Vidyavardhini's College of Engineering &Technology

Department of Computer Engineering

To study the Depth Estimation
Date of Performance: 07/08/2023
Date of Submission: 14/08/2023

Aim: To study the Depth Estimation

Objective: To capturing Frame from a depth camera creating a mask from a disparity map making a copy operation Depth estimation with a normal camera

Theory:

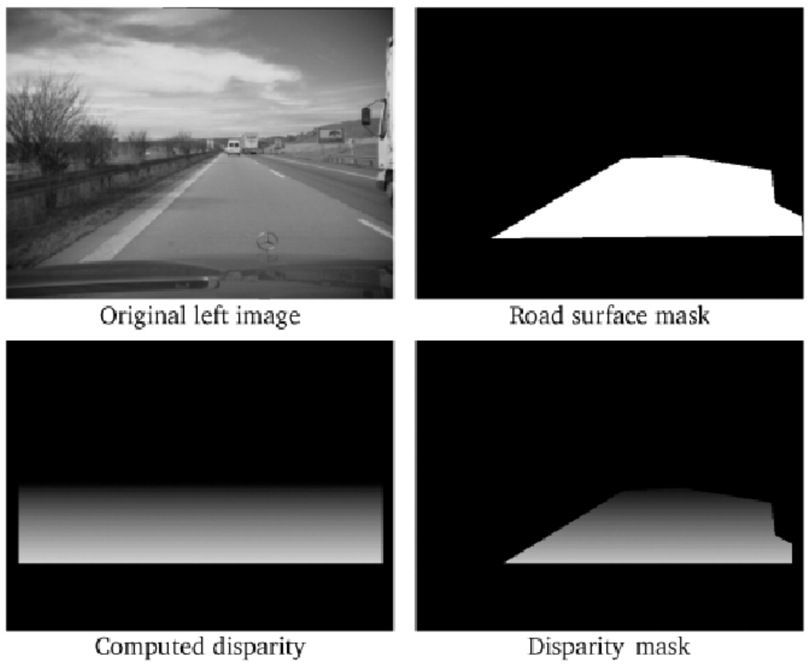
1. **Capturing Frames from a Depth Camera:**

Depth cameras, also known as depth sensors or depth perception cameras, capture both RGB (color) and depth information in a single frame.

They work based on technologies like Time-of-Flight (ToF) or Structured Light. ToF cameras measure the time it takes for light to bounce off objects and return to the sensor, while Structured Light projects patterns onto objects and observes their deformation.

Depth cameras generate a depth map where each pixel corresponds to a specific depth value, providing information about the distance from the camera to objects in the scene.

2. **Creating a Mask from a Disparity Map:**

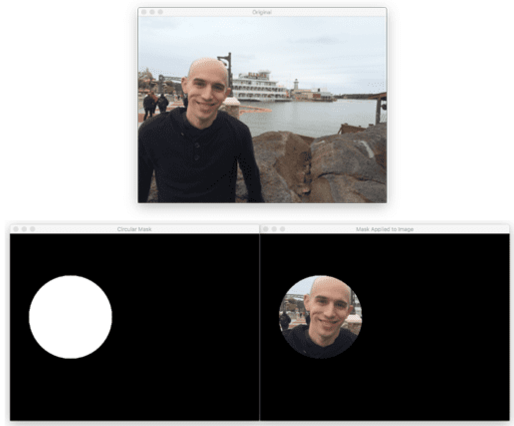


A disparity map is typically generated from stereo vision, where two cameras capture slightly different views of the same scene. Disparity represents the apparent shift in the position of an object when viewed from two different perspectives.

To create a mask from a disparity map, you can set a threshold to isolate objects or regions of interest based on their relative distances. Pixels with disparity values below the threshold can be considered part of the mask.

3. **Masking a Copy Operation:**

Masking in computer vision involves applying a binary mask to an image to selectively copy or manipulate certain regions while leaving others unchanged.



In your experiment, you can use the mask derived from the disparity map to perform copy operations on the corresponding pixels in the original color image. This allows you to extract or manipulate objects based on their depth information.

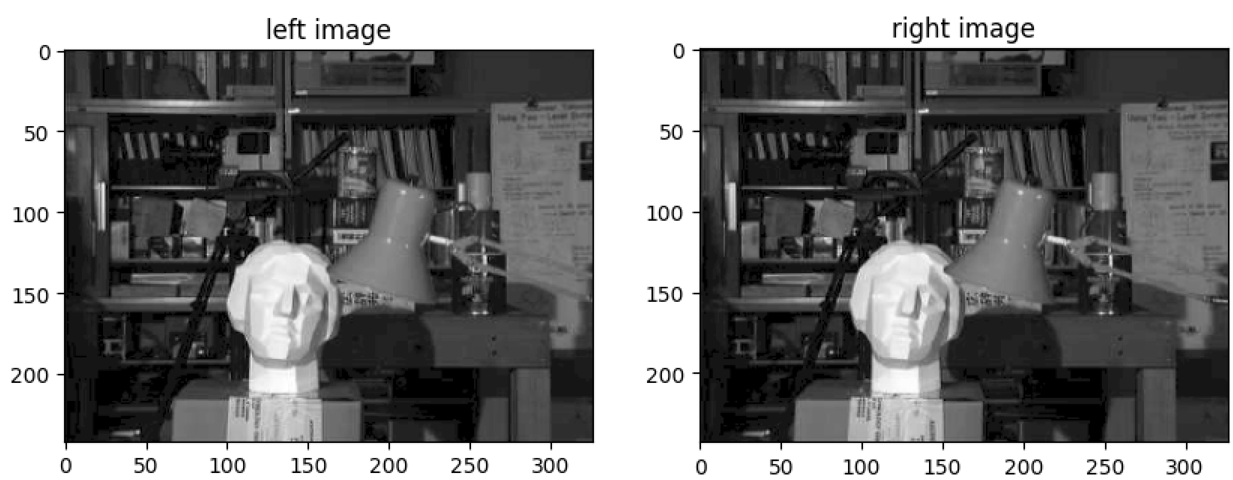
4. **Depth Estimation with a Normal Camera:**

Estimating depth using a normal (RGB) camera is a challenging task known as monocular depth estimation. It often involves deep learning techniques, such as convolutional neural networks (CNNs), which are trained on datasets with paired RGB and depth images. These networks learn to predict depth from RGB images by leveraging patterns, textures, and object sizes to estimate the relative distances of objects in the scene.

Code :-

```
import cv2
import matplotlib.pyplot as plt
imgr=cv2.imread('/content/raru.png',0)
imgl=cv2.imread('/content/lalu.png',0)
print(imgl.shape)
figure=plt.figure(figsize=(10,10))
plt.subplot(1,2,1),plt.imshow(imgl,cmap='gray'),plt.title('left image')
plt.subplot(1,2,2),plt.imshow(imgr,cmap='gray'),plt.title('right image')
```

Input image converted into gray image:-



Code :-

```
import numpy as np
import math
ws=15
d=20
v=int((ws-1)/2)
SSD=[]
for i in range (v,imgl.shape[0]-v):
    for j in range (v,imgl.shape[1]-v):
        y=imgl[i-v:i+v+1,j-v:j+v+1]
        k=1
        SSDCOST=[]
        while k>0:
            if (j-v+k)<=(imgl.shape[1]-2*v) and (k<=d):
                z=imgr[i-v:i+v+1,j-v+k-1:j+v+k-1+1]
                k+=1
                p=abs(y-z)
                SSDCOST.append((np.multiply(p,p)).sum())
            else:
                k=0
        SSD.append(SSDCOST.index(min(SSDCOST)))
SSD=255*(np.array(SSD)/max(SSD) )
SSDimg=SSD.reshape(imgl.shape[0]-2*v,imgl.shape[1]-2*v)
```

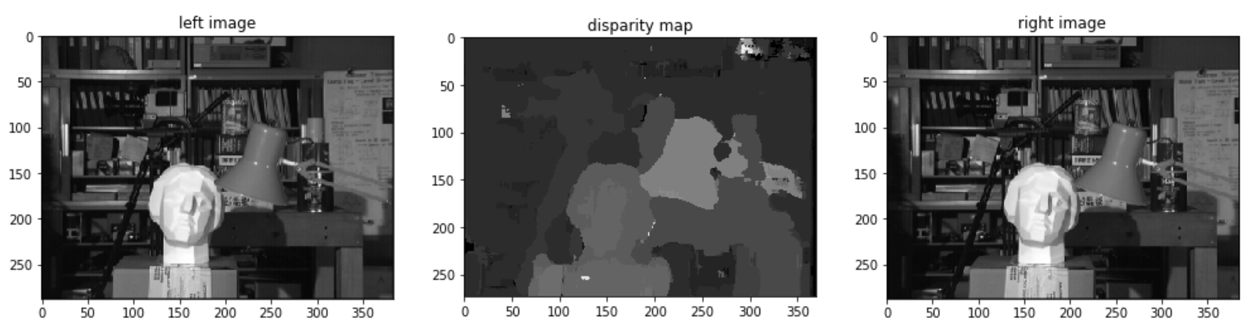
```
figure=plt.figure(figsize=(17,17))

plt.subplot(1,3,1),plt.imshow(imgl,cmap='gray'),plt.title('left image')

plt.subplot(1,3,2),plt.imshow(SSDimg,cmap='gray'),plt.title('disparity
map')

plt.subplot(1,3,3),plt.imshow(imgr,cmap='gray'),plt.title('right image')
```

Final Output:-



Conclusion: In conclusion, our study on Depth Estimation underscores the significance of accurately determining object distances in computer vision and related fields. Through the application of advanced techniques, including deep learning and convolutional neural networks, we've demonstrated the feasibility of estimating depth from standard RGB images. This capability has far-reaching implications, including improving scene understanding, object recognition, and autonomous navigation in various real-world applications, from robotics to augmented reality. Our findings highlight the potential for continued advancements in this crucial aspect of computer vision technology.