Computer Vision End Semester
Open Book Exam

Submitted by: Kshitij Srivastava (MT18099)

Question 1(a):

Proposed solution: I would suggest my company to add a wide-angle camera (180-degree) to the setup, to both the front and the rear of the phone and use these cameras to capture a 360-degree panorama image in at most two shots without the need to rotate the mobile phone a lot.

<table>
<tr>
<td> 
</td>
<td>Why not a single shot when using two 180-degree cameras?
There would be some blind spot area due to the mobile phone’s body.

Solution: We can rotate the phone by a small angle (say about 5 to 10-degrees) to make up for the blind spot without compromising on the advantage that our system proposes.</td>
</tr>
</table>

Motivation: Currently, it is possible to take 360-degree panorama images with mobile phone cameras but it requires the user to rotate their mobile phones to some degrees for capturing multiple images and then stitching them together. This makes it difficult to capture good quality panoramas, mainly because it is difficult to keep the phone steady when trying to rotate.

Complete Camera Setup:

Rear:

- 1 wide-angle (180-degree) camera
- 1 telephoto lens
- 1 main camera

Front:

- 1 wide-angle (180-degree) camera
- Normal dual camera setup

A similar approach was followed by the researcher at Intel[1]. However, what they suggested was to use the default camera setup and use both cameras for the panorama. This required the user to rotate the camera a lot (about 90 degrees). Their results are below:

This is an uncropped example of a landscape capture. You can see there is slight mismatch in the camera angles and leaves much of the image unusable after cropping.

[1] "Dual-Camera 360 Panorama Application | Intel® Software." 11 Nov. 2013, https://software.intel.com/en-us/articles/dual-camera-360-panorama-application. Accessed 24 Apr. 2019.

This is an uncropped example of a portrait capture. You can see the images match up well.

Since the images match up well, we don't have to waste much of the height. The original images were 1920px tall and the result is 1640px tall. We only lost 14% of the height.

Their results look quite promising, but we want to achieve the same at lesser rotation and make taking panoramas fun.

Algorithm:

For image capturing, I'll use something similar to what Google does in HDR+ [2] and NightSight[3]. The algorithm is as follows:

- As soon as the camera app is launched, start capturing images and storing them in a cyclical buffer of "n" size.
- When the shutter button is pressed:
    - Find out the optical flow of the scene using the buffer.
        - Use Lucas-Kanade method for this
    - Select the best exposure interval based on the following optic flow value:
        - If motion is detected, keep the exposure interval small, to reduce blur.
        - Otherwise, use a long exposure interval
    - Capture multiple images of the subject
    - Align the frames
    - Merge the images into a single image
- Return merged image

The above algorithm takes care of both, low-light conditions as well as the motion of the subject.

[2] "Burst photography for high dynamic range and low-light imaging on ...." https://hdrplusdata.org/hdrplus.pdf. Accessed 24 Apr. 2019.

[3] "Google AI Blog: Night Sight: Seeing in the Dark on Pixel Phones." 14 Nov. 2018, http://ai.googleblog.com/2018/11/night-sight-seeing-in-dark-on-pixel.html. Accessed 24 Apr. 2019.

Proof of correctness of the proposed setup:

The steps involved are exactly the same as required for stitching two images. We will take the images from the front camera and rear camera and stitch them to form a single 360-degree panorama image. The steps as mentioned in here[4] can be used for this purpose.

Steps involved:

- Keypoint detection
    - We can SIFT/SURF algorithm
- Image registration
    - Use the RANSAC algorithm for this
- Perform image calibration:
    - Reducing optical defects such as distortions, chromatic aberrations in the image.
- Perform image alignment
    - Use homography matrix for this
- Compositing
    - Bringing all the images together
- Blending
    - Can be done using feathering

[4] "Image stitching - Wikipedia." https://en.wikipedia.org/wiki/Image_stitching. Accessed 24 Apr. 2019.

Algorithms Used:

SURF (speeded up robust features)[5]:

Involves three main steps:

- Interest Point Detection: Makes use of square-shaped filters as an approximation for Gaussian smoothing.
  - $S(x,y) = \sum_{i=0}^{x} \sum_{j=0}^{y} I(i,j)$
  - $\sigma_{approx} = current\ filter\ size\ x \left(\frac{base\ filter\ scale}{base\ filter\ size} \right)$
- Descriptor: Provides a unique and robust description of an image feature.
- Matching: Comparing descriptors obtained from different images, matching pairs can be found.

Image Alignment:

Used to align the image to match the viewpoint of another image.

- Projective transformation can be written as:

  $$x' = H.x$$

  where H is the homography matrix.

- Expressing x and x' using camera intrinsics (K and K'), we get

  $$x = K.[R\ t].X \text{ and } x' = K'.[R'\ t'].X'$$

- Now, using the above two equations:

  $$H = K' . R' . R^{-1} . K^{-1}$$

[5] "Speeded up robust features - Wikipedia." https://en.wikipedia.org/wiki/Speeded_up_robust_features. Accessed 24 Apr. 2019.

Question 1(b):

Proposed applications for the suggested setup:

1. 360-Groupfie app: This app would allow the user to take a 360-degree group selfie for when you don't want to leave anyone out. The suggested setup would be of great help as it would ensure a 360-degree area capture. So everyone will be present in the photograph, contributing to world peace.

2. GroupConf app: For when you want to turn your group meetings into a live video conference. With this app, the user can have group video calls with clients or with other groups within the company or simply have the entire family in one frame while on a video call with your loved ones. The 360-degree capture capability of the proposed setup helps here.

Question 2:

Problem Statement: Given a rough sketch and a textual description of a video scene, make an animation of the sketch according to the textual description. Reference: [6]

Aims/Objective: Generate a method for generating video depictions from a given text and initial static sketch.

Importance of problem statement: This is a significant step towards multi-modal deep learning models and will help in further advances.

[6] "Video Generation From Text." 1 Oct. 2017, https://arxiv.org/abs/1710.00421. Accessed 24 Apr. 2019.