Building 3d scenes from 2d images using ml algorithms

Junyi TANG U45214536

***Problem Statement*—** 3D modeling is coming to be a heat filed which is widely used in various areas such as architecture modeling，game developing, medical equipment producing and so on. Even if so much attention is drawn on this 3D modeling filed, the nowadays 3D modeling methodology is still mainly relay on manual work. And the 3D modeling work to so time-consuming that an engineer may spend tens of days to build up a tiny 3D model from the beginning. And several days or hours may be still needed to check or improve that model design. Under this kind of circumstance, the method of automatically building up 3D models based on a simple 2D image becoming a important and valued topic to study. The process of reconstructing the 3D scenes from a simple 2D images can be simplified as two steps including pattern and object recognition, and the 3D object generation. Machine learning algorithms are also introduced into this topic to ensure the accuracy of the whole 2D to 3D conversion process. With a well-trained model, we can get a delicate 3D model by simply input a 2D image within a short time. This technology can help those manufacturing industry and even the whole human society work way more efficient.

***Applications*—**As the 3D modeling field is commonly used in all different areas of whole society, its real-life applications also vary in different aspects. For manufacturing industry, engineer may have to use some software like CAD to draw design map of some electronic components. (Figure 1) We can use this technology to get the 3D model and use 3D printing to do massive production within a short time.

图示

描述已自动生成

**Figure 1**

**For architecture design or real estate industry, reconstructing 3D scenes from 2D images can help architect build up the general 3D models from simple blueprint which can save lots of time and help company to cut down the design cost and run more efficiently. (Figure 2) This technology can also help real estate agent or decoration company to demonstrate a vivid scene of their designs which can bring customers more direct and deeper feelings and help these company become more competitive in the market.**

图示

描述已自动生成

**Figure 2**

**For entertainment area, reconstructing 3D scenes from 2D images is widely used in film making or software development. A big box film like The Avengers series pay billions of moneys in CG scenes making. (Figure 3) With a matured 2D to 3D conversion technology, Marvel can save countless money and present a better visual treatment.**

游戏的截图

中度可信度描述已自动生成

**Figure 3**

**And one of the biggest game developing company Blizzard Entertainment also use this technology to generate game characters or playing scenes. (Figure 4)**

游戏的截图

中度可信度描述已自动生成

**Figure 4**

**In general, the technology of reconstructing 3D scenes from 2D images can bring much convenience to item producing and help the whole manufacturing industry even the whole human society works more efficiently.**

**Meanwhile this technology is a good method to do 3D character modeling or scenes reconstructing for entertainment field which can enrich and improve human’s entertainment life. Study shows a rich and peaceful entertainment life can bring people sense of satisfaction thus bring stable public order to whole human society.**

As for all the application fields mentioned before, I am most interested in 3D scenes reconstructing in CG or in game development.

***Literature review***

**As there are already many research and papers talking about this 3D reconstructing from 2D images. I only take few work to talk about the whole process about how this technology is developed step by step.**

**Traditional approaches to 3D reconstruction rely on an intermediate representation of depth maps prior to estimating a full 3D model of a scene. To convert a 2D object into 3D model, the most important part is to predict the depth of this object. The established approach to address this task is SLAM or SfM[1],which reconstructs 3D scenes based on feature-point correspondence with consecutive frames or multiple views. That means we will try to fetch some most significant points to represent and detect these 2D objects and then to predict the depth of these object to form 3D models.**

**But this method requires multiple views of a single object. As long as we do not have multiple views images, the feature point method works poorly when different view images are not enough. So, people introduce a monocular depth estimation. However, as shown in Fig 5, existing monocular depth estimation methods[2,3] lone are unable to faithfully recover an accurate 3D point cloud.**

图示

描述已自动生成

**Figure 5**

But the monocular depth estimation requires leveraging high level scene priors, data-driven approaches have become the key solution to this problem. [4,5]

Recent works have shown promising results by training deep neural networks on diverse in-the-wild data, e.g. web stereo images and stereo videos [6,7]

However, the diversity of the training data also poses challenges for the model training, as training data captured by different cameras can exhibit significantly different image priors for depth [8] estimation

To overcome these challenges, I found a paper using a novel monocular scene shape estimation framework that consists of a depth prediction module and a point cloud reconstruction module.

It use two-stage single image 3D shape estimation pipeline is illustrated in Fig. 6. It is composed of a depth prediction module (DPM) and a point cloud module (PCM). The two modules are trained separately on different data sources, and are then combined together at inference time.[9]

图示

描述已自动生成

Figure 6

***Open source research and duplicate results*—**As for the open sources in internet about this topic, I want to introduce the most commonly used one, PyTorch3D.

PyTorch3D is to help accelerate research at the intersection of deep learning and 3D. 3D data is more complex than 2D images and while working on projects such as [Mesh R-CNN](https://github.com/facebookresearch/meshrcnn) and [C3DPO](https://github.com/facebookresearch/c3dpo_nrsfm).

This open source project is developed by Facebook AI lab team, and the most attracting part of this open source project is that it still maintains a relative high update speed. Its latest activity is 2 days ago which means developers do not have to worry about if the api if up to date.

For my interested part, the character generation, PyTorch3D provide a metho to deform a sphere mesh to any object you want. As the Fig 7 shows, its tutorials show how to deform a sphere mesh to a dolphin. With a flexible mesh deforming method, we can generate any 3D character in a convenient way.

图片包含 游戏机, 女人, 过滤网

描述已自动生成图片包含 文本

描述已自动生成

**Figure 7**

**Also, after generating the object,** PyTorch3D provide texture and render method for developer like me to make a great 3D model from a 2D images of some characters.

**Reference**

**[1] Richard Hartley and Andrew Zisserman. Multiple view geometry in computer vision. Cambridge university press, 2003.**

**[2] David Eigen, Christian Puhrsch, and Rob Fergus. Depth map prediction from a single image using a multi-scale deep network. In Proc. Advances in Neural Inf. Process. Syst., pages 2366–2374, 2014.**

**[3] Lijun Wang, Jianming Zhang, Oliver Wang, Zhe Lin, and Huchuan Lu. Sdc-depth: Semantic divide-and-conquer network for monocular depth estimation. In Proc. IEEE Conf. Comp. Vis. Patt. Recogn., pages 541–550, 2020.**

**[4] Zhengqi Li and Noah Snavely. Megadepth: Learning singleview depth prediction from internet photos. In Proc. IEEE Conf. Comp. Vis. Patt. Recogn., pages 2041–2050, 2018.**

**[5] Ren´e Ranftl, Katrin Lasinger, David Hafner, Konrad Schindler, and Vladlen Koltun. Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset**

**transfer. IEEE Trans. Pattern Anal. Mach. Intell., 2020.**

**[6] Weifeng Chen, Zhao Fu, Dawei Yang, and Jia Deng. Singleimage depth perception in the wild. In Proc. Advances in Neural Inf. Process. Syst., pages 730–738, 2016.**

**[7] Weifeng Chen, Shengyi Qia n, David Fan, Noriyuki Kojima, Max Hamilton, and Jia Deng. Oasis: A large-scale dataset for single image 3d in the wild. In Proc. IEEE Conf. Comp. Vis. Patt. Recogn., pages 679–688, 2020.**

**[8] Jose M Facil, Benjamin Ummenhofer, Huizhong Zhou, Luis Montesano, Thomas Brox, and Javier Civera. Camconvs: camera-aware multi-scale convolutions for singleview depth. In Proc. IEEE Conf. Comp. Vis. Patt. Recogn., pages 11826–11835, 2019**

**[9] Learning to Recover 3D Scene Shape from a Single Image**