Computer Vision Group Project

Monocular Depth Estimation

Rahul Allam, Akash Thorat, Lorena Maria Genua and Gaurav Bhosale

Depth Information is important for autonomous systems to perceive environments and estimate their own state. Depth estimation is a classic task in computer vision, which is of great significance for many applications such as augmented reality, target tracking, and autonomous driving.

In this project we are trying to solve depth estimation from a single image frame with help of neural networks. Traditional depth estimation methods of image-based depth estimation are usually based on stereo camera, which calculates the disparity of two 2D images through stereo matching and triangulation to obtain a depth map.

The problem can be framed as: given a single RGB image as input, predict a dense depth map for each pixel. A single RGB image can be viewed as follows. Let I be the space of RGB images and D the domain of real-valued depth maps. Given a training set $T = \{(I_i , D_i)\}^M{}_{i=1}$, $I_i \in I$ and $D_i \in D$, the task is to learn a non-linear mapping $\Phi : I \rightarrow D$.

Approach:

When only a single image is available, it is not possible to use epipolar geometry. Instead, algorithms have to rely on pictorial cues: cues that indicate depth within a single image, such as texture gradients or the apparent size of known objects. Monocular cues such as: Texture and Gradient Variation, Shading , color/Haze, and defocus aid in accurate depth estimation. These are complex statistical models which are susceptible to noise.

With the arrival of stronger hardware and better machine-learning techniques – most notably Convolutional Neural Networks (CNN) – it is now possible to learn pictorial cues rather than program them by hand. One of the earliest examples of monocular depth estimation using machine learning was published in 2006 by Saxena et al.

We use aim to use CNN and Encoder-Decoder architecture to learn these parameters, while experimenting with different architectures to get the best results such as Resnets, VGG etc. We will be using the NYU depth dataset V2.

Applications:

Monocular depth estimation has been useful, and it has vast application demands due to the availability of only one single camera in most application scenarios. The applications are numerous from depth estimation from your mobile camera, mobile robots, autonomous robots etc. getting the monocular depth estimation as good as stereo will remove excessive hardware costs with respect to cameras.