Depth Estimation

SK.BOO

August 1, 2023

Abstract

This paper presents an overview of the recent advances in monocular depth estimation techniques. Depth estimation is the process of estimating the distance between a camera and each pixel in an image. This information can be used for a variety of applications, such as autonomous driving, robotics, and virtual reality. Traditionally, depth estimation techniques have been based on geometric cues by image structures. However, these techniques are often limited by the quality of images and the complexity of scenes. Recently, deep learning has been used to achieve state-of-the-art results in monocular depth estimation. Deep learning models can learn to estimate depth from a large dataset of images and labels. This allows them to generalize to new scenes and conditions that are not present in the training data.In this paper, we review the different approaches to monocular depth estimation. We also discuss the challenges and limitations of these approaches. Finally, we provide an overview of the future directions of research in this area.

Keywords— monocular depth estimation, deep learning, image structure

1 Introduction

Human beings perceive three-dimensional information based on the input received from both eyes, enabling depth perception. Similarly, by utilizing two cameras in tandem, a 3D image can be generated. When observing a single image, we can roughly estimate perspective based on geometric features such as shadows and occlusions, extracting these characteristics to measure depth. However, variations in lighting conditions and resolution can significantly affect the accuracy of such depth estimation methods.

In December 2014, at the NIPS conference, it was demonstrated that depth estimation from a single image is feasible using deep learning techniques. Since then, numerous studies on monocular depth estimation using deep learning have been conducted, leading to significant advancements.

This paper examines depth estimation through image structure and depth estimation through deep learning.

2 Image Structure

Image structure models based on image statistics used a wide range of applications, such as image indexing and similarity calculation, and research on natural image models. In this paper, we describe two levels of image structure based on the second-order statistics of images. The first level is the size of the global Fourier transform of an image, which only contains non-localized information about the major directions and scales that make up the image. The second level is the size of the local wavelet transform, which spatially approximates the major directions and scales within an image.

2.1 Unlocalized Image Structure and Spectral Signatures

The discrete Fourier transform(FT) of an image is defined as:

$$I(\mathbf{f}) = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} i(\mathbf{x}h(\mathbf{x})e^{-j2\pi}$$

The image is analyzed by spectral analysis using Fourier transform.

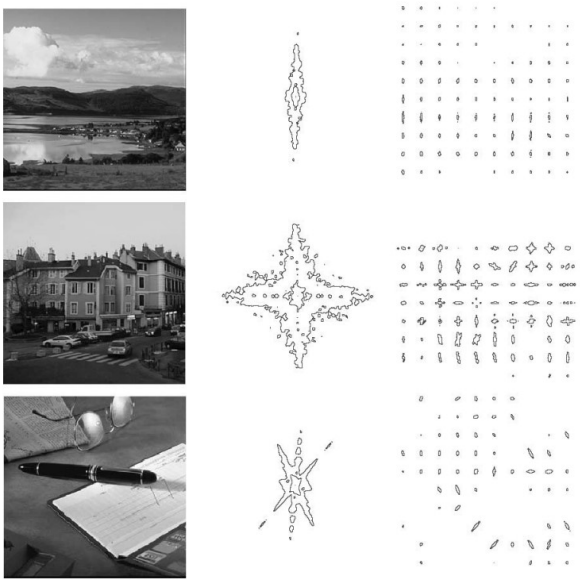

Figure 1: fourier transform

2.2 Spatial Organization and Local Spectral Signatures

2.3 Image Structure as a depth cue

The average depth of the image is mentioned as a measure or scale of spatial extent. The variability observed in the spectrum characteristics is attributed as a primary factor associated with the average depth.

3 Monocular Depth Estimation technology based on Deep Learning

Deep learning has greatly advanced the field of image recognition and has recently been applied to monocular depth estimation. Monocular depth estimation is a technique for estimating the distance to an object using only one image. Research is being done based on stereo images or the relationships between frames within a single image, without using sensors such as LiDAR, such as in self-driving cars.

Deep learning-based monocular depth estimation techniques have advanced significantly in recent years. Representative deep learning algorithms include Monodepth, DenseDepth, and FastDepth. Research is also being actively conducted to improve performance by combining with the latest models such as transformers and self-supervised learning.

3.1 Monodepth

Monodepth is a deep learning-based monocular depth estimation algorithm that uses epipolar geometry constraints to train with image reconstruction loss and generates stereo images.

Epipolar geometry is the geometry of the relationship between two images of the same scene taken from different viewpoints. This relationship is governed by the epipolar constraint, which states that the corresponding points in the two images lie on lines called epipolar lines.

Monodepth uses the epipolar constraint to train a deep neural network to estimate the depth of each pixel in an image. The network is trained on a dataset of stereo images, where the depth of each pixel in one image is known. The network learns to predict the depth of each pixel in the other image by minimizing the image reconstruction loss.

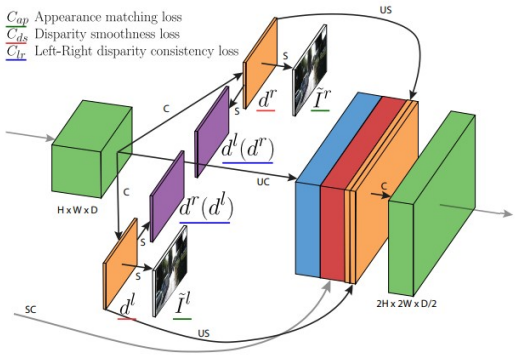

Figure 2: monodepth model structure

4 Conclusion

Spectral analysis is a method of analyzing the structure of an image by looking at its frequency distribution. You can use the frequency distribution of images to distinguish between artificial objects and their natural environment. This is because artificial objects generally have different frequency distributions from their natural environments. You can also use spectroscopy to estimate the average depth of an image. Therefore, it is possible to additionally provide situation information on an image using this. This will then be available when analyzing and learning 3D images.

A method of estimating the relationship between images through learning image data rather than human-designed features using deep learning is constantly being researched. Various model structures and loss functions have been introduced, and recently, research is being conducted through new structures based on transformers. This can be applied anywhere dealing with three dimensions.