Boston University

EC 601

Project 1

Haoxiang Sun

Building 3D scenes from 2D images using ML algorithms

Abstract

Human beings are entering the information age, and computers will enter almost every field more and more widely. On the one hand, more and more people without computer training need to use computers. On the other hand, the functions of computers are becoming more and more powerful, and the methods of using computers are becoming more and more complicated. This creates a sharp contradiction between the flexibility of conversation and communication and the rigor and rigidity required in the use of a computer. People can exchange information with the outside world through vision and hearing, language, and can express the same meaning in different ways, but the computer is required to write programs strictly by various programming languages, only in this way can the computer run. To enable more people to use the complex computer, it is necessary to change the past that people adapt to the computer, to memorize the rules of computer use. But in turn, let the computer adapt to people's habits and requirements, in the way that people are used to information exchange with people, that is, let the computer with vision, hearing and speech and other abilities. Then the computer must have the ability of logical reasoning and decision making. A computer with these capabilities is an intelligent computer.

1. Introduction

Due to the improvement of people's living standards nowadays, society's demand for technological development has shown exponential growth, such as the conversion of two-dimensional pictures into three-dimensional models. This technology will be applied in many fields, simple examples such as 3D perception of household sweeping robots, or game modeling, the manufacture of machine parts, etc. So 2D to 3D technology is especially important, and more importantly, the tools needed in it, such as point cloud, machine learning, etc. The point cloud is a very important tool, it refers to a massive collection of points on the surface characteristics of a target. The point cloud obtained according to the principle of laser measurement includes three-dimensional coordinates (XYZ) and laser reflection intensity (Intensity). The point cloud obtained according to the principle of photogrammetry includes three-dimensional coordinates (XYZ) and color information (RGB). The point cloud is obtained by combining the principles of laser measurement and photogrammetry, including three-dimensional coordinates (XYZ), laser reflection intensity (Intensity), and color information (RGB). After obtaining the spatial coordinates of each sampling point on the surface of the object, what is obtained is a collection of points called "Point Cloud".

These tools will help us get the 3D models people need more quickly and save costs, which are an important part of many industries. The manufacture of auto parts and the development of artificial intelligence machine tools can greatly save labor.

2. Applications

Because people's economic level is greatly improved, their demand for life and entertainment products has also increased. These technologies have also become an indispensable part of the game. For example, in the game "Onmyoji" developed by a Chinese company, players can see their favorite characters dancing by scanning a picture, which is also one of the unique selling propositions of this game.

Boston Dynamics BigDog's (Fig 1)stereo vision system is also one of the applications. It can help police or soldiers to detect some dangerous places, like finding criminals who are hiding, etc., which greatly reduces the casualty rate. Similar to BigDog, there is also an engineering vehicle, which can go to places that humans cannot reach, such as extremely cold or hot ones. There are many similar applications. It saves human resources, reduces casualties, and can also entertain the public and improve the standard of living of human beings. However, I am most interested in computer vision for autonomous driving. During autonomous driving, it will automatically scan the road conditions and can do things that humans cannot do in dangerous situations, such as emergency braking, etc.

Fig. 1

3. Literature review

Computer vision is the soul of autonomous driving. It is the "eye" of a car. So what is computer vision? It is a science that studies how to make machines "see". Furthermore, it refers to the use of cameras and computers instead of human eyes to recognize, track, and measure targets, and then perform image processing. It is more suitable for the human eye to observe or transmit the image to the instrument for inspection. Under normal circumstances, a car will install two cameras to construct the road conditions and vehicle conditions ahead through real-time shots. So Camera calibration is a very important task in 3-D computer vision particularly when metric data are required for applications involving accurate dimensional measurements.

Calibrating a camera involves determining its intrinsic parameters (generally, the two coordinates in the image frame of the intersection of the optical axis with the image plane, the two-scale factors along the vertical and horizontal axes of the image frame, and, if need be, the lens distortion parameters) and its position and orientation concerning an arbitrary world reference frame. Calibrating a stereovision sensor made up of two cameras involves determining the intrinsic parameters of each camera and the relative position and orientation between the two cameras (Fig. 2). These calibration data are required to compute, by triangulation, the 3-D coordinates of a point corresponding to matched pixels on the two images. (1)Through these camera captures and simple calculations of the system, the driving route during autonomous driving can be obtained. (Fig. 2)

Fig. 2

Since it comes to cameras, image processing technology is also very necessary. In the article Autonomous High-Speed Road Vehicle Guidance by Computer Vision1, an image processing system BVV is mentioned:

The image processing system (Fig. 3) has been developed as a tool for the study of basic algorithms and architectures for real-time image sequence analysis. This system allows 14 parallel processors (PPS) to access digitized images, and each processor pulls out whatever part of the image it happens to be interested in. It is usually a collection of line elements detected using the correlation method of terminal masks and reports their positions and slopes to a higher-level processor that can integrate the results of multiple PPS to identify perspective projections of known objects in a three-dimensional world. When the processor receives information that contradicts its real-world model, it can redirect a PP to a new image region and, upon successful pass, command the PP to track its characteristics independently on the image sequence. Multiple PP Windows can move and overlap freely throughout the image area without memory access interference. The degree of function parallelism can be obtained by having PP study the same image region with different algorithms. Where a feature should be visible to support ongoing object assumptions. Hence the state of the vehicle relative to the road. (2)These technologies will be of great help to our lives and there are many places where they can be applied, such as

(1) Control process, for example, an industrial robot;
(2) Navigation, for example, mobile robots;
(3) Events detected, such as video surveillance and people counting;
(4) Organizational information, for example, an index database for images and image sequences;
(5) Modeling objects or environments, such as medical image analysis systems or terrain models;
(6) Interaction, for example, when input to a device for computer-human interaction;
(7) Automatic detection, for example, in manufacturing applications.

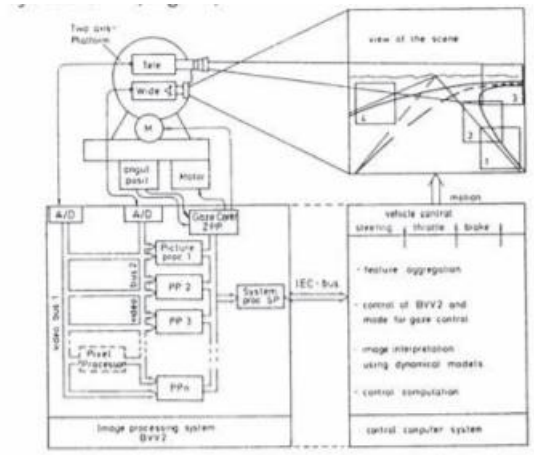

Fig. 3

4. Open Source research

Regarding open source, we can use PyTorch3D, PyTorch3D provides efficient, reusable components for 3D Computer Vision research with PyTorch. It has many functions such as Data structure for storing and manipulating triangle meshes Efficient operations on triangle meshes (projective transformations, graph convolution, sampling, loss functions)

A differentiable mesh renderer. PyTorch3D is designed to integrate smoothly with deep learning methods for predicting and manipulating 3D data.. so it can
1) be implemented using PyTorch tensors
2) Can handle mini-batches of heterogeneous data
3) Can be differentiated
4) Can utilize GPUs for acceleration
5) Within FAIR, PyTorch3D has been used to power research projects such as Mesh R-CNN(3)

5. Conclusions

Computer vision is not only an engineering field but also a challenging and important research field in the scientific field. Computer vision is a comprehensive discipline, it has attracted researchers from various disciplines to participate in its research. Computer vision is closely related to human vision. A correct understanding

of human vision will be very beneficial to the research of computer vision. Our 2D to 3D is also part of it, and it is also our research direction.

Reference

1. Orteu, Jean-José. "3-D computer vision in experimental mechanics." Optics and lasers in engineering 47.3-4 (2009): 282-291.
2.Dickmanns, Ernst D., and Alfred Zapp. "Autonomous high speed road vehicle guidance by computer vision." IFAC Proceedings Volumes 20.5 (1987): 221-226.
3. Github link: https://github.com/facebookresearch/pytorch3d