JURIHUI LI
lijrh@bu.edu
EC601 A2 Project 1

Photogrammetry: From Images to 3D point clouds

Computer Vision, Imaging and Optical Science

1. Introduction

Photogrammetry and 3D point clouds are contributing to the visualization and statistics of many real-world data. Photogrammetry is a technology that uses the image of the subject to reconstruct the spatial position and three-dimensional shape of the object. [1] The main task of photogrammetry is to measure topographic maps of various scales, establish a database, and provide basic data for various geographic information and housing construction.

A point cloud is a set of data sets in space: in 3D measurement, relying on measuring instruments to collect data on the target object, a massive set of points on the target surface can be obtained. [2] Each point contains geometric coordinates, intensity values and other information. [3] These points are combined to form a point cloud. The point cloud can more realistically restore the three-dimensional effect of the target object and realize visualization. [3]

The point cloud image construction contains richer visual information and can improve the accuracy of visual measurement. With these advantages, point cloud image construction can improve the quality of visual output in application scenarios; augmented reality can get more accurate visual effects. The applications of 3D point cloud technology are 3D environmental simulation of dense areas, autonomous driving, industrial manufacturing, medical 3D reconstruction, etc. [4]

2. Current Research

At present, the main research on obtaining 3D point cloud data from images can be divided into two types from the objects to be photographed, one is large scenes or objects, and the other is relatively small objects. [5] Regardless of the type, however, computing power increases exponentially with the number of points, which puts a lot of pressure on computing time and computer memory.

The traditional computational method, the Iterative Closest Point Algorithm (ICP) iteratively performs correspondence estimation and least squares optimization for transform estimation. [6] In the current state of the art, the scene is represented as a neural radiation field for view synthesis. Then embeds an entire scene into the weights of a feedforward neural network. Trained by backpropagation through a differential volume rendering procedure. [7] Achieves 3D state-of-the-art synthesized scenes using a sparse set of input images.

3. Methodology

Current established approaches to Photogrammetry follow a multi-step pipeline to transform a series of images of a subject into a 3D model. Used to create everything from small scale 3d models to large surveying projects. Images can be captured with any camera, from cell phones to drones. Boulch et al. placed cameras at multiple locations, generating point cloud data from multiple RGB and depth photos. They then processed these photos pixel-by-pixel using a 2D segmentation network, and the resulting data from the RGB and depth images were further fused and analyzed using residual correction. [8] Tatarchenko et al. introduced tangent convolution for dense point cloud segmentation, the geometry around each point cloud is first projected onto a virtual tangent plane, and then the tangent convolution is applied directly to the geometry. [9]

4. Analysis and solutions

Here we discuss methods for obtaining 3D point cloud data using devices such as cameras or imagers. First, we scan and record the objects to be photographed through multiple cameras, so that we can obtain image object data. After that, 3D models of image data and point cloud data are created using open-source algorithms, etc. Finally, the 3D point cloud data is exported, and the neural network algorithm is used for image training and synthesis.

5. Limitations and Drawbacks

1) Lack of data: it is impossible to make every data point available; scanned model objects may cause data loss due to irregularities or occlusions.
2) Noise: The sensor will generate noise, which will lead to errors and data anomalies, it takes time to select the correct data points and remove outliers.
3) There is also a graph representation in the point cloud, and feature extraction is performed by constructing a graph neural network and graph convolution. But the disadvantage of graph neural network is that it is computationally time-consuming, which will be a serious limitation for practical applications. [5]
4) The acquisition of point clouds is generally obtained by 3D imaging sensors, which are subject to many hardware limitations in this process. [10] For example, there will be problems in the output image, such as reflection on the surface of the object, transparent objects, etc. will cause the loss of image data.
5) 3D point cloud data extraction: In the images obtained by shooting, different 3D imaging sensors will have different pixel values, and different pixel values will cause differences and changes in point cloud data. [10] So, the amount of extracted 3D point cloud data can vary greatly depending on the sensor.
6) 3D point clouds have unstructured properties.

6. References

[1] https://en.wikipedia.org/wiki/Photogrammetry
[2] https://en.wikipedia.org/wiki/Point_cloud
[3] Y. Guo, F. Sohel, M. Bennamoun, M. Lu, and J. Wan, “Rotational projection statistics for 3D local surface description and object recognition,” IJCV, 2013.
[4] X. Chen, H. Ma, J. Wan, B. Li, and T. Xia, “Multi-view 3D object detection

network for autonomous driving," in CVPR, 2017.
[5] Treating Point Cloud as Moving Camera Videos: A No-Reference Quality Assessment Metric. arXiv:2208.14085v1(2022)
[6] C. Chen, G. Li, R. Xu, T. Chen, M. Wang, and L. Lin, "ClusterNet: Deep hierarchical cluster network with rigorously rotationinvariant representation for point cloud analysis," in CVPR, 2019
[7] Mildenhall, Ben, et al. "Nerf: Representing scenes as neural radiance fields for view synthesis." Communications of the ACM 65.1 (2021): 99-106.
[8] A. Boulch, B. Le Saux, and N. Audebert, "Unstructured point cloud semantic labeling using deep segmentation networks." in 3DOR, 2017.
[9] M. Tatarchenko, J. Park, V. Koltun, and Q.-Y. Zhou, "Tangent convolutions for dense prediction in 3D," in CVPR, 2018.
[10] F. Zhang, D. Hall, T. Xu, S. Boyle and D. Bull "A Simulation Environment for Drone Cinematography." arXiv:2010.01315v1