# LearnAF Dataset

Last updated: 11/02/2020

## Overview

Data is captured using a custom rig (shown below) that captures images of a scene using 5 phone cameras simultaneously. Each camera captures a 49 image focal sweep, in which the focus distance of each image is changed. We share the RGB images from all 5 cameras, their computed pose, the dual-pixel data and the depth maps.

## Organization

Data is split into “test” and “train” folders. There are 351 focal sweeps in “train” and 47 focal sweeps in “test”. Each of these has the following subdirectories:

- *scaled_images/<capture_name>/<focal_slice_idx>/result_scaled_image_{center|left|right|top|bottom}.jpg*: Five RGB images from the 5 phones scaled down to 378x504 resolution.
- *scaled_camera_pose/<capture_name>/<focal_slice_idx>/result_scaled_camera_pose_*.txt*: Camera poses corresponding to the 5 images above. See Camera Pose Description for details.
- raw_left_pd/*<capture_name>/<focal_slice_idx>*/result_pd_left_*.png: 16 bit PNG images corresponding to the raw left PD data captured from the sensor of size 756 x 2016. This data has a white level of 1023, so all pixels will be in the range [0, 1023]. This also corresponds to the data captured by our app, but in the app the raw pixels have been multiplied by 65535 / 1023, so that they use the entire range of 16 bit values.
- raw_up_left_pd/*<capture_name>/<focal_slice_idx>*/result_up_pd_left*.png: Left dual-pixel image of size 1512x2016 resized to have the same aspect ratio as the RGB image.
- raw_right_pd/*<capture_name>/<focal_slice_idx>*/result_pd_right_*.png: 16 bit PNG images corresponding to the raw right PD data captured from the sensor.

- raw_up_right_pd/<*capture_name*>/<*focal_slice_idx*>/result_up_pd_right*.png: Right dual-pixel image of size 1512x2016 resized to have the same aspect ratio as the RGB image.
- *merged_depth/<capture_name>/result_merged_depth_*.png*: Depth map corresponding to the five images. See Depth Map Description to understand how the depth is stored.
- *merged_conf/<capture_name>/result_merged_conf_*.exr* : Confidence of the depth maps above in the range [0, 1]. The test dataset also contains these as binary numpy arrays.

# Camera Pose Description

## Camera Model

A camera transforms a homogeneous 3D point ***X*** into a point ***x*** on its image plane. The mapping can be divided into the following 3 steps:

1. Transform the 3D point ***X*** into a ray ***u*** in the local coordinate system

```
u = R * (X - c)
```

2. Project ***u*** to ***v*** in the normalized image coordinate system.

```
v = u / u[2]
```

3. Map ***v*** to ***x*** in the pixel coordinate system.

```
r = v[0]^2 + v[1]^2

distortion_factor = 1 + k1 * r + k2 * r * r + k3 * r * r * r;

xd = v[0] * distortion_factor
yd = v[1] * distortion_factor

x = [f * xd + s * yd + px; f * a * yd + py]
```

where
***X*** : Homogeneous point.
***c*** : Position of the camera.
***R*** : Orientation of the camera.
*f* : Focal length.
*s* : Skew.
*a* : Pixel aspect ratio (a = scale_factor_y / scale_factor_x).
*px, py* : Principal point.

*k1, k2, k3* : Radial distortion coefficients.
***x*** : Image of X.

## Camera File

A typical provided camera pose file is as follows:

```
position: -0.034968384543875804
position: 0.00015154158334801444
position: 0.010699204117784251
orientation: 0.0033926260744923686
orientation: 0.0075903968782610378
orientation: -0.012258027817996359
focal_length: 860.3425758403057
pixel_aspect_ratio: 1
principal_point: 378.94709474340851
principal_point: 505.62880599068296
radial_distortion: 0.026503988319729721
radial_distortion: 0.0079358472962738225
radial_distortion: 0
skew: 0
size_x: 756
size_y: 1008
projection_type: PERSPECTIVE
```

where:

position: Position of the camera (***c***)
orientation: Orientation (***R***) of the camera expressed in Angle-Axis notation.
focal_length: Focal length (*f*) in pixels .
pixel_aspect_ratio: Pixel aspect ratio (*a*). Always 1.
principal_point : Principal point location (*px, py*) in pixels.
radial_distortion: Parameters for radial distortion (*k1, k2, k3*).
skew: Skew (s). Always 0.
size_x: Image width.
size_y: Image height.
projection_type: Type of camera model. Always PERSPECTIVE.

# Depth Map Description

Depth maps are obtained by inverse perspective sampling of depth in the range [0.2, 100] meters.

To convert a given depth map to inverse perspective depth in the range [0, 1]:

```
depth_map = depth_map / 255.0
```

To further convert to depth in meters:

```
(max * min) / (max - (max - min) * depth_map)
```

where max = 100.0 and min = 0.2

Even though the depth can be converted into meters, it's not guaranteed to be metric correct since the absolute scale of the reconstruction is unknown.

# Focus Distances

| Slice Index | Focus Distance (mm) | Slice Index | Focus Distance (mm) | Slice Index | Focus Distance (mm) |
|---|---|---|---|---|---|
| slice_00 | 3910.92 | slice_17 | 274.13 | slice_34 | 142.35 |
| slice_01 | 2289.27 | slice_18 | 261.53 | slice_35 | 138.98 |
| slice_02 | 1508.71 | slice_19 | 247.35 | slice_36 | 134.99 |
| slice_03 | 1185.83 | slice_20 | 237.08 | slice_37 | 131.23 |
| slice_04 | 935.91 | slice_21 | 225.41 | slice_38 | 127.69 |
| slice_05 | 801.09 | slice_22 | 216.88 | slice_39 | 124.99 |
| slice_06 | 700.37 | slice_23 | 207.10 | slice_40 | 121.77 |
| slice_07 | 605.39 | slice_24 | 198.18 | slice_41 | 118.73 |
| slice_08 | 546.23 | slice_25 | 191.60 | slice_42 | 116.40 |
| slice_09 | 486.87 | slice_26 | 183.96 | slice_43 | 113.63 |
| slice_10 | 447.99 | slice_27 | 178.29 | slice_44 | 110.99 |
| slice_11 | 407.40 | slice_28 | 171.69 | slice_45 | 108.47 |
| slice_12 | 379.91 | slice_29 | 165.57 | slice_46 | 106.54 |

| | | | | | |
|---|---|---|---|---|---|
| slice_13 | 350.41 | slice_30 | 160.99 | slice_47 | 104.23 |
| slice_14 | 329.95 | slice_31 | 155.61 | slice_48 | 102.01 |
| slice_15 | 307.54 | slice_32 | 150.59 | | |
| slice_16 | 291.72 | slice_33 | 146.81 | | |