

# **3D Topologically-Aware Semantic Scene reconstruction and depth map / wireframe / Point-Cloud from single RGB panorama scene (or Two views)**

Computer Vision (CS6350)

TPA - 11

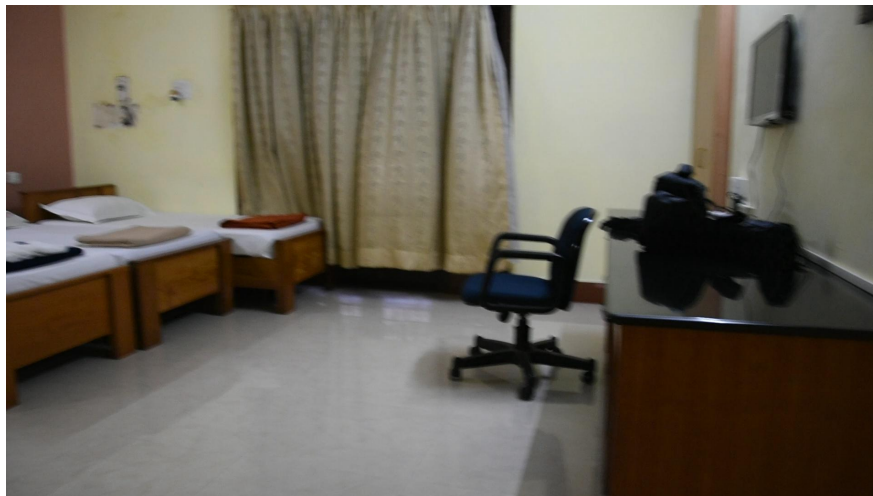
## **1. Problem Statement**

The purpose of this project is to develop algorithms capable of three-dimensional topologically-aware semantic Scene Reconstruction from a single panorama scene (constructed from a set of frames) and its corresponding semantic segmentation mask. The basic steps of a typical 3D reconstruction process using depth map 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) preserving structures of the objects present in the image (topologically-aware) [11] with semantic colors of the various classes in the segmentation mask of the panorama. Depending on the model used, alternative methods can be adopted.

## **2. Input**

- A panorama scene (or a pair of stereo frames)
- Segmentation mask of the panorama (or separately for both frames)

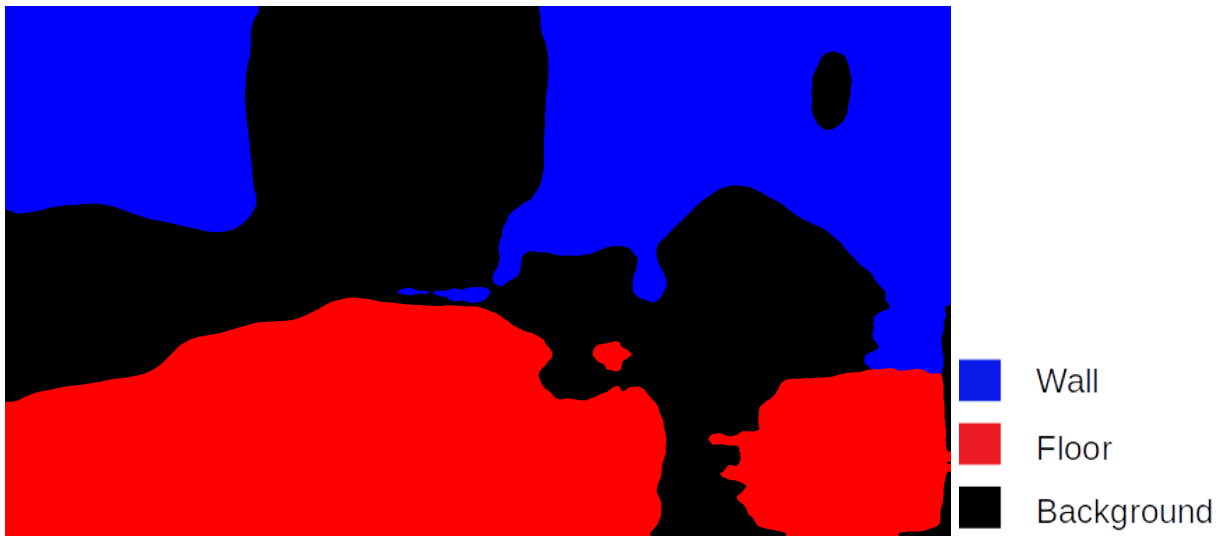
See examples in the following:



**Figure 1 :Input scene (not Panorama) of a room**



**Figure 2: Input Scene (Panorama) of outdoor (OAT,IIT M)**



**Figure 3 : Segmentation mask of figure 1**



**Figure 4: input[12]**

### 3. Datasets

- KITTI Dataset

link

-[http://www.cvlibs.net/datasets/kitti/eval\\_object.php?obj\\_benchmark=3d](http://www.cvlibs.net/datasets/kitti/eval_object.php?obj_benchmark=3d)

- NYU Depth v2 Dataset link - [https://cs.nyu.edu/~silberman/datasets/nyu\\_depth\\_v2.html](https://cs.nyu.edu/~silberman/datasets/nyu_depth_v2.html)
- SapeNet Dataset link - <https://www.shapenet.org/>  
**Caution/Warning:** Semantic scene Reconstruction from pan-video may be considered more challenging than from arbitrary (not perfect) stereo ; the former may get you more marks.
- PIX 3D [13]: <http://pix3d.csail.mit.edu/>
- CUB-200-2011 [14]: [http://www.vision.caltech.edu/datasets/cub\\_200\\_2011/](http://www.vision.caltech.edu/datasets/cub_200_2011/)

## 4. Output

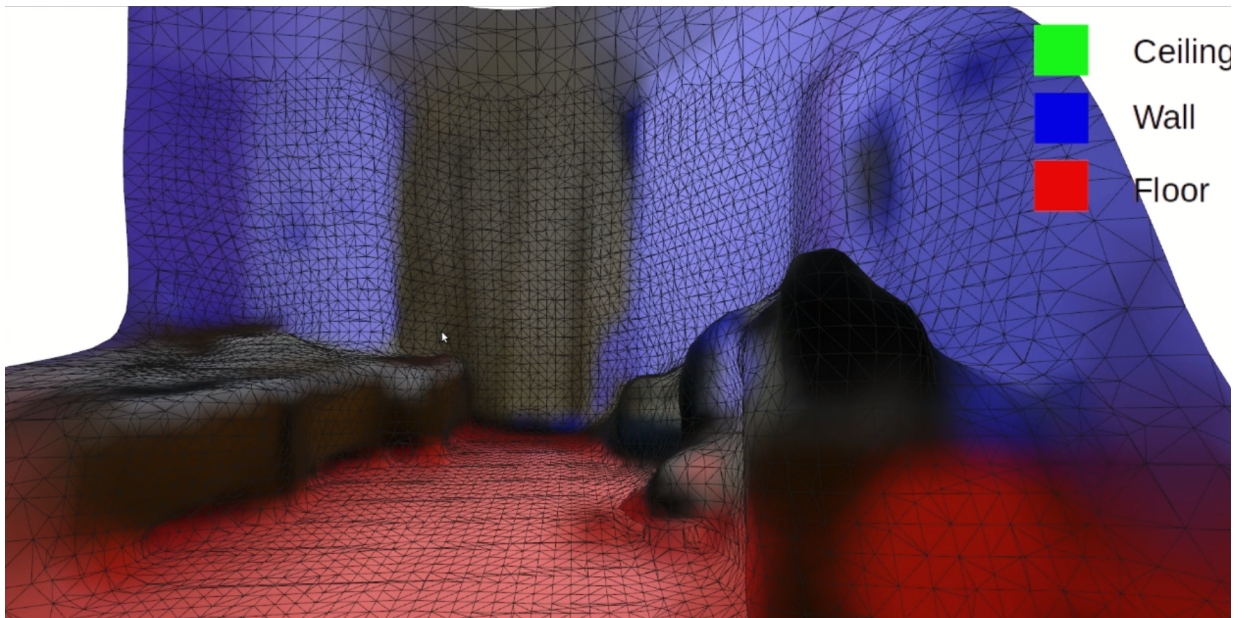
- Depth Map , Wireframe with semantic segmentation colors , Rendered 3D semantic scene with novel views



**Figure 5: Depth map of figure 1**



**Figure 6: 3D Scene reconstruction of figure 1**



**Figure 7: Semantic 3D wireframe of figure 1**



**Figure 8: 3D Semantic rendered Scene of figure 1**



**Figure 9: 3D reconstruction of figure 4**

## 5. 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. Guřduřkbay and F. Durupınar, “Three-dimensional scene representations: Modeling, animation, and rendering techniques,” in *Three-Dimensional Television*. Springer, 2008, pp. 165– 200.
- [3] Zeng, W., Karaoglu, S., & Gevers, T. (2020). Pano2Scene: 3D Indoor Semantic Scene Reconstruction from a Single Indoor Panorama Image. In *BMVC*.



- [4] X. Han, H. Laga, and M. Bennamoun, "Image-based 3D object reconstruction: State-of-the-art and trends in the deep learning era," *IEEE transactions on pattern analysis and machine intelligence*, 2019.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] Chuhan Zou, Alex Colburn, Qi Shan, and Derek Hoiem. Layoutnet: Reconstructing the 3d room layout from a single rgb image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2051–2059, 2018.
- [11] Ummenhofer, B., Zhou, H., Uhrig, J., Mayer, N., Ilg, E., Dosovitskiy, A., & Brox, T. (2017). Demon: Depth and motion network for learning monocular stereo. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 5038-5047). Code at: <https://github.com/lmb-freiburg/demon>
- [12] Duggal, Shivam, and Deepak Pathak. "Topologically-Aware Deformation Fields for Single-View 3D Reconstruction." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022.
- [13] Xingyuan Sun, Jiajun Wu, Xiuming Zhang, Zhoutong Zhang, Chengkai Zhang, Tianfan Xue, Joshua B Tenenbaum, and William T Freeman. Pix3d: Dataset and methods for single image 3D shape modeling. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.
- [14] P. Welinder, S. Branson, T. Mita, C. Wah, F. Schroff, S. Belongie, and P. Perona. Caltech-UCSD Birds 200. Technical Report CNS-TR-2010-001, California Institute of Technology, 2010