UE17CS490A – Capstone Project Phase – 1

Project Progress Review #2
(Project Requirements Specification and Literature Survey)

Project Title	:	Extracting and Rendering 3D Structure and Orientation of Objects From 2D Images
Project ID	:	PW21KS03 (3_129_276_1525)
Project Guide	:	Prof. K. S. Srinivas
Project Team	:	Ashwin R Bharadwaj PES1201700003 Hardik Gourisaria PES1201700129 Hrishikesh V PES1201700276 K. Shrinidhi Bhagavath PES1201701525

Abstract and Scope

Rendering 3D structure of real world objects **using only 2D images** of the object **without the aid of any special sensors** such as the Kinect sensor.

Applications include:

- 3D printing real world objects based on images
- Landscape simulations for AR and VR.
- Improve human interaction of automation robots by helping them navigate and interact better with their surroundings

Abstract and Scope

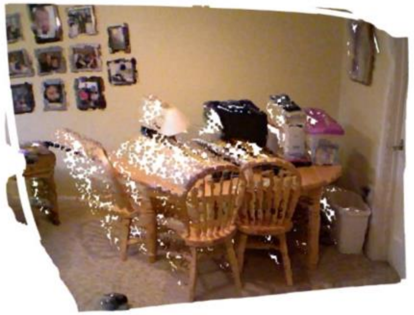
Abstract:

Extraction of spatial orientation and geometric structure of objects and landscapes from 2D images to render the structure in 3D space to aid in various application in the field of Virtual and Augmented Reality

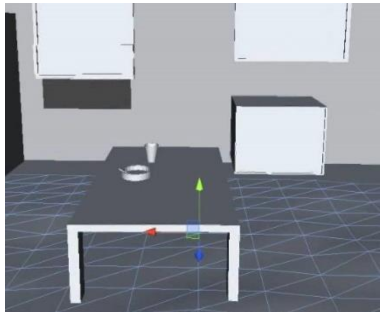
Abstract and Scope

Input Image(s)

Depth Map(s)



Generate Point Cloud

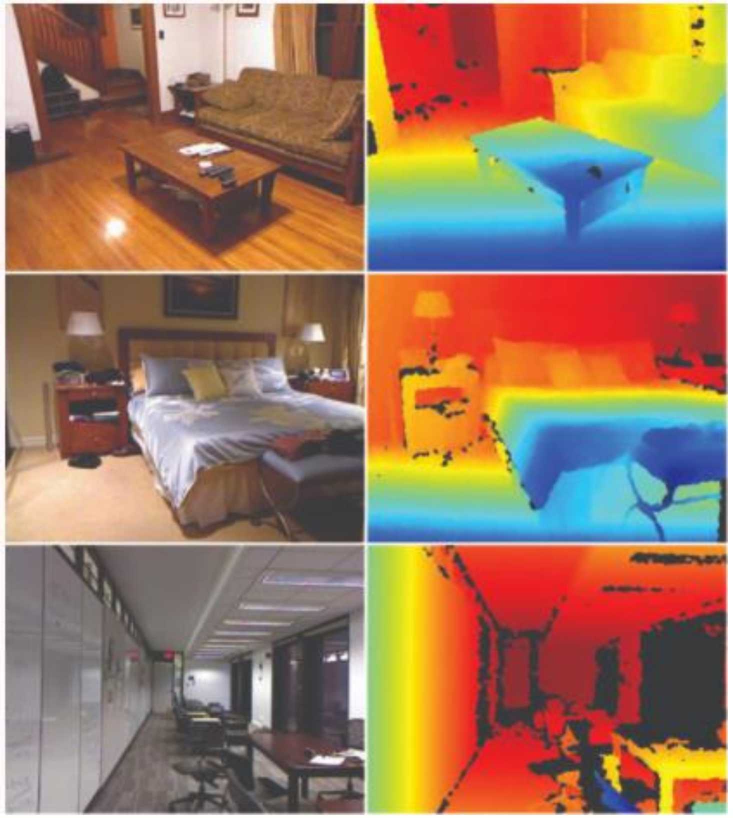


Generate Solid 3D Objects/Landscapes (Optimistic Goal)

Suggestions from Review - 1

Datasets were recommended in the Review 1

- NYU Depth dataset v2
- DIML RGB+D dataset



User Classes and Characteristics

- Architects may use the product to model a room in 3D space using only 2D images and make changes to the 3D model as desired

- 3D print real world objects based on images

- Landscape simulations for AR and VR for devising military simulations and strategy

- Improve human interaction of automation robots by helping them navigate and interact better with their surroundings

PES UNIVERSITY

Constraints / Dependencies / Assumptions

Hardware Dependency:

- Machines with good Graphics Processing Unit (GPU)
- Good amount of RAM
- Rent cloud GPUs for the above hardware requirements

Assumption:

We assume that one of the Deep Learning methods being explored will provide us good results.

Functional Requirements

- Takes 2D RGB images as inputs
- Convert 2D RGB image to 2D Grayscale image
- Generates the Depth Map for the 2D image input
- Uses the generated Depth Map for the 2D images to generate a 3D point cloud

Non - Functional Requirements

- Model developed should be light weight and have low latency
- Model should not have any hardware dependency
- Model developed should work across multiple platforms and should be portable

Literature Survey

1. **Zhao, Chaoqiang & Sun, Qiyu & Zhang, Chongzhen & Tang, Yang & Qian, Feng. (2020). Monocular Depth Estimation Based On Deep Learning: An Overview.**

   <u>Main Idea:</u>
   Analyses and compares various Monocular Depth Estimation Techniques
   <u>Pros:</u>
   Introduction and summary to multiple techniques that can be explored

Literature Survey

2. **Wofk, Diana and Ma, Fangchang and Yang, Tien-Ju and Karaman, Sertac and Sze, Vivienne, "FastDepth: Fast Monocular Depth Estimation on Embedded Systems," in IEEE International Conference on Robotics and Automation (ICRA), 2019**

   Main Idea:
   Using Mobile Net as an encoder to build a lightweight efficient model to obtain depth map for 2D images.
   Pros:
   Developed a lightweight model to generate depth map for 2D images
   Cons:
   Accuracy of the model reduces (MSE increases)

Literature Survey

3. **K. G. Lore, K. Reddy, M. Giering, and E. Bernal, “Generative adversarial networks for depth map estimation from RGB video,” pp. 1258–12588,06 2018**

Main Idea:

Using Optical Flow of the images to generate a Depth Map using Conditional GAN.

Pros:

Improved the accuracy on existing GAN based models

Cons:

- Used very limited dataset and there is high chance of overfitting
- No guarantee that the model will generalize well

Literature Survey

4. **J. Facil, B. Ummenhofer, H. Zhou, L. Montesano, T. Brox, and J. Civera, “Cam-convs: Camera-aware multi-scale convolutions for single-view depth,” 04 201**

   Main Idea:
   Using Camera Parameters to improve the accuracy of the Depth Map generated by the Deep Learning Model

   Pros:
   Model generalizes well across images captures using different cameras.

Literature Survey

5. S. Sabour, N. Frosst, and G. E. Hinton, “Dynamic routing between capsules,” 2017.

6. X. Luo, J.-B. Huang, R. Szeliski, K. Matzen, and J. Kopf, “Consistent video depth estimation,” 2020.

7. V. Harman, J. Flack, S. Fox, and M. Dowley, “Rapid 2d-to-3d conversion,” in Stereoscopic displays and virtual reality systems IX, vol. 4660, pp. 78-86, International Society for Optics and Photonics, 2002.

Literature Survey

8. Y. Zhao, T. Birdal, H. Deng, and F. Tombari, “3d point capsule networks,” in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1009-1018, 2019.

9. L. Zhang, M. Edraki, and G.-J. Qi, “Cappronet: Deep feature learning via orthogonal projections onto capsule subspaces,” in Advances in Neural Information Processing Systems, pp. 5814-5823, 2018.

10. R. Saqur and S. Vivona, “Capsgan: Using dynamic routing for generative adversarial networks,” 2018.

PES UNIVERSITY

Literature Survey

11. X.-F. Han*, H. Laga*, M. B. S. Member, and IEEE, “Image-based 3d object reconstruction: State-of-the-art and trends in the deep learning era,” arXiv preprint arXiv:1906.06543, 2019.

12. H. Xie, H. Yao, X. Sun, S. Zhou, S. Zhang, H. I. of Technology, S. Re-search, and P. C. Laboratory, “Pix2vox: Context-aware 3d reconstruction from single and multi-view images,” arXiv preprint arXiv:1901.11153v2, 2019.

13. Z. Li and N. Snavely. Megadepth: Learning single-view depth prediction from internet photos. In Computer Vision and Pattern Recognition (CVPR), 2018

14. Y. Kuznietsov, J. St ̈uckler, and B. Leibe. Semi-supervised deep learning for monocular depth map prediction. In Proc.of the IEEE Conference on Computer Vision and Pattern Recognition, pages 6647-6655, 201

15. C. Godard, O. Mac Aodha, and G. Brostow. Digging into self-supervised monocular depth estimation. arXiv preprintarXiv:1806.01260, 2018

Progress So Far

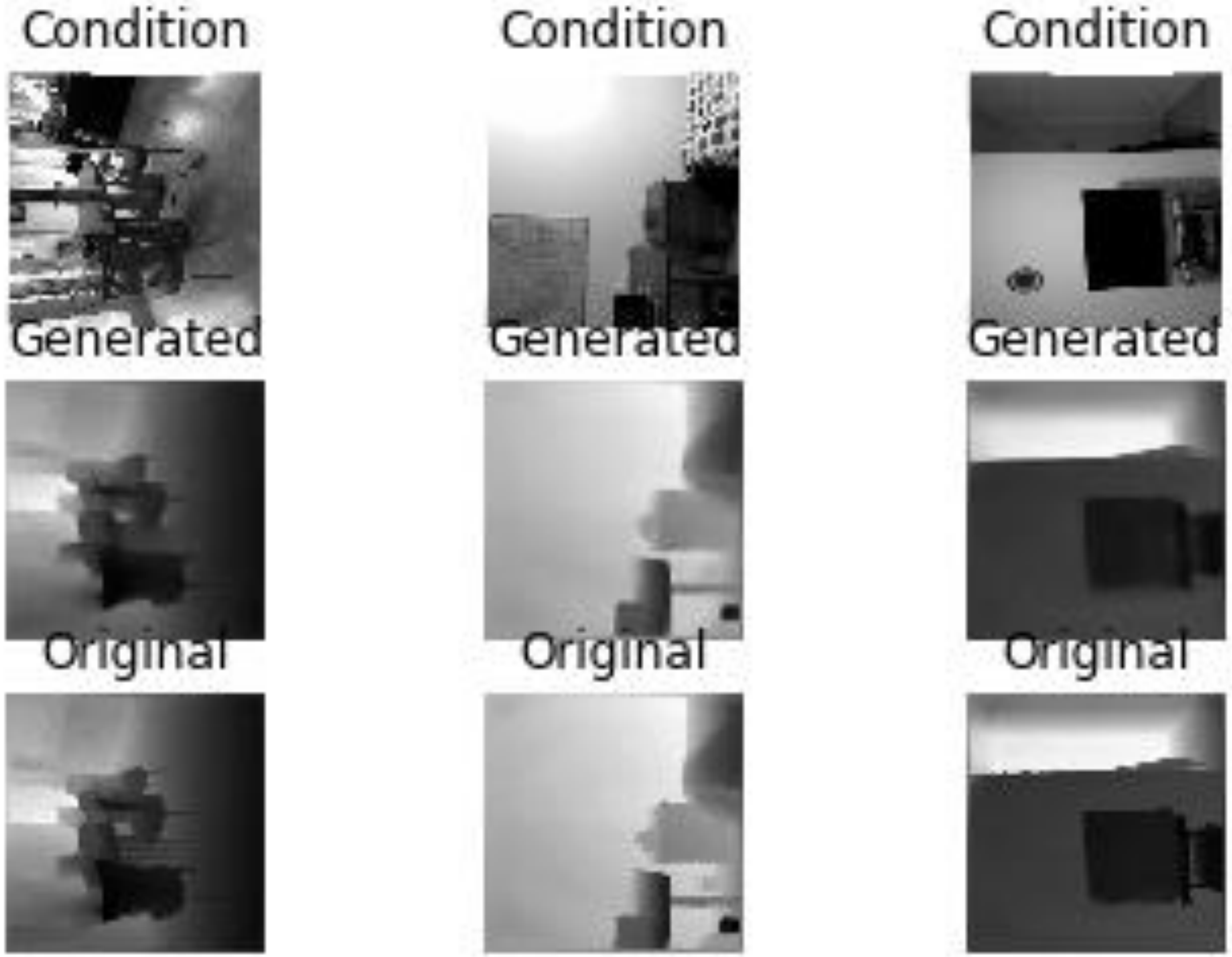
Model Structure:
Auto Encoder-Decoder CNN with Skip Level Connections (U-Net) to preserve shape edges and features

Optimizer:
Adam Optimizer
Learning rate = 0.002   Alpha = 0.5

Loss Function Tries:
MSE, MAE, Berhu Function

Input and Output:
2D single channel grayscale image to Single channel depth map

Progress So Far



Thank You