IVDC INDUCTION

PS1: MONOCULAR DEPTH ESTIMATION

TASK1:

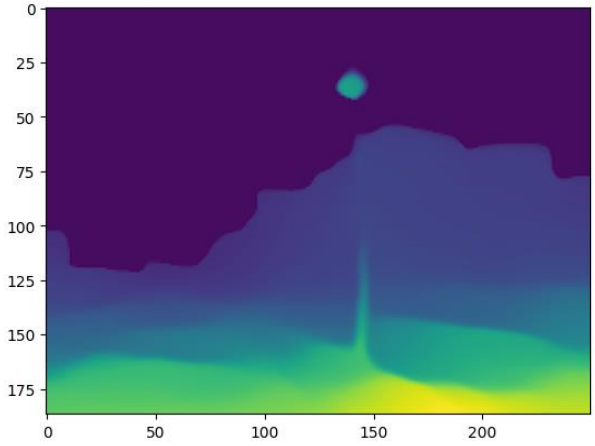
Traditional Methods:

1. Structure from Motion (SfM): SfM involves estimating both the camera motion and the 3D structure of the scene by analyzing multiple images or video frames taken from different positions.
2. Stereo Matching: This technique uses multiple images of the same scene taken from slightly different positions to calculate the depth by finding the corresponding points between the images and using the concept of parallax theorem.
3. Depth from Defocus: This method analyzes the blur or defocus information in an image to estimate depth by understanding the relationship between depth, focal length, aperture size, and object distance.
4. Time-of-Flight(ToF): Time-of-Flight cameras emit coloured pattern light or infrared signals and measure the time taken for the signals to bounce back, allowing for depth estimation based on the round-trip time and position.

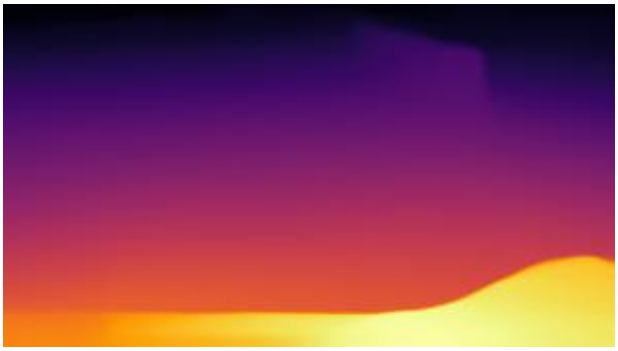
Machine Learning-based Methods:

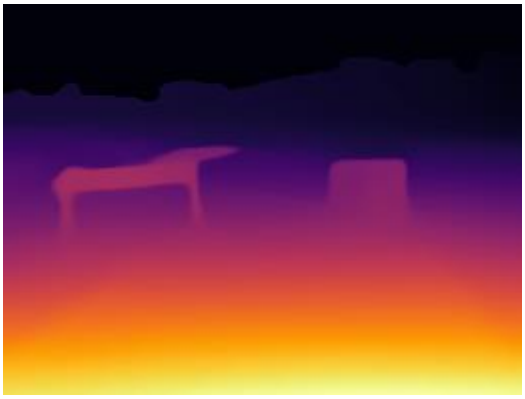
1. Convolutional Neural Networks (CNN): CNNs have been used for depth estimation by training them on large datasets of paired RGB and depth images. The network learns to map RGB images to corresponding depth maps.
2. Recurrent Neural Networks (RNN): RNNs, particularly Long Short-Term Memory (LSTM) networks, have been employed for depth estimation tasks by modeling the sequential dependencies in image sequences or video frames.
3. Fully Convolutional Networks (FCN): FCNs are designed to output dense pixel-wise depth maps by utilizing convolutional layers and skip connections to capture multi-scale information.
4. Point Cloud-based Methods: These methods convert images to 3D point clouds and estimate depth by training deep neural networks on the point cloud data or by employing point cloud registration techniques.

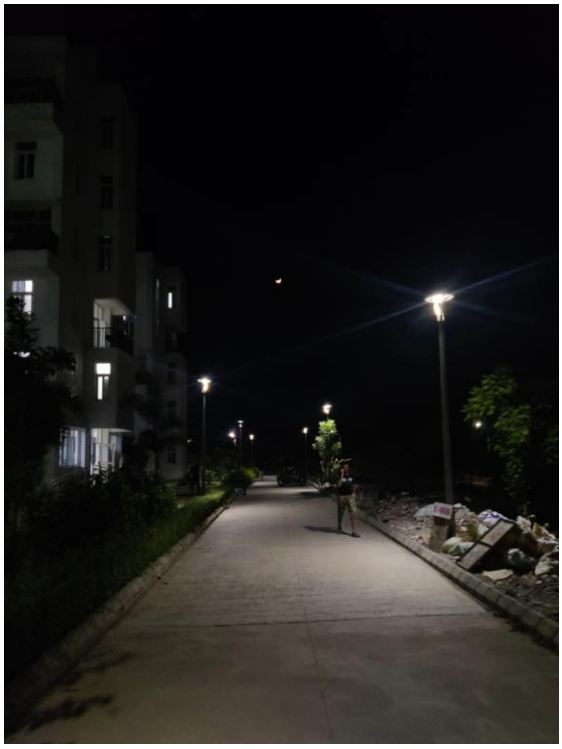
TASK2:

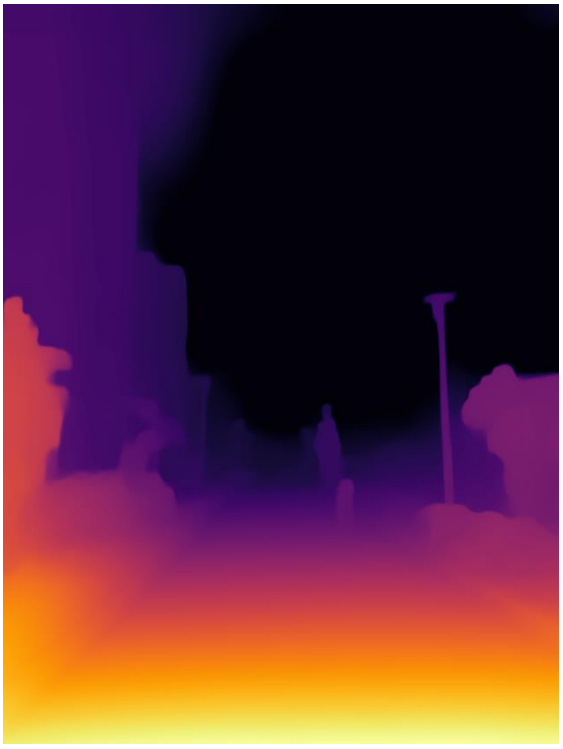

0
25
50
75
100
125
150
175
0
50
100
150
200


University of New South Wales IGVC2015 Advanced Course Run









All the images were processed using MiDaS in online google colab. The first version was processed in a different version of MiDaS and hence we can see the variation.

Here the lighter colour refers to closer pixels and the darker ones for the farther pixels.

Pre trained model of MiDaS was used for processing all the above images. Since incorporating videos in word file is more difficult it has been explicitly uploaded (output_sound.mp4) in the github repo. Kindly look into them.

-GIRIDHAR A P

220002033

ELECTRICAL ENGG.

IIT INDORE