Final Project Check-in: Dolly Zoom

Teammates: Beatrice Hoang and Jessica Wan

We have implemented the digital zoom proposed in **The "Vertigo Effect" on Your Smartphone: Dolly Zoom via Single Shot View Synthesis**

Our pipeline based on the paper is this:

- ☑ Given an image I, generate its depth map using the ZoeDepth depth estimation model
- ☑ Use equation (4) from the paper to digitally zoom the input image I and depth map D. We are zooming from the image's focal length (f_1) to a desired focal length (f_1^A). We are using this sample image taken with a Sony a7ii with a 35mm lens →


Because we know our camera's settings and focal length, we have chosen to hard code $f_1 = 35$. We will have to find a workaround for this (possibly camera calibration with a chessboard).

$$\boldsymbol{u}_1^A = (f_1^A/f_1)\boldsymbol{u}_1 + \left(1 - (f_1^A/f_1)\right)\boldsymbol{u}_0 \qquad (4)$$

Originally, the equation maps every pixel from I ($\mathbf{u_1}$) to a pixel position in the output image ($\mathbf{u_1}^A$) based on the distance between a pixel and the center of the image $\mathbf{u_0}$. However, implementing this created gaps in our image since some pixels in the output image did not have a corresponding input pixel. In order to fix this, we had the equation solve for $\mathbf{u_1}$:

$\mathbf{u_1} = (\mathbf{u_1}^A - \mathbf{u_0}) / k$ where $k = f_1^A/f_1$ (modified equation 4)

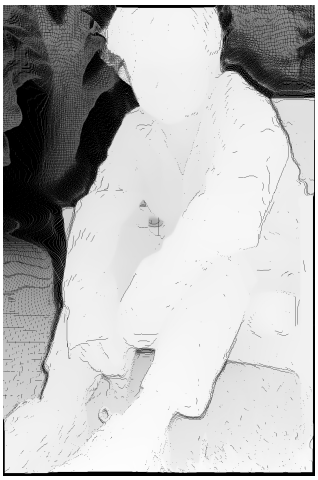
This ensures that the output image has a value. This equation is applied to the input image to return I_a, and to the depth map to return D_a.

Here is our sample output, changing the focal length from 35mm to 40mm →


- ☑ Use equation (3) to find the synthesized dolly zoom. Similar to equation (4), this equation is based on the position of pixels relative to the center of the image. This equation uses I_a and D_a

$$\boldsymbol{u}_1^B = \frac{D_1^A(D_0 - t)}{D_0(D_1^A - t)}\boldsymbol{u}_1^A + \frac{t(D_1^A - D_0)}{D_0(D_1^A - t)}\boldsymbol{u}_0 \,. \qquad (3)$$

Outputs I_1^{DZ} and D_1^{DZ} →

next, we will generate a second set of dolly-zoom images in order to fill in the gaps produced here.