Computer Vision

End Sem Exam

Name: **Kaustav Vats**
Roll Number: **2016048**

Question 1a

Given that the camera has 2 different setups- **Setup1**: 2 camera TOF based setup, **Setup2**: 3 camera TOF based setup. These multiple cameras setup provide an effect like bokeh effect and many other features.
For 3 camera TOF based setup, camera alignment can affect the result of the image captured by the cameras. Setup2 with all 3 camera aligned horizontally and vertically can capture much wider range image, with better depth effect than setup1. Well the above effect can be achieved with one camera (Google Pixel). But the problem is that it might not provide a bokeh effect in low light conditions. So to achieve these kind of effects, we can have a camera setup as **1 SLR camera with an Infrared camera setup.**
How can these effects be achieved using only a single camera? [Ref1]

In order to provide depth effect for images, we first need identify the foreground and the background in an image. Pixel uses segmentation to partition the objects, but it still doesn't provide how much need to be blurred. Training a convolutional neural network would help us identify the person in the image. Google has trained CNN's on millions of photos, including person with different clothing style, wearing hat, mask etc. From this trained model it identifies the person in an image and is able to segment it. SLR camera provides a shallow depth of effect. Which means that objects that are closer as sharper than the objects that are farther away. After detecting the person, we don't want to blur the objects that are near. We sharpen those nearby objects. To avoid them getting blur in later steps. Using the google pixel technology called dual pixel auto focus.

Cameras are made such that left side and right side of lens can capture images with very small disparity. Using camera capturing burst of two image from both side of lens, stereo Depth effect can be calculated. Burst images can be used to reduce the noise and any other deformity. After calculating the depth effect, final depth map can be combined such that avoiding the person region and the pixels that are closer to the camera.

An infrared camera can also be used to calculate depth effect instead of using a single camera. Infrared camera can identify the depth and illuminate/brighter the regions that are closer to camera. Signals that are sent by infrared emitter can be detected by infrared cameras. After identifying the closer regions in an image, various sharpening filters can be used to enhance the image and provide much better low light quality images.

Now when the subject is in motion. An autofocus object tracking detection system can be used to capture the moving object with better quality. First we need to identify the moving object in consecutive frames. To do this we first extract the features and do feature tracking and calculating the optical flow of consecutive frames and identifying the region in the optical image which highlights the moving region, since there can be multiple moving regions like a person moving in front of a lake or a beach side video.

We can use above trained depth estimation model to identify the objects that are closer to the camera. Based on this we can track the objects by guessing the next position. And focus on that region for better video quality. This setup will work because we have neglected the background moving object as noise and only focusing on the actual moving objects.

Question 1b

With the above setup we can add these as a feature-

1. Adjustable depth effect- In this feature a person can adjust a lever to increase or decrease depth effect. On decreasing the depth effect region that is farther away can be deblurred and vica versa. This feature would provide personalized filter effect, which seems more pleasing to the user than the original proposed depth effect image.
   One we have identified depth in the image using the infrared sensor or by above explained models(to do depth estimation using only single camera) we can do thresholding blurring, if a pixel has a depth that is above the specified threshold then we can blur those identify such pixels and apply gaussian blurring on those set of pixels.

2. A feature that is able to identify the depth in a room and is able to numerically approximate the depth using above proposed depth calculation techniques. This feature can be used to calculate distance from camera to object. We need to manually specify the distance value based on threshold of depth map values for a pixel. Or we can train a model by manually creating a dataset for accurate result of depth based on the depth value

Question 2

Title: Content based filtering

Problem Statement: To identify and understand the content present in the video and do filtering based on it.

Aim: To identify images or videos with confidence that contains a particular objects. I want to provide censorship for particular topics in a country or for a personalized experience of viewing content on the internet.

1. Spatial component - Creating a Convolutional neural network model to identify object of interest.
2. Temporal Component - Creating a 3D Convolutional neural network model to identify content in videos. This models can identify the moving object as well by learning filters and doing 3d convolution on some consecutive frames of the image.

Why this is important?
There are some topics are necessary to be filtered for people belonging to different age or for personalize experience. For example- Sometime we don’t want to see spoilers for a particular tv series, a movie etc. Sometimes we just want to avoid content like Tik Tok videos, politics or some sensitive topics. By training the models on the dataset of particular topics we can do filtering.