ECE697 - Capstone Project

Dynamic Depth of Field with Eye Tracking

Neel Kelkar, Suchith Suresh, Nawal Dua
College of Engineering, University of Wisconsin - Madison, WI

Students: ndkelkar@wisc.edu, suchithsures@wisc.edu, ndua2@wisc.edu

Professor: matthew.malloy@wisc.edu

ABSTRACT

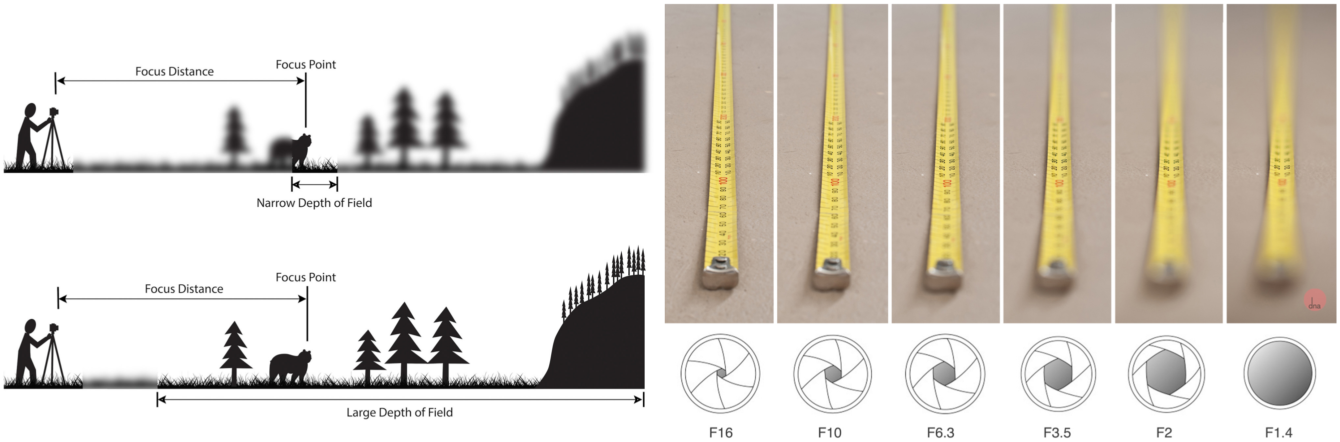
For a camera, Depth of Field(DoF) is the distance between the nearest and farthest object in focus. The depth of field in a video depends on numerous factors such as focal length, aperture, distance to subject, etc. The human eye too has a depth of field. DoF is generally extremely wide for objects more than a few meters away, meaning that everything is in focus at the same time.

For near objects, the DoF is very shallow, giving rise to the cinematic focus effect we see in movies. The DoF in cinematic videography is done using specialized lenses and scripted positioning of subjects. Mobile phone videos do not have this positioning planned beforehand and thus the cameras have a wide DoF as the default. Mobile phone camera's don't have readily available interchangeable lenses either, preventing daily users to use the optimal lens for the subject.

In this project, we intend to first generate depth maps from input video. Using these depth maps, we plan to create refocusable videos.

With these refocusable videos, we can simulate human vision but with a more cinematic DoF using eye tracking technology where the video will place more emphasis on whichever subject a viewer looks at.

If time permits, after we build the eye tracking DoF system, we plan to compare a few depth estimation methods and possibly build more applications using the depth maps.



(a) Narrow v/s Large DoF[4] (b) Aperture affecting DoF[2]

Figure 1. Representation of Depth of Field

BACKGROUND

The depth maps are generated using Google's "Depth from moving people" [3] open-source github code, this code provides us with a pre-trained model with options for future training if required. The first goal of the project will be to use the pre-trained model to generate the depth maps. The second goal time permitting would involve re-training the network for our application or improve the generated depth maps

Accurate calculations of Depth of Field (DoF) use complex equations which are excessive for our application, hence we will use simplified equations [1] to estimate the blur levels necessary in the out of focus regions of each frame.

$$b = d * m\frac{x_d}{D} \qquad \text{Equation 1.}$$

where,

b : blur

d : aperture value

m : magnification of the subject

x_d : relative distance between the subject(region of interest) and the background or foreground

D : distance between the foreground and background

The simplified equations assume paraxial approximation, hence we can use gaussian filters to simulate the expected blur.

As our focus is on the post-processing of stored videos for improved immersion, we can ignore the quantities representing aperture and magnification as they remain constant for the duration of the video, hence we get:

$$b \propto \frac{x_d}{D} \qquad \text{Equation 2.}$$

We obtain the relative depths of the objects (or regions), in a scene by using the generated depth maps[3]

BLUR GENERATION

Depth Levels

Instead of blurring based on the exact depth levels, we divide a frame into 10 levels of depth between the minimum and maximum values in the depth map. These depth levels are indexed between 0 and 9.

Blurring

We first select a focus point in an image. We take the depth level of this focus point and blur the depth levels before and after it using integer approximations of a Gaussian curve.
For example, for a point at a depth of 0.5, we use a Gaussian function with a mean of 0.5 and adjustable variance. We take 10 values of this Gaussian corresponding to the depth levels with it's peak at the depth level the point is in. After rounding up the values, we subtract this array from the maximum value in the array to get something like an upside down Gaussian curve. We then multiply each value by an even multiplier and add 1 to get an array of odd values. We use these values as the size of Gaussian blur kernels that we apply to the corresponding depth levels.
For a focus at 0.5, we generate the array [5, 3, 1, 1, 1, 3, 5, 5, 7, 7] with a multiplier of 2. By adjusting the multiplier we can adjust how strong the blur is and by adjusting the variance we can adjust how wide the depth of field will be.

EXPECTED WORK TO BE COMPLETED

1. Find a monocular depth estimation method.
   (a) Check the performance of Google's "Depth from moving people" [3] using the github repository provided. ✓
2. Use the depth map to generate refocusable video.
   (a) Try different Gaussian blurring methods to simulate desired DoF ✓
3. Develop a system to refocus a pre-recorded video real-time.
   (a) Develop a system to focus on any selected subject real-time.✓
4. Use eye-tracking technology with this system to simulate human vision.
   (a) Simulate eye tracking using mouse pointer tracking for the system ✓
   (b) Move the pointer over the video by tracking where a user is looking at on the screen.
5. Survey performance of different depth estimation methods.

Challenges

- Depth maps generated by pre-trained models are not suitable
  - Solution 1: Modify the parameters of the pre-trained model
  - Solution 2: Process the generated depth maps to address the issue ✓
  - Improvement 1: Re-train the model for our application
  - Improvement 2: Try to implement improved Depth Map generation algorithms
- Real time blurring is not possible
  - Solution 1: Process the stored videos and store the data for each possible blur level for each frame ✓
- Eye-tracking hardware is unavailable
  - Solution 1: Exclusive Eye-Tracking hardware based eye-tracking replaced with webcam based eye-tracking (Webcam based solution is not accurate and performance will suffer)

WORK DONE

Figure 2 provides a visual representation of the work completed

1. We are using the pretrained models provided by Google to generate depth maps for videos.
2. We have a system for refocusing at any selected depth using multiple Gaussian masks.
3. We have figured out how to apply different filters onto real time live video.
4. We are currently in the process of figuring out GPU acceleration or reducing the computation time of the blurring.
5. To achieve a smooth video output for eye tracking we are now pre-processing the blur values required for the stored video and storing the data locally.

DATASETS

1. MannequinChallenge - A Dataset of videos with still people and a moving monocular camera. This is the dataset that Google used to train the DOF we will be using.
2. TUM RGBD Dataset - Dataset containing RGB-D data and ground-truth data
3. DAVIS dataset - Densely Annotated Video Segmentation. This is the dataset we will be running Google's model on and creating our prototype on. When the model is run on a video from this data set. This dataset will almost function as test dataset because we will be using the model on this dataset to see how our algorithm works.

The datasets will primarily be used for testing the proposed algorithm and will be used for re-training if required

RESOURCES NEEDED

1. Euler Cluster - If we need to train our own model or do high level computation. It is likely we will have to use this.
2. Gaming Desktop with capable CPU and GPU. This meets all our computation needs if we use a pretrained model.
3. Eye tracking device - Currently we are looking at Tobii Eye Tracker 5 for $229. Tobii makes state of the art eye tracking devices for multiple purposes such as gaming and scientific research. It has head and eye-tracking capabilities for screens up to 27" big.

(a) Original image

(b) Distant, shallow DoF

(c) Near, shallow DoF

(d) Distant, shallow DoF w/Mouse Tracking

(e) Near, shallow DoF w/Mouse Tracking

Figure 2. Work Done

REFERENCES

[1] Jeff Conrad. "Depth of Field in Depth". In: 2004, pp. 12–13. URL: https://www.largeformatphotography.info/articles/DoFinDepth.pdf.

[2] Bill Dobbins. "THE MYTH OF DEPTH OF FIELD". In: 2019. URL: https://onphotography.me/2019/02/17/the-myth-of-depth-of-field.

[3] Bill Freeman et al. "Learning the Depths of Moving People by Watching Frozen People". In: 2019.

[4] Laviolette. "Depth of Field". In: 2017. URL: https://sites.google.com/a/ocsb.ca/teh-2016-awq2or-3/calendar/week-9-and-10-exposure.