. 348–357.

[7] N. Xue, T. Wu, S. Bai, F. Wang, G.-S. Xia, L. Zhang, and P. H. Torr, "Holistically-attracted wireframe parsing," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 2788–2797.

[8] Y. Yao, N. Schertler, E. Rosales, H. Rhodin, L. Sigal, and A. Sheffer, "Front2back: Single view 3d shape reconstruction via front to back prediction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 531–540.