An Introduce of Building 3D scenes from 2D images using ML algorithms

Chenbo Zhang
Department of ECE
Boston University
zhangcb@bu.edu

Abstract

Using machine learning algorithms to build 3D scenes out of 2D images is widely used in nowadays industries. From auto-drive to robotic, from face recognition to Geological exploration, this technology is widely use in reconstructing 3D information using 2D camera information. The main focus of this algorithm is to reconstruct 3D information from 2D images, that is, to predict the third dimensional information out of two dimensional information. This paper is aimed to introduce this technology, its application, and various ways and tools to achieve it.

1 Introduction

Using machine learning algorithms to build 3D scenes out of 2D images is widely used in nowadays industries. From auto-drive to robotic, from face recognition to Geological exploration, this technology is widely use in reconstructing 3D information using 2D camera information. The main focus of this algorithm is to reconstruct 3D information from 2D images, that is, to predict the third dimensional information out of two dimensional information. This paper is aimed to introduce this technology, its application, and various ways and tools to achieve it.

1.1 Advantages

As a 2D camera is financially more affordable than a dense camera or a radar, if 3D information constructed by 2D camera can be with high confidence and accuracy, the industry can save cost by replacing radars and 3D cameras with 2D cameras.

2 Application

This technology is widely used in various industries. For face recognition, Nguyen, Kim Trong and Zitzmann's paper *Face Spoofing Detection for Smartphones using a 3D Reconstruction and the Motion Sensors*[1] provide a way of face recognition.
For ground vehicle robotics, David Nist´er present the method *Visual Odometry*[2], which can reconstruct 3D scene and guide robot with real time input from simple camera.
For 3D object creation, Zhang present a way of multi-view generation of 3D object in the paper *Using vanishing points for camera calibration and coarse 3D reconstruction from a single image*[3], which is used in NVIDIA Omniverse to fasten 3D model creation.
For auto-drive implementation, Dickmanns provide a way to guide high speed vehicles in his paper *Autonomous High Speed Road Vehicle Guidance by Computer Vision*[4].

Project 1 Report for Course EC601, *Prod Des Ece* in Boston University

3 Approaches

3.1 Inverse Graphics Network

In NeurIPS 2015, Kulkarni present this method in paper *Deep Convolutional Inverse Graphics Network*, provide a method to learn an interpretable representation of images, disentangled with respect to three-dimensional scene structure and viewing transformations such as depth rotations and lighting variations.[5]
In this paper, Kulkarni introduced Deep Convolutional Inverse Graphics Network (DC-IGN), a network with a encoder and a decoder, transferring input observed image X to output $P(X|Z)$, where z_i of Z consists of scene latent variables such as pose, light, texture or shape.
The paper showed that in their test, the encoder process the data x to get latents z_i thus images can be re-rendered to different viewpoints, lighting conditions, shape variations, etc. Thus the 3D scene is predicted using the 2D information.

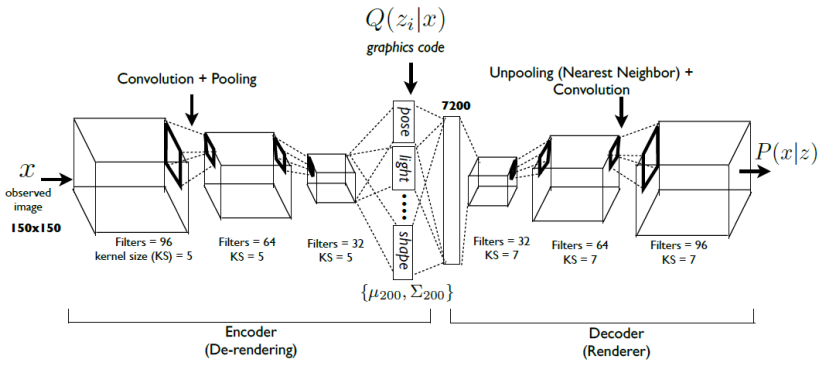

Figure 1: Model Architecture of Kulkarni's approach.[5]

3.2 Vanishing Points

In 2000 Guillou present this interesting method of reconstructing 3D scene in his paper *Using vanishing points for camera calibration and coarse 3D reconstruction from a single image*.[6]
Guillou provided a method recover the geometry and the photometry of objects from a single image. Though is under several restraint, such as it require the single image contains at least two vanishing points, etc, the algorithm still did a significant job rendering 3D scene from single image.
This method determines two vanishing point by extracting the edges of objects in the image. After vanishing points was determined, the mapping between edges in the image and 3D location can be established.

4 Software and Database

4.1 Pytorch3D

Pytorch3D is a Pytorch-based 3D computer vision library for deep learning with 3D data[7]. It is a Facebook's research with a BSD license. It is a powerful tool for machine learning and deep learning on 3D object such as 3D point cloud, 3D mesh, etc.

4.2 NVIDIA Omniverse

NVIDIA Omniverse is a powerful multi-GPU real-time simulation collaboration platform for 3D production pipelines based on Pixar's Universal Scene Description and NVIDIA RTX™ technology.[8]
It is under copyright of NVIDIA CORPORATION. In Omniverse NVIDIA provide a useful tool called GANverse3D which featuring GANs under the hood, that can turns 2D images into 3D objects. In NVIDIA GANverse3D Demo, Omniverese can directly generate 3D car model out of car pictures from different perspectives.

4.3 Point Cloud Library

The Point Cloud Library (PCL) is a standalone, large scale, open project for 2D/3D image and point cloud processing. PCL is released under the terms of the BSD license, and thus free for commercial and research use.[9] This library provide a variety of models sets such as filters, features, key-points, surface, etc.

4.4 Shapnet

ShapeNet is an ongoing effort to establish a richly-annotated, large-scale dataset of 3D shapes. For instance, its Render for CNN can generate millions of training images for high-capacity models such as deep CNNs.[10]

5 Conclusion

As shown above, building 3D scenes from 2D images is a method not only important but also widely used in industries such as face recognition, robotics, 3D model generating, auto-drive, etc. There are variety of ways to realize it, and there are various powerful libraries, platforms, and databases that can help us construct 2D to 3D machine learning algorithms.

References

[1] Kim Trong Nguyen, Cathel Zitzmann, Florent Retraint, Agnes Delahaies, Frédéric Morain-Nicolier, and Hoai Phuong Nguyen. Face spoofing detection for smartphones using a 3d reconstruction and the motion sensors. In *ICISSP*, pages 286–291, 2018.

[2] David Nistér, Oleg Naroditsky, and James Bergen. Visual odometry. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, volume 1, pages I–I. Ieee, 2004.

[3] Yuxuan Zhang, Wenzheng Chen, Huan Ling, Jun Gao, Yinan Zhang, Antonio Torralba, and Sanja Fidler. Image gans meet differentiable rendering for inverse graphics and interpretable 3d neural rendering. *arXiv preprint arXiv:2010.09125*, 2020.

[4] Ernst D Dickmanns and Alfred Zapp. Autonomous high speed road vehicle guidance by computer vision. *IFAC Proceedings Volumes*, 20(5):221–226, 1987.

[5] Tejas D Kulkarni, Will Whitney, Pushmeet Kohli, and Joshua B Tenenbaum. Deep convolutional inverse graphics network. *arXiv preprint arXiv:1503.03167*, 2015.

[6] Erwan Guillou, Daniel Meneveaux, Eric Maisel, and Kadi Bouatouch. Using vanishing points for camera calibration and coarse 3d reconstruction from a single image. *The Visual Computer*, 16(7):396–410, 2000.

[7] Nikhila Ravi, Jeremy Reizenstein, David Novotny, Taylor Gordon, Wan-Yen Lo, Justin Johnson, and Georgia Gkioxari. Accelerating 3d deep learning with pytorch3d. *arXiv:2007.08501*, 2020.

[8] NVIDIA. Nvidia omniverse™ platform, Aug 2021.

[9] Radu Bogdan Rusu and Steve Cousins. 3D is here: Point Cloud Library (PCL). In *IEEE International Conference on Robotics and Automation (ICRA)*, Shanghai, China, May 9-13 2011. IEEE.

[10] Hao Su, Charles R. Qi, Yangyan Li, and Leonidas J. Guibas. Render for cnn: Viewpoint estimation in images using cnns trained with rendered 3d model views. In *The IEEE International Conference on Computer Vision (ICCV)*, December 2015.