Obtaining 3D Information from 2D Images Efficiently using Feature Extraction and Laser Dot Projection

Pratham Gandhi, Samuel Schuur

Bronx, New York

Abstract

A system to extract 3D information about a space was created that could be trained simply in a matter of minutes. The system, nicknamed "Constellation", was designed to be implemented in situations requiring rapidly updating computations and environmental awareness, to run on relatively inexpensive hardware, and to perform its tasks with limited human intervention. To achieve this, Constellation uses a combination of hardware and software components, focused around a set of artificial neural networks to extract features and a physical grid of points refracted from a single laser. A Python 3.6 implementation of the system was tested on two machines in a variety of environments, and produced an average of x times faster environment generation as other comparable algorithms or approaches to similar problems.

NYCSEF 2018

1. Introduction

The current most popular approach to creating a three dimensional informational estimate from a two dimensional image is triangulation, or binocular, stereo vision approach. Such approaches take several images of the same scene from different locations and angles in order to generate a spatial awareness. Such systems, however, are limited in practical applications in our increasingly evolving digital world, due to the limiting factor that they require multiple images, the acquisition of which is sometimes impractical. Additionally, they can sometimes be quite slow, in that they need to first identify common features in the different photos, find out the positional relation of those features, and then only begin to start seeing the greater picture of the full space. The purpose of Constellation was to solve both of these problems with a generally applicable approach which can be implemented in a variety of different use-cases.

We chose to approach this problem first through the somewhat traditional technique of using artificial neural networks, mathematical functions modeled after nature which specialize in taking in large amounts of input to produce outputs, to identify features in a two dimensional image. Neural networks are are perfect for navigating the fuzzy problems of feature extraction from images. Where our system differs is in its use of a grid of refracted laser points into the scene to judge the distances and orientations of various objects in relation to the camera. This allows superior environment generation in a shorter amount of time, due to the omission of the multiple-image stereo aspect. This approach and its implemented system was named Constellation for the star-like appearance of the laser dot grid it relies on.

2. Background

2.1. Binocular Computer Stereo Vision

In order to perceive depth the human eye uses many visual cues, both monocular and binocular. Of these visual cues stereopsis, the perception of depth resulting from the brain comparing the differences in images produced by both eyes, is in many ways one of the easiest depth perceptions methods to mimic through software, as it is the least reliant on external sources of stimuli and human experience. This is what Stereo Computer Vision attempts to do; building depth maps by comparing scenes from multiple vantage points. More specifically Binocular Computer Stereo Vision does this by comparing the images from two cameras and by extracting and comparing common features from both images it builds a depth map of the scene in question.

2.1.1. Removing Distortion from Images

2.2. Structure-From-Motion Pipelines

3. Approach

4. Implementation

4.1. Hardware

4.1.1. Refracting Grid of Laser Points

4.1.2. Computers

4.2. Software

4.2.1. Object Oriented vs. Imperative Neural Network Design

4.2.2. Feature Extraction

4.2.3. Distance Estimation

5. Results

5.1. Algorithmic Analyses

5.1.1. Computer Stereo Vision

5.1.2. Structure-From-Motion Pipeline

5.1.3. Constellation

6. Conclusion