POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

IACV Homework - 2024/25

Report for the "Image Analysis and Computer Vision" homework

Author: **Paolo Ginefra**

GitHub Repository: **PaoloGinefra/I-ACV_Homework_2024**

Person code: 10765882
Advisor: Prof. Vincenzo Caglioti
Academic Year: 2024-25

Contents

Introduction & problem definition

This is the report of Paolo Ginefra's solution to the "Image Analysis and Computer Vision" homework 2024/25. All the referenced code is available in the GitHub Repository.

0.1. Problem definition

In this section, the problem at hand will be described in detail referencing the "Homework Assignment 2024-25" document.

0.1.1. Scene description

A piece of furniture is a rectangular parallelepiped, whose width (along the X-axis) is $l = 1$. The other dimensions, namely the depth m along the Y axis and the height h along the Z-axis are unknown. In addition, a horizontal circumference (i.e., parallel to the X-Y plane) is visible. Furthermore, an unknown horizontal planar curve is also visible, placed at midheight $h/2$.

Figure 1: The Scene Description

0.1.2. Image Description

A single image is taken of the above rectangular parallelepiped by an uncalibrated, zero skew, camera. (Its calibration matrix K depends on four unknown parameters, namely f_x, f_y and the two pixel coordinates U_O, V_O of the principal point). A set of lines parallel to X-axis are visible, and their images l_1, l_2 and l_3 are extracted; a set of lines parallel to the Y-axis are visible and their images m_1, m_2, m_3, m_4, m_5 and m_6 are extracted; a set of vertical lines (i.e., parallel to the Z axis) are also visible and their images h_1, h_2, h_3 and h_4 are extracted. In addition, both the image C of the circumference and the image S of the unknown horizontal curve are also extracted.

Figure 2: The Scene Description

0.1.3. Part1 - Theory

1. From the l_i and m_i lines, find the vanishing line l'_∞ of the horizontal plane.
2. Using the results of the previous point, find a (Euclidean) rectification mapping H_R for a horizontal plane (e.g., the lower horizontal face of the parallelepiped), and compute the depth m of the parallelepiped.
3. From the results of the previous points, use the lines h_i to find the calibration matrix K.
4. Using the results of the previous points, determine the height h of the parallelepiped.
5. Using S and the results of previous points, compute the X-Y coordinates of a dozen points

(at your choice) of the unknown horizontal curve.

6. Using K, localize the camera with respect to the parallelepiped.

0.1.4. Part2 - Matlab

1. Consider the image "Look-outCat.png". Using feature extraction techniques (including those already implemented in Matlab) plus possible manual intervention, extract the images of useful lines and both the image C, of the circumference and the image S of the other planar curve.

2. Write a Matlab program that implements the solutions to problems $1-6$ and show the obtained results.

3. Plot the rectified curve S and show different views of the recovered 3D model of the rectangular parallelepiped.

0.2. Document structure

Since the Theory and the Matlab code have a significant overlap, they are presented in parallel following the numbering of the Matlab part. The $n-th$ Theory task takes the numbering $2.n$. For each part, the actual code is not reported since it is easily accessible and documented in the GitHub Repository.

1 | Feature Extraction

1.1. Line extraction

In order to extract the most significant lines:

1. Crop the image to the desired portion

Figure 1.1: The image cropped

2. Convert to HSV color space.
3. Perform the SVD on the RGB channels plus the Saturation and Value.

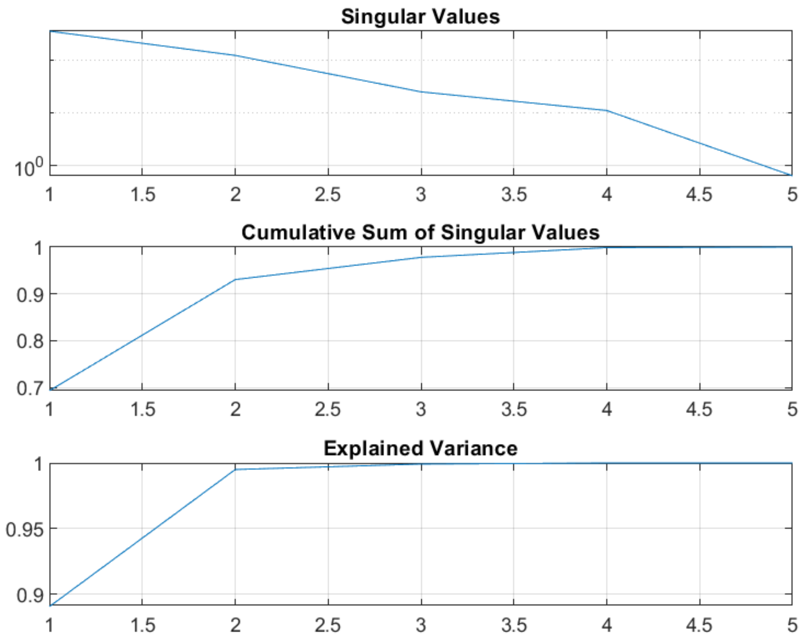

Figure 1.2: The singular values distribution

4. Compute the Principal Components

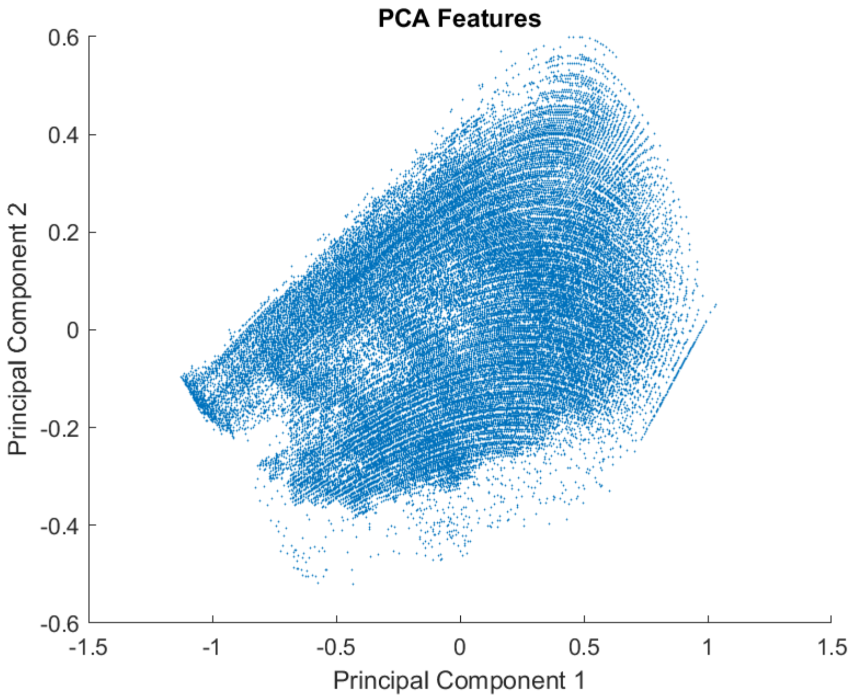

Figure 1.3: The first 2 principal components

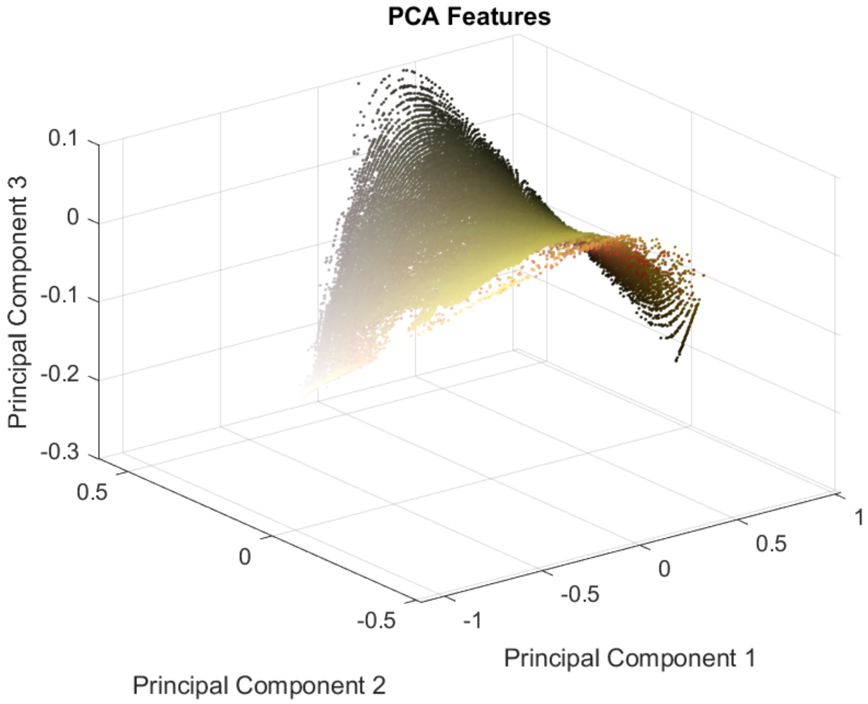

Figure 1.4: The first 3 principal components

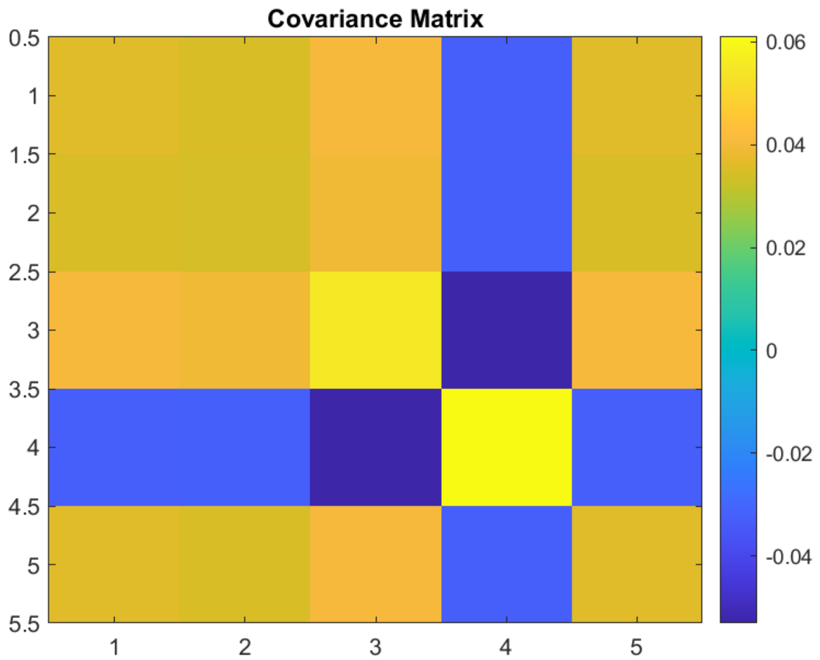

Figure 1.5: The Covariance matrix between principal components

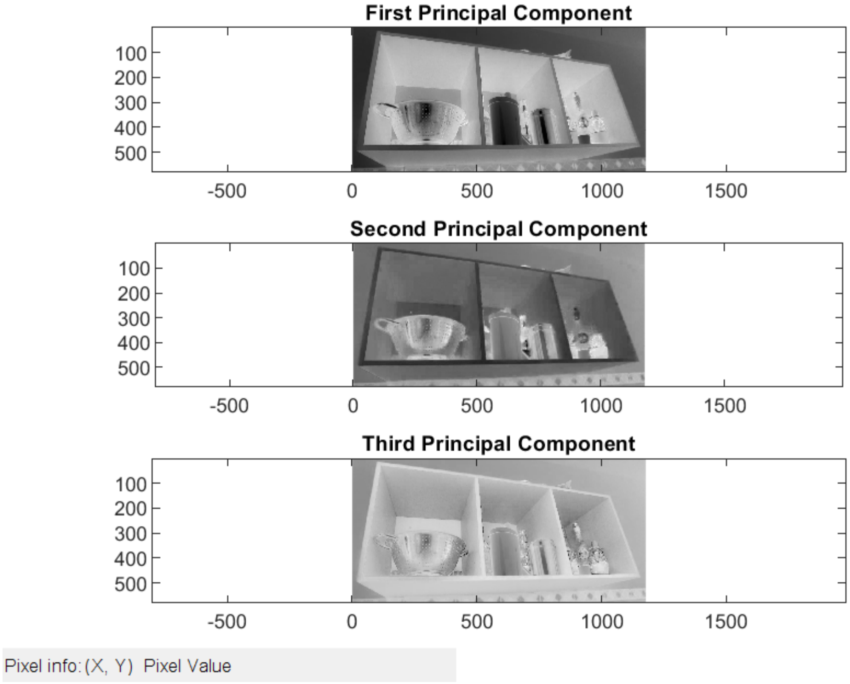

Figure 1.6: The first three principal components in image form

5. Apply Canny Edge detection to the first principal component

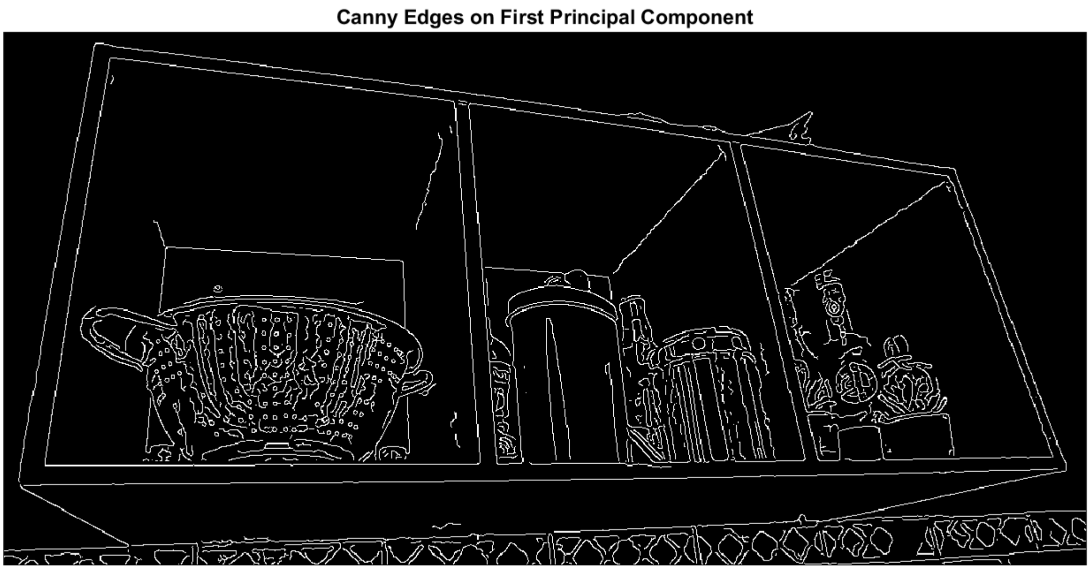

Figure 1.7: Canny edge detection on the first principal component

6. Remove edges that are too crowded, like the holes of the colander. To do that, first heavily blur the image then threshold the image to select the sparser areas. Use this to mask the edges.

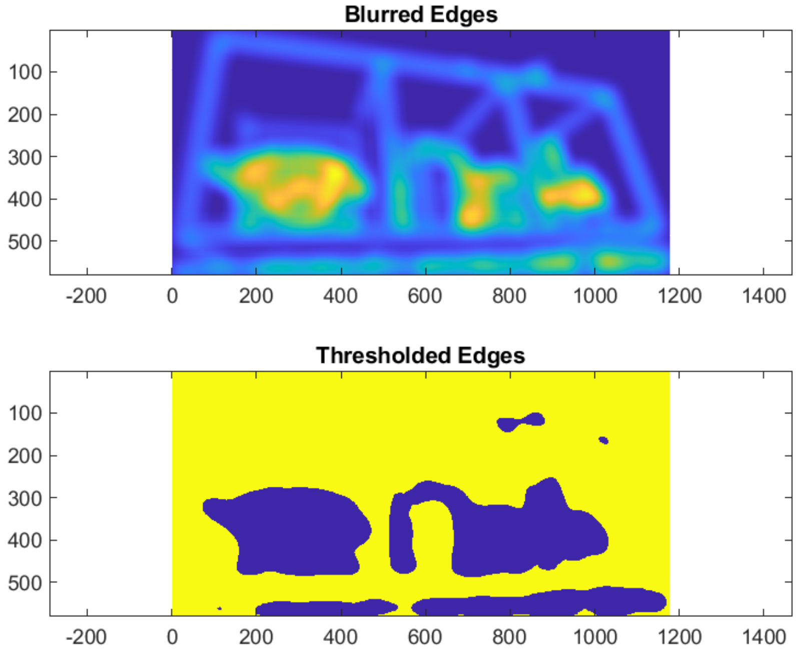

Figure 1.8: Create the edge mask

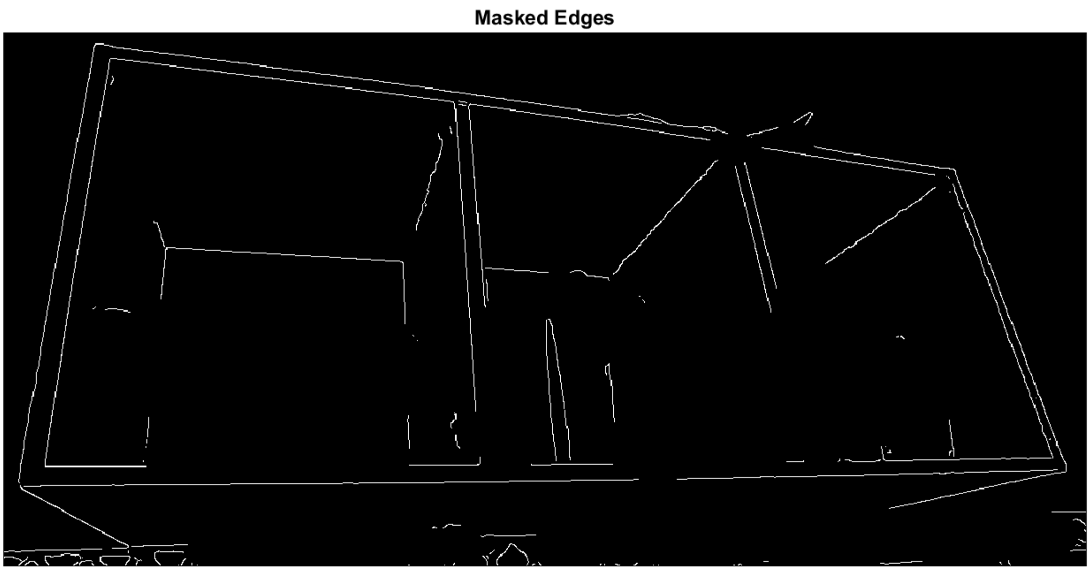

Figure 1.9: Canny edge detection on the first principal component

7. Apply the Hough Transform to the masked edges

Figure 1.10: Hough Transform result

1.2. Conic Extraction

In order to extract the conic C:

1. Crop the image to the desired portion

Figure 1.11: The image cropped

2. Convert to gray scale

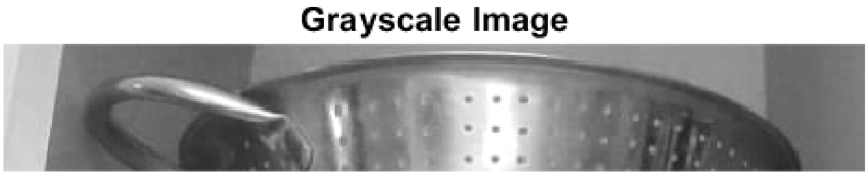

Figure 1.12: The image in grey scale

3. Canny edge detection ($\sigma = 2$, threshold $= [0, 0.15]$)

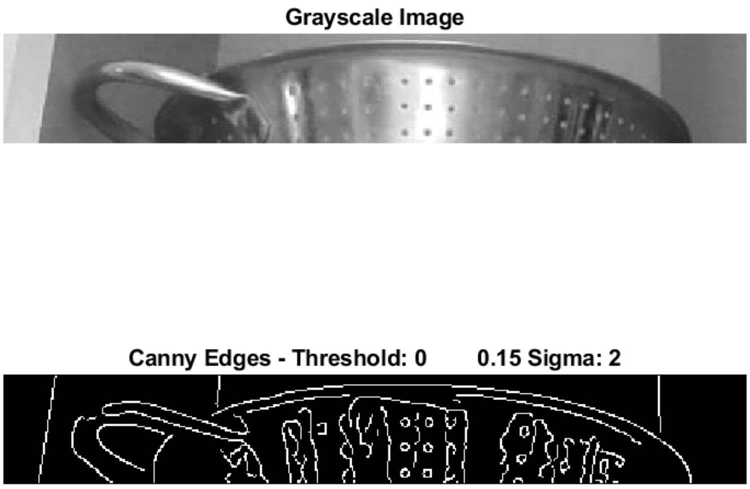

Figure 1.13: Canny Edge detection

4. Remove edges that are too crowded, like the holes of the colander. To do that, first heavily blur the image ($\sigma = 10$) then threshold the image to select the sparser areas (≤ 0.1). Use this to mask the edges

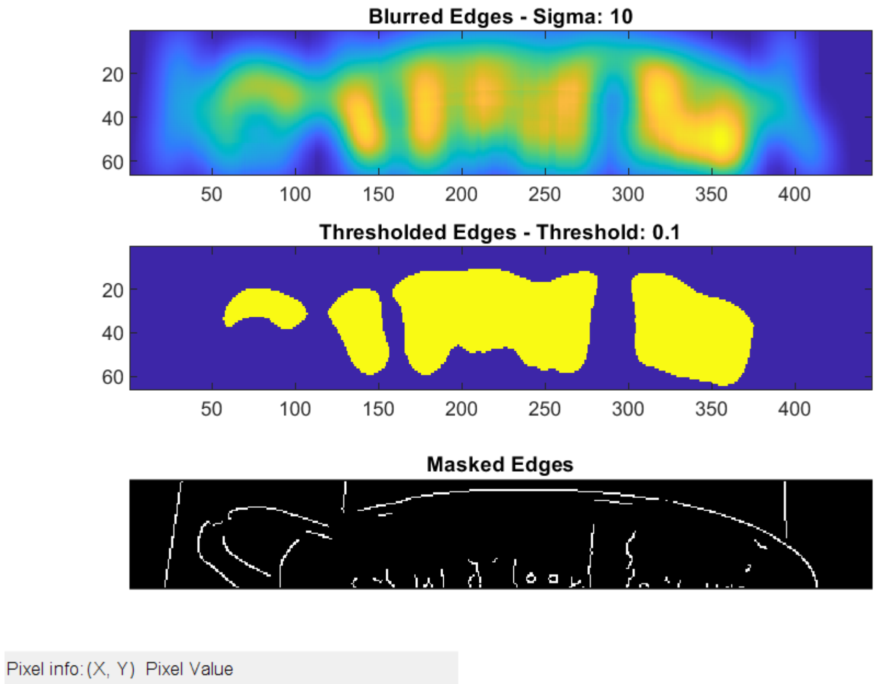

Figure 1.14: Masking the edges

5. Extract the remaining points $\neq 0$ and use a slightly modified RANSAC to extract the Conic with the most inliers

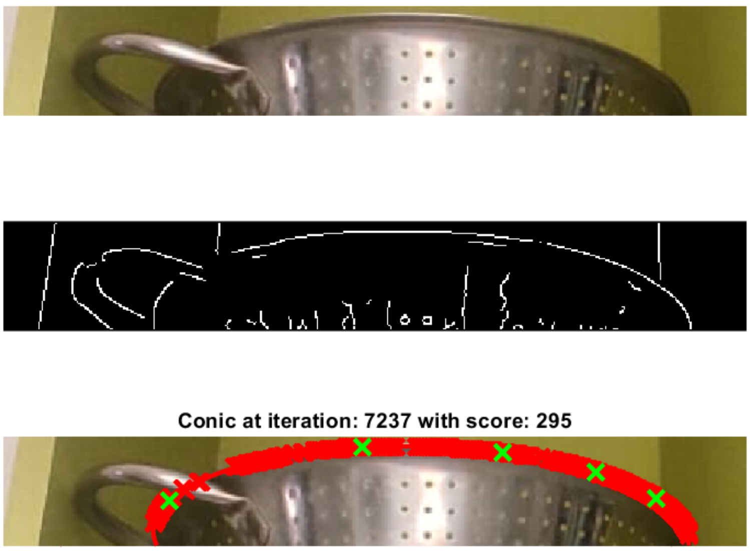

Figure 1.15: Conic RANSAC

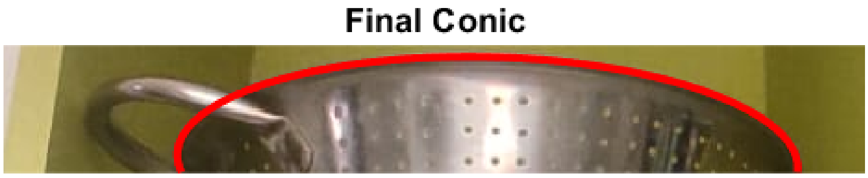

Figure 1.16: Final Conic

1.3. Extraction of S curve

In order to extract the curve S:

1. Crop the image to the desired portion
2. Convert to grayscale
3. Canny edge detection ($\sigma = 1$, threshold $= [0.2, 0.8]$)

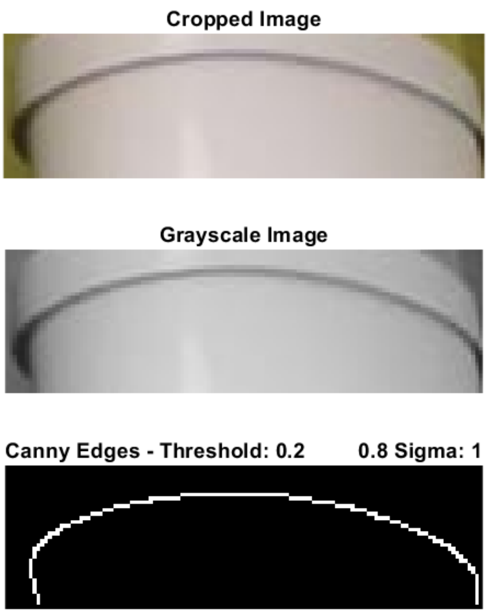

Figure 1.17: The S extraction process

2 | Scene Geometry Reconstruction

2.1. Finding the vanishing line l'_∞ of the horizontal plane

Procedure 1: Robust finding of the intersection between multiple intersecting lines

his procedure aims at finding the intersection point P of a given set of $n \geq 2$ intersecting lines $l_i\, \forall i \in \{1, ..., n\}$. Both the point P and the lines l_i are provided in homogenous coordinates. Thus

$$\underline{P} \triangleq \begin{bmatrix} P_x \\ P_y \\ 1 \end{bmatrix}$$

and

$$\underline{l_i} \triangleq \begin{bmatrix} l_{ix} \\ l_{iy} \\ 1 \end{bmatrix} \forall i \in \{1, ..., n\}$$

. The matrix L is defined as

$$L = \begin{bmatrix} \underline{l_1} & \underline{l_2} & \dots & \underline{l_n} \end{bmatrix}$$

Were all the lines perfectly intersecting in a single point it would hold

$$L^T \underline{P} = \underline{0}$$

and thus

$$\underline{P} \in RNS(L)$$

Unfortunately, in the real world, seldom do n "intersecting" lines actually intersect. An approximated solution is needed and thus the problem becomes an optimization one:

$$\underline{P} \triangleq \underset{\underline{x}}{\operatorname{argmin}} \, ||L^T \underline{x}||_2$$

This problem can be solved using the Singular Value Decomposition (SVD) on the L

matrix:

$$
\begin{gathered}
L = U \Sigma V^T \\
\text{where} \\
U, \Sigma, V \in \mathbb{R}^{3x3}, \\
U \cdot U^T = I, \\
V \cdot V^T = I, \\
\Sigma_{ii} = \sigma_i \in \mathbb{R} \; \forall i \in \{1, ..., 3\}, \\
\Sigma_{ij} = 0 \; \forall i \in \{1, ..., 3\}, \forall j \in \{1, ..., 3\}, i \neq j, \\
|\sigma_i| >= |\sigma_j| \; \forall i, j \in \{1, ..., 3\}, i < j
\end{gathered}
\tag{2.1}
$$

Since the σ_i are in absolute nonincreasing order it can be easily derived that the optimal solution up to scale is:

$$\underline{P} \triangleq V_n := \text{n-th column of V}$$

The image of the vanishing line l'_∞ of the horizontal plane can be found as the line going through the images of two distinct vanishing points of the horizontal plane. Luckily, the vanishing point associated with a direction in the horizontal plane is shared with all parallel planes, and thus the ms and ls lines will all meet respectively in the vanishing points v_m and v_l.

Using Procedure 1 the best approximation for v_m and v_l can be computed.

The line l'_∞ is now defined as :

$$
\begin{aligned}
{l'_\infty}^T v_m = 0 \\
{l'_\infty}^T v_l = 0
\end{aligned}
$$

and thus:

$$l'_\infty \triangleq v_m \times v_l$$

The results of the extraction are the following:

$$l'_\infty \triangleq \begin{bmatrix} -8.0288 \cdot 10^{-05} \\ -0.0011 \\ 1 \end{bmatrix}$$

Figure 2.1: The Extracted image of the vanishing line

2.2. Metric Rectification and depth estimation

To rectify the upper face of the cabinet a stratified approach has been chosen. The rectification is the outcome of three steps:

1. Affine Rectification
2. Affinity to make C a circle
3. Affinity to align the axis and allign the plot (optional)

2.2.1. Affine Rectification

The purpose of the affine rectification is to find the Homography H_{aff} such that

$$H_{aff}^{-T} \cdot l'_{\infty} = l_{\infty} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Consequently H_{aff} has to be of this form:

$$H_{aff} = \begin{bmatrix} * & * & * \\ * & * & * \\ & l_{\infty}'^{T} & \end{bmatrix}$$

In order to make the affine reconstructed image of a more manageable size the following homography has been chosen:

$$H_{aff} = \begin{bmatrix} \frac{1}{10} & 0 & 0 \\ 0 & \frac{1}{10} & 0 \\ -8.0288 \cdot 10^{-05} & -0.0011 & 1 \end{bmatrix}$$

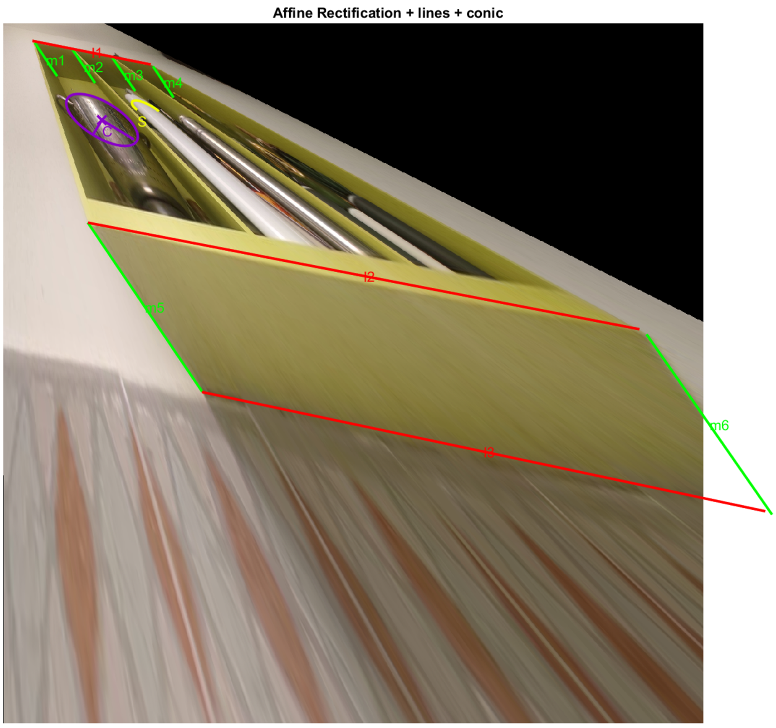

Figure 2.2: The result of the affine rectification

2.2.2. Affinity to make C a circle

Given the extracted conic C in homogeneous matrix form, one can apply the affine rectification obtaining:

$$C_{aff} = H_{aff}^{-T} \cdot C \cdot H_{aff}^{-1}$$

From that the length (a and b) and direction (θ) of the ellipse' axis as well as the ellipse center's coordinates (C_c) can be extracted:

$$
\begin{aligned}
a &= 74.0106 \text{ pixels} \\
b &= 31.8570 \text{ pixels} \\
\theta &= 0.5529 \text{ rad} \\
C_C &= \begin{bmatrix} 180.4210 \\ 176.4675 \\ 1 \end{bmatrix} \text{ pixels}
\end{aligned}
$$

The Homography necessary to turn C_{aff} back into a circle can be factored into an isometry U and a scaling S as $H_{circle} = USU^{-1}$. The two homographies can be built as follows:

$$
\begin{aligned}
U &= \begin{bmatrix} \cos\theta & -\sin\theta & \\ \sin\theta & \cos\theta & C_c \\ 0 & 0 & \end{bmatrix} \\
S &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & a/b & 0 \\ 0 & 0 & 1 \end{bmatrix} \\
\theta &= 0.5529 \text{ rad}
\end{aligned}
$$

Thus a shape reconstructing homography can be:

$$H_{metric} = H_{circle} \cdot H_{aff}$$

2.2.3. Affinity to align the axis and align the plot (optional)

To facilitate subsequent tasks another affinity is necessary to align the l lines to the x-axis and to center the plot. This is encapsulated in H_{offset}

The final Rectifing Homography is thus:

$$H_{metric} = H_{off} \cdot H_{circle} \cdot H_{aff} = \begin{bmatrix} 0.1041 & -0.2222 & 138.2097 \\ -0.0286 & 0.1633 & -14.5239 \\ -0.0001 & -0.0011 & 1 \end{bmatrix}$$

The result of the rectification is the following:

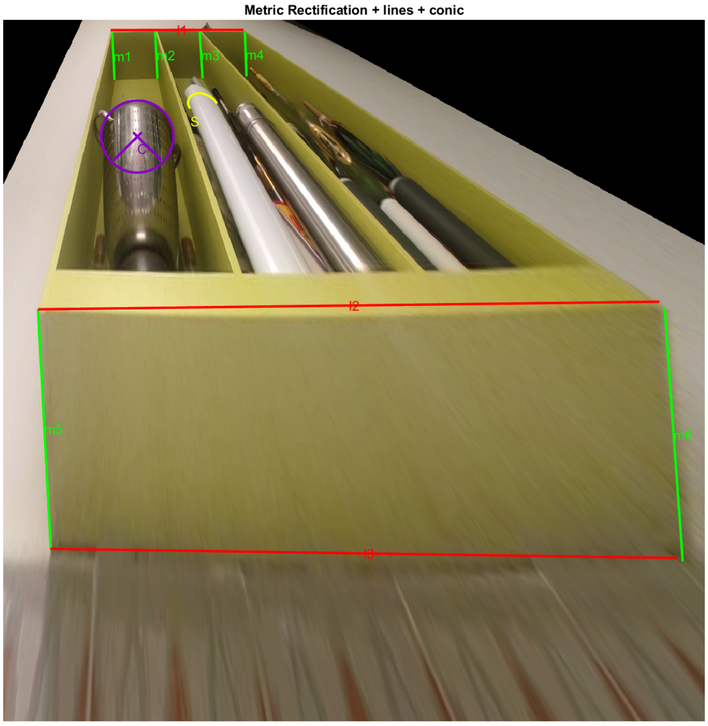

Figure 2.3: The result of the metric rectification

2.2.4. Depth estimation

Estimating the depth m is only a matter of measuring the right things. m_5 and m_6 are the only lines spanning the whole depth and they are both coplanar to l_2 and l_3 that are both 1 unit long in the scene. Since lines that share the same horizontal plane have the same scaling factor in the metric rectification, follows that by calling $m^m_{5_1}$, $m^m_{5_2}$, $m^m_{6_1}$, $m^m_{6_2}$ $l^m_{2_1}$, $l^m_{2_2}$, $l^m_{3_1}$, $l^m_{3_2}$ the homogenous coordinates of the pixel coordinated of the endpoints of the segments in the metric rectified image, the depth is:

$$m = \frac{||m^m_{5_1} - m^m_{5_2}|| + ||m^m_{6_1} - m^m_{6_2}||}{||l^m_{2_1} - l^m_{2_2}|| + ||l^m_{3_1} - l^m_{3_2}||} = 0.39571 \text{ units}$$

With the exact same reasoning translated to the $m_{1:4}$ and l_1 lines, the internal depth of the cabinet can be computed:

$$m_{internal} = 0.35073 \text{ units}$$

With this two information, the depth of the back plate of the cabinet can be estimated (This will come in handy for the 3D model reconstruction).

$$m_{backplate} = m - m_{internal} = 0.044985 \text{ units}$$

2.3. Intrinsic Calibration

The goal of this task is to find the calibration matrix K built as follows:

$$K = \begin{bmatrix} f_x & 0 & U_0 \\ 0 & f_y & V_0 \\ 0 & 0 & 1 \end{bmatrix}$$

Instead of estimating K directly, it is easier to find $\omega \triangleq (KK^T)^{-1}$. This matrix is built as follows:

$$\omega = \begin{bmatrix} a^2 & 0 & -U_0a^2 \\ 0 & 1 & -V_0 \\ -U_0a^2 & -V_0 & f_y^2 + a^2U_0^2 + V_0^2 \end{bmatrix}$$

By knowing an homography $H = [h_1, h_2, h_3]$ that goes from the coordinate on a plane π to the image space and the vanishing point v along the direction orthogonal to π, the following conditions must hold:

$$\begin{aligned} h_1^T\omega h_2 &= 0 \\ h_1^T\omega h_1 - h_2^T\omega h_2 &= 0 \\ v^T\omega h_1 &= 0 \\ v^T\omega h_2 &= 0 \end{aligned}$$

And thus the entries of ω can be estimated as solutions of a linear system of equations.

K can then be reconstructed with the Cholesky decomposition of ω^-1.

In this case, the homography H used is equal to H_{metric}^{-1} thus setting the plane π to be horizontal. The vanishing point v can be computed with Procedure 1 on the lines $h_{1:4}$.

The estimated calibration matrix is:

$$K = \begin{bmatrix} 782.3411 & 0 & 800.8240 \\ 0 & 789.2100 & 534.6531 \\ 0 & 0 & 1 \end{bmatrix}$$

2.4. Height estimation

By multiplying H with K^{-1}, a set of basis of π $r1$ and $r2$ and the offset o_π are obtained, expressed in the reference frame of the camera:

$$r_{cam} = \begin{bmatrix} r_1 & r_2 & o_\pi \end{bmatrix} = K^{-1}H$$

These coordinates are known up to scale, which is the key for selecting planes at a given height. That's because it is known that the segments $l_{1:3}$ are 1 unit long in the scene, thus by dividing the vectors by the length of the metric rectified l_1^m ($|l_1^m|$) or l_2^m ($|l_2^m|$), a set of 3D origin and coordinates of the upper and lower face of the cabinet can be found. Calling $\hat{r_3}$ a unit vector orthogonal to both r_1 and r_2, the height can be simply computed as:

$$h = \hat{r_3}^T \cdot (\frac{o_\pi}{|l_1^m|} - \frac{o_\pi}{|l_2^m|}) = (\frac{1}{|l_1^m|} - \frac{1}{|l_2^m|})\hat{r_3}^T \cdot o_\pi = 0.3842 \text{ units}$$

2.5. Coordinates of S

Starting from the points of the curve in the original image expressed in homogenous coordinates in the matrix S. First, they need to be metric rectified obtaining $S_{metric} = H_{metric} \cdot S$. They can then be expressed in the camera 3d coordinate system as $S'_{cam} = r_{cam} \cdot S_{metric}$.

Now the appropriate scaling factor s to place the points at $h/2$ needs to be found. The scaling s is one such that by scaling the offset of the plane of the top face of the cabinet it moves by $h/2$ units along the $\hat{r_3}$ direction

$$\hat{r_3}^T \cdot (\frac{o_\pi}{|l_1^m|} - s\frac{o_\pi}{|l_1^m|}) = h/2$$
$$(1 - s) = h/2$$
$$s = 1 - \frac{h \cdot |l_1^m|}{2 \cdot \hat{r_3}^T \cdot o_\pi}$$

Thus $S_{cam} = s \cdot S'_{cam}$

Then the 3D coordinates of the chosen origin (the bottom left corner of the cabinet closer to the camera) must be found. Luckily, the coordinate of that point in the metric rectification is $l_{2_1}^m$, thus:

$$O_{cam} = \frac{r_{cam} \cdot l_{2_1}^m}{|l_2^m|}$$

The last piece is to find the $\hat{X}_{cam}$ and $\hat{Y}_{cam}$ versors in the camera space. They are aligned with r_1 and r_2.

$$\hat{X}_{cam} = r_1/|r1|$$

$$\hat{y}_{cam} = r_2/|r2|$$

Thus

$$S_{XY} = \begin{bmatrix} \hat{X}_{cam}^{T} \\ \hat{Y}_{cam}^{T} \end{bmatrix} \cdot (S_{cam} - O_{cam})$$

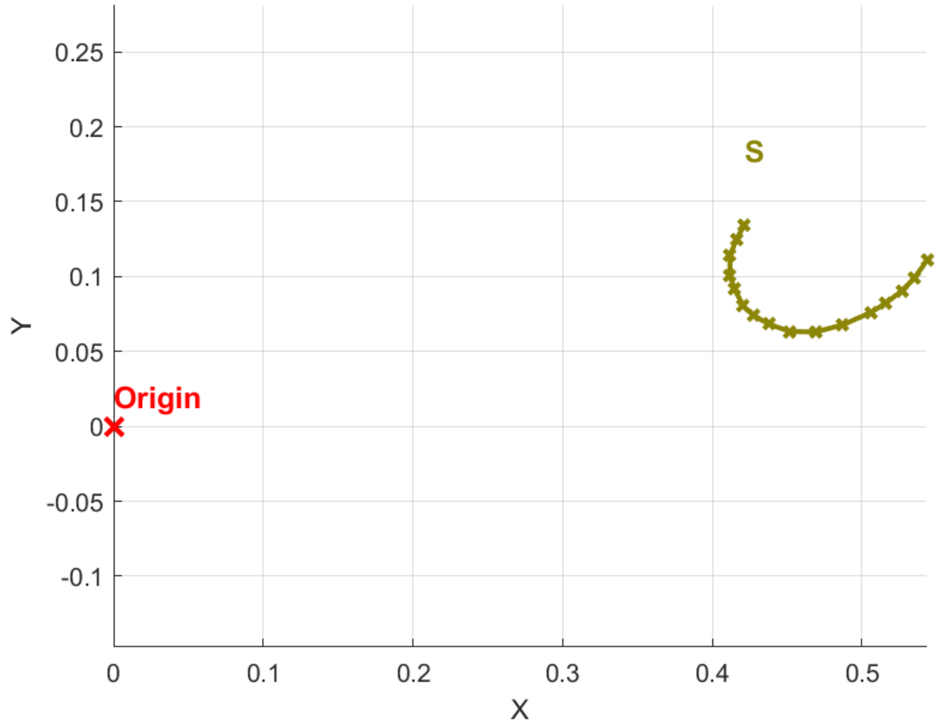

Figure 2.4: The X Y Coordinates of S

2.6. Camera Localization

Since both the chosen origin and the chosen axis are known in the camera reference frame:

$$R^{-1} = \begin{bmatrix} \hat{X}_{cam} & \hat{Y}_{cam} & \hat{r_3} & O_{cam} \end{bmatrix}$$

The computed R matrix is:

$$R = \begin{bmatrix} 0.9539 & 0.0510 & 0.2956 & 0.1479 \\ -0.2924 & 0.3790 & 0.8780 & -0.5679 \\ -0.0673 & -0.9240 & 0.3765 & -0.1038 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

3 | 3D model

In order to construct the 3D Model of the cabinet the last pieces of information needed are the width of the cabinet panels and the width of one "cell" of the cabinet. To do so, the metric rectification $H_f ront$ of the front face of the cabinet is needed. That is easily obtained using the calibration matrix as:

$$H_{front} = K \cdot \begin{bmatrix} r_1 & \hat{r_3} \cdot |r_1| & o_\pi \end{bmatrix}$$

Once the image is rectified it is all a matter of taking the right measurements: the width of the cabinet panels is 0.0157 units long while the width of one "cell" is 0.3167 units.

The rectified image and the taken references are:

Figure 3.1: The Metric rectification of the vertical face

Using all the information gathered so far a 3D model of the scene can be built:

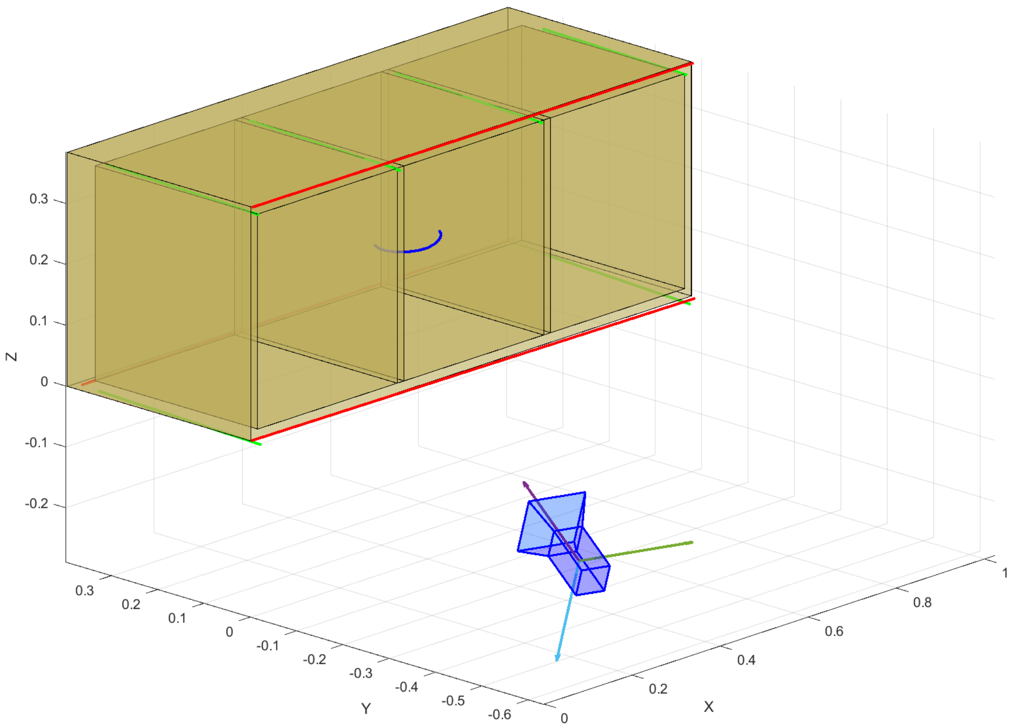

Figure 3.2: A view of the 3D model

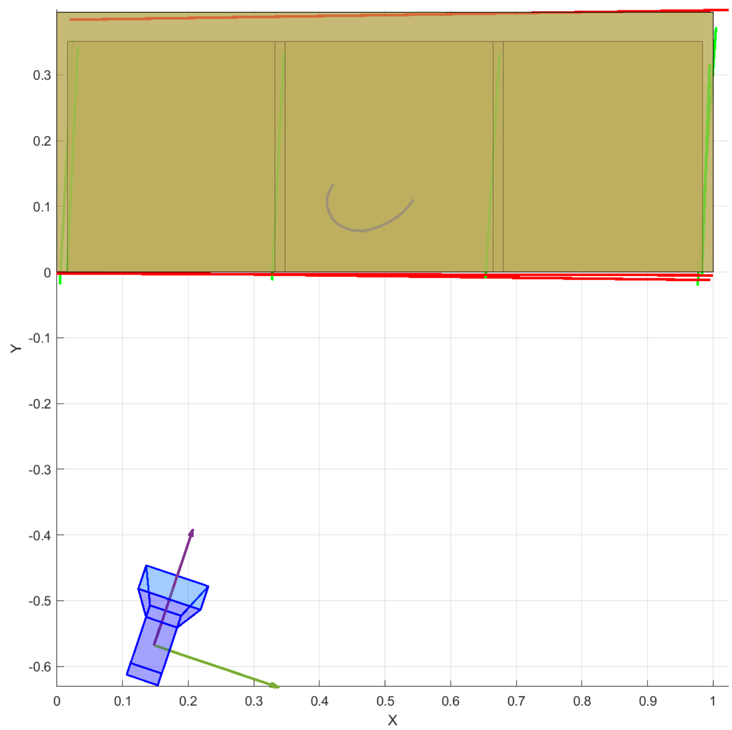

Figure 3.3: A view of the 3D model from the top

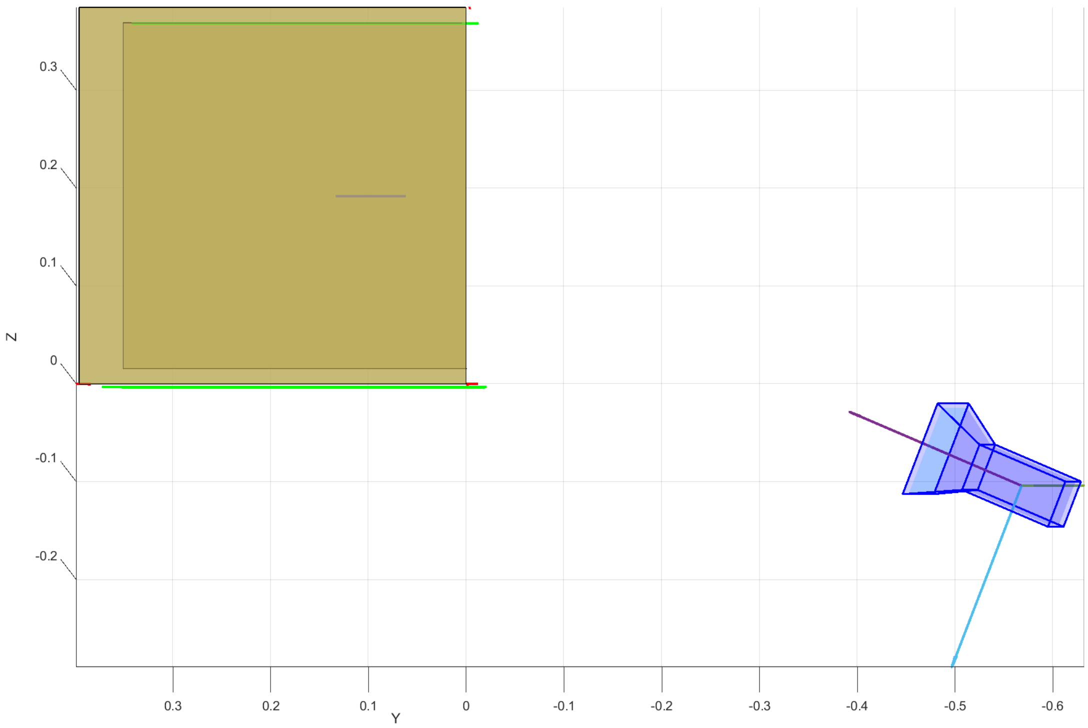

Figure 3.4: A view of the 3D model from the side

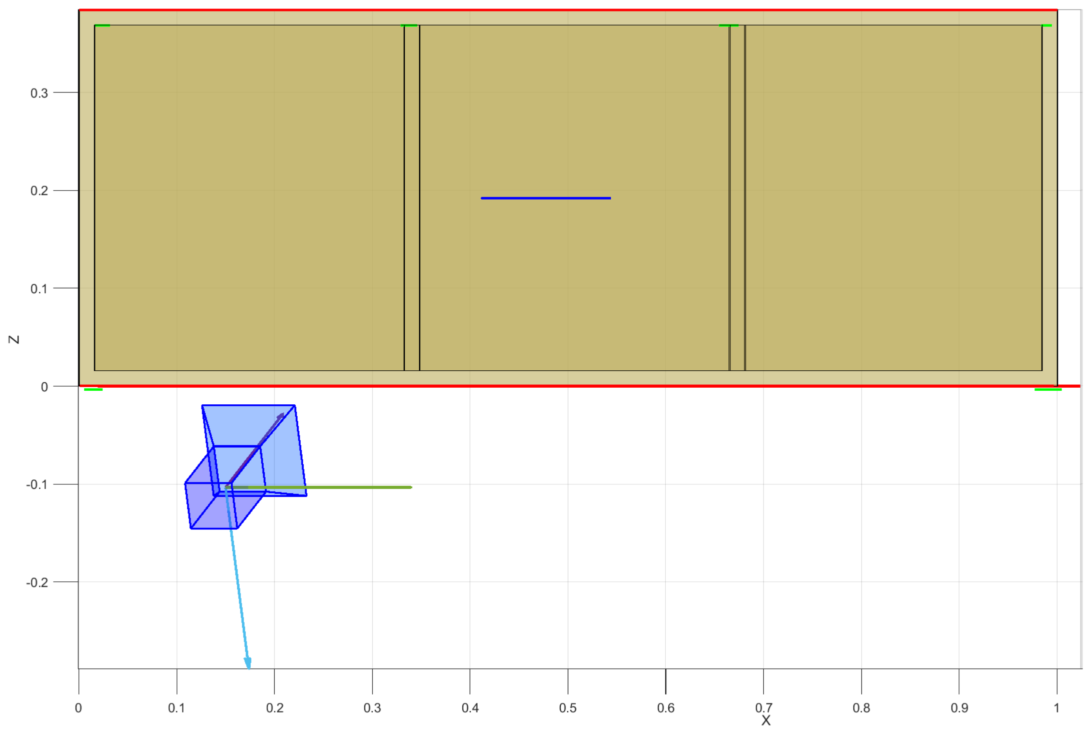

Figure 3.5: A view of the 3D model from the front

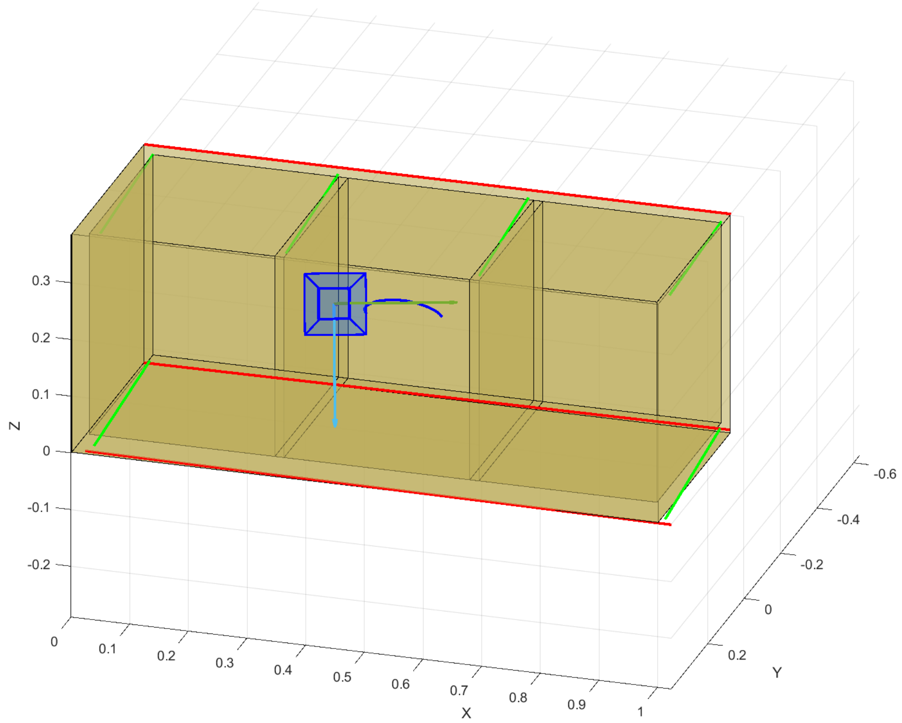

Figure 3.6: A view of the 3D model from the camera's perspective (orthographic)

List of Figures