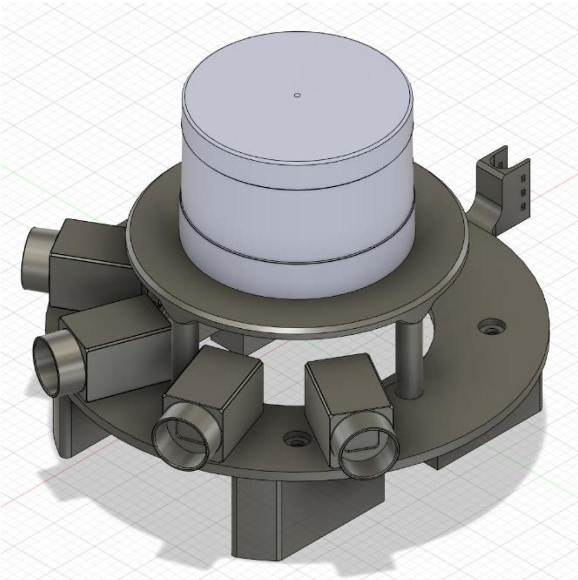
# Introduction to the ELTE 3D sensor pack

# (Version 2.2*)

The currently used version of the sensor pack consists of several cameras, a LiDAR device and some miscellaneous sensors. The cameras and the LiDAR are mounted on a 3D printed fixture, so their positions are constrained to each other.

The fixture allows the sensor pack to be mounted on multiple vehicles, mostly used on the roof rack of the car, but it can be mounted to a go-kart for indoor use, or a shopping cart, for manual movement. The fixture in its current form allows the use of five cameras as shown below.

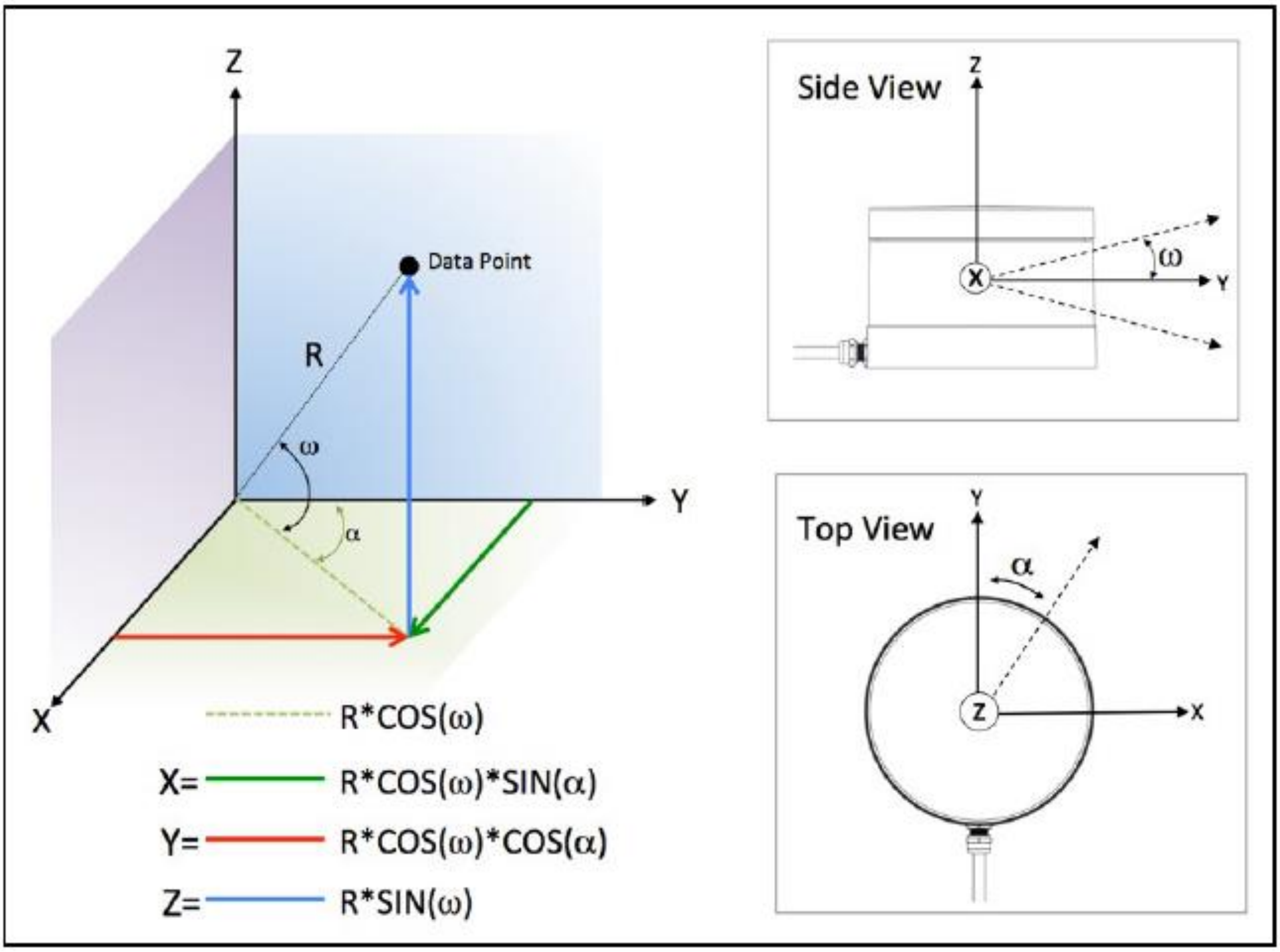



The optics for the cameras are not shown in the Computer Aided Design (CAD). In the current configuration four cameras are facing forward, 20°and 60° angled relative to the forward axis of the car. These four cameras are equipped with a “normal” optic (so it can be approximated using the pinhole model).

The cameras currently in use are HikVision/HikRobot MV-CA020-20GC, the exact documentation can be found here. Calibration parameters and information about the optics can be found later in this document.

**: V2.2 is the latest iteration as of 2023.05.29, but there is constant development for the software and mounting mechanism*

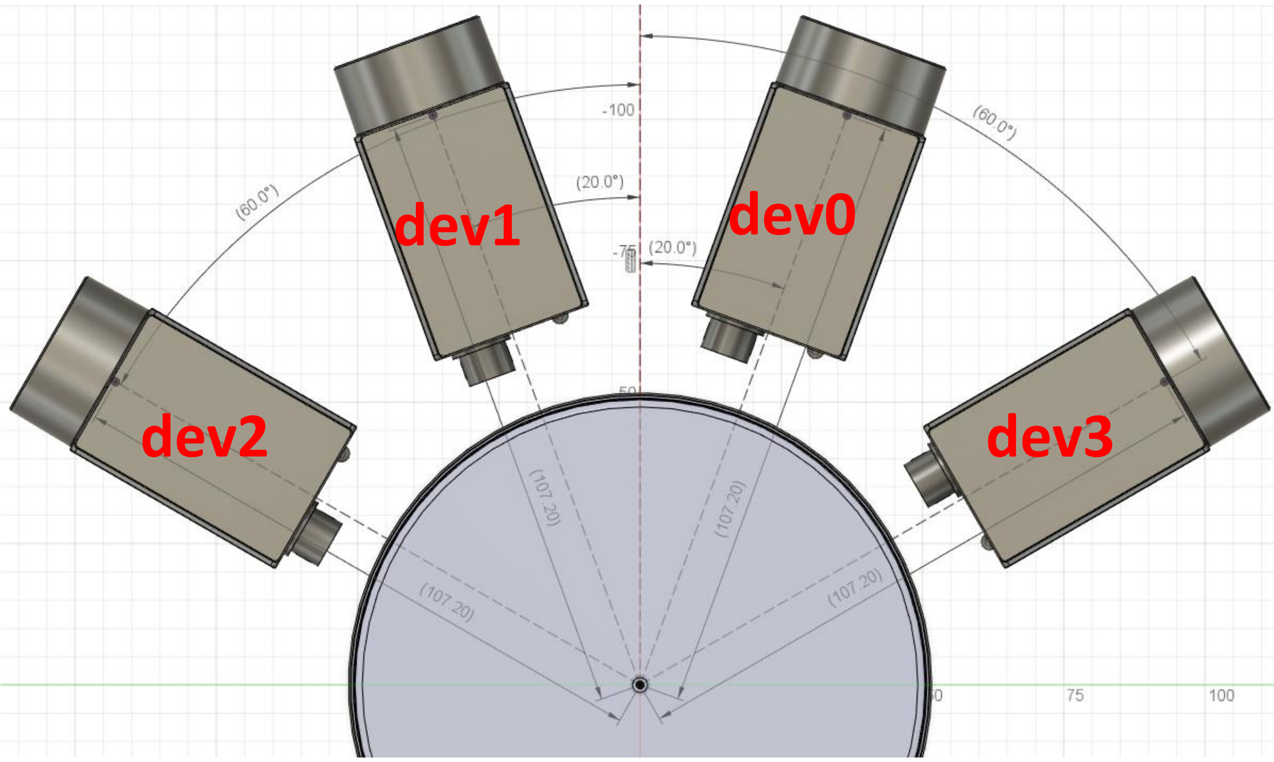
The LiDAR is mounted on top of the main 3D printed carrier. The LiDAR used here is a Velodyne VLP-16, (later renamed as PUCK). It uses 16 channels to scan in the vertical direction, with a vertical FOV of 30° (and horizontal FOV of 360°). For reading the exact parameters and documentation of the LiDAR click here.

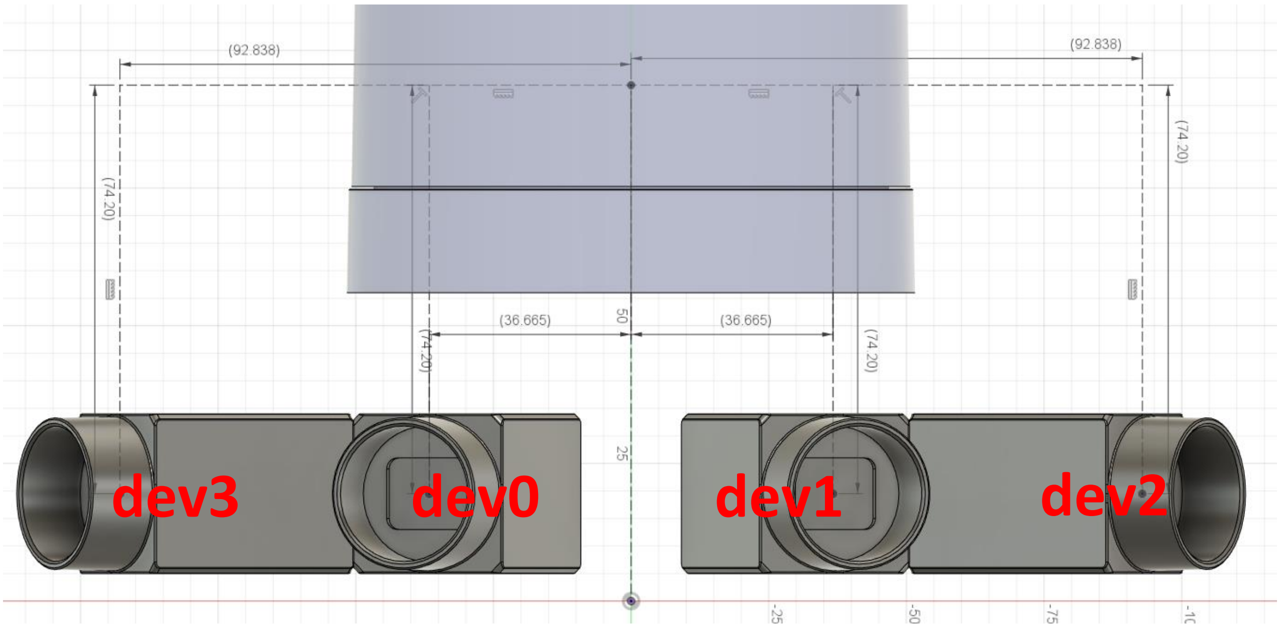



In most of our measurements we use the LiDAR's coordinate system as a reference (world coordinate system). It uses a standard right-handed coordinate system, as shown above. The data coming off the LiDAR uses a polar coordinate system, with a rotation angle for the horizontal and vertical direction, and a distance measurement for the exact position. Our software converts this to XYZIII coordinates, when using the .xyz file format. The I in the XYZIII format stands for intensity, and it is the measured reflectivity of the given point. It is the same number 3 times, to allow some compatibility with XYZRGB point cloud parsers. **In the competition, all the published point clouds are in this .xyz format.**

The sensor pack also contains a GPS module connected to the LiDAR, and an IMU. The LiDAR uses the GPS module only for time synchronization, but the full data coming from the GPS is available to log. Our software logs the latest GPS position for every saved point cloud. The GPS module used is an Ardusimple RTK2B which uses an U-Blox ZED-F9P series GNSS receiver chipset, with active RTK corrections, to provide high accuracy positioning. In our test the positioning is precise to 50-20cm when moving, 10-2cm when stationary. **The position data from the GPS module will be recorded for every dataset, and will be used to evaluate the solutions given in the competition, and it will be included in *some* datasets to allow participants to test their solutions**

To reference each camera's position and coordinate systems, there is a mounting diagram. Each measurement shown is relative to the optical centre of the LiDAR, which is the 0,0,0 coordinate is the LiDAR's coordinate system, **<u>all values are given in mm.</u>**



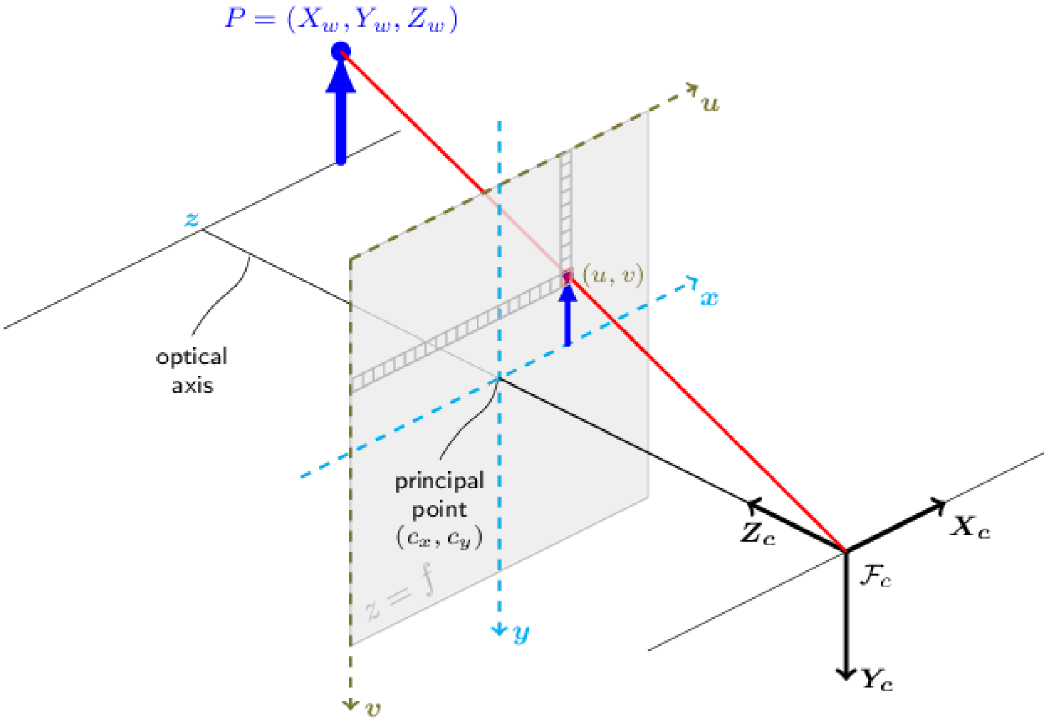




Each of the available camera's coordinates in the LiDAR (world) coordinate system is in the table below.

XYZ is the coordinates relative to the LiDAR system D is the absolute distance, <u>all values are given in mm</u>.

| | X | Y | Z | D |
|---|---|---|---|---|
| dev0 | 36,665 | 100,735 | -74,2 | 130,374 |
| dev1 | -36,665 | 100,735 | -74,2 | 130,374 |
| dev2 | -92,383 | 53,6 | -74,2 | 130,374 |
| dev3 | 92,838 | 53,6 | -74,2 | 130,374 |

The points measured on the cameras in the above diagram is the optical centre of the cameras, the middle point of the sensor plane, which is **NOT** the same as the focus point as the focal length is 7mm.




The coordinate system used on the cameras are the OpenCV camera coordinate system. It is originated at the focal point of the camera, and uses a left-handed coordinate system, as shown above. The illustration is taken from the OpenCV documentation, for more information, look at the official OpenCV documentation here. Because of the differences in the systems, a coordinate conversion is needed.

In the current version of the software and hardware, the cameras are limited in framerate, because our PC cannot handle all the data coming from the cameras, while still maintaining time sync. The cameras are time synchronized via an external trigger source connected to them. The speed of the trigger source determines the framerate. For every full frame that comes from the camera, the latest full rotation of the LiDAR is logged, along with the latest GPS position. **All of the data published for the competition is recorded at 4FPS on the cameras and 1200RPM (20RPS) on the LiDAR, and the resolution of the cameras are set to full, which is 1920×1200px.** The camera has a resolution of 1920×1200px, and a pixel size of 4.8×4.8µm.

## Camera-Lidar calibration

### Camera intrinsic parameters

As it is well-known in computer vision, the intrinsic parameters of digital cameras are usually represented by an upper triangular matrix

$$K = \begin{bmatrix} kf & 0 & u_0 \\ 0 & kf & v_0 \\ 0 & 0 & 1 \end{bmatrix},$$

where $f$ is the focal length, $k$ is the pixel size and $[u_0, v_0]^T$ is the principal point.

If there is a spatial point $P_C = [X_C, Y_C, Z_C]$ given in the camera coordinate system, the projection is given by

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \sim \begin{bmatrix} kf & 0 & u_0 \\ 0 & kf & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_C \\ Y_c \\ Z_c \end{bmatrix}$$

where $\sim$ denotes equality up to scale operator, and the projected coordinates $[u, v]$ can be obtained by a homogeneous division.

For the applied optics, the focal length is 6mm. The pixel size is 4.8×4.8µm. Therefore, $k = \frac{1}{4.8 \cdot 10^{-6}} pixel/m$. Then

$$\mathrm{kf} = \frac{6 \cdot 10^{-3}}{4{,}8 \cdot 10^{-6}} = 1250$$

Moreover, as we use high-quality Fujinon SV-0614H optics, the principal point is at the midpoint of the image: $[u_0, v_0] = [960{,}600]$.

Thus, the intrinsic parameter matrix is as follows:

$$K = \begin{bmatrix} 1250 & 0 & 960 \\ 0 & 1250 & 600 \\ 0 & 0 & 1 \end{bmatrix}$$

### LiDAR-Camera and Camera-Camera poses

In our system, the world coordinates are fixed to the LiDAR device. The rigid transformation between the LiDAR and the cameras or a camera and the other cameras are represented by an orthonormal matrix $R$ and a translation vector $t$. If a 3D location is given in the world as $P_L = [X_L, Y_L, Z_L]$, the corresponding location in the camera system can be given as:

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = RP_L + t = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} X_L \\ Y_L \\ Z_L \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}.$$

**For the competition, the $[\boldsymbol{R}|\boldsymbol{t}]$ matrices are given in the calibration (text) file:**

$$[\boldsymbol{R}|\boldsymbol{t}] = \begin{bmatrix} \boldsymbol{r_{11}} & \boldsymbol{r_{12}} & \boldsymbol{r_{13}} & \boldsymbol{t_x} \\ \boldsymbol{r_{21}} & \boldsymbol{r_{22}} & \boldsymbol{r_{23}} & \boldsymbol{t_y} \\ \boldsymbol{r_{31}} & \boldsymbol{r_{32}} & \boldsymbol{r_{33}} & \boldsymbol{t_z} \end{bmatrix}.$$

**Remark that there are three different $[\boldsymbol{R}|\boldsymbol{t}]$ matrices for the four cameras, they are denoted by $\boldsymbol{Rt1}$ , $\boldsymbol{Rt2}$, $\boldsymbol{Rt3}$ and $\boldsymbol{Rt4}$ in the calibration file.**

## Calibration data

All the sensors are calibrated by MATLAB toolboxes, if you would like to interpret the values, read the related documentation(s).