International Institute
Of Information
Technology Hyderabad

Team: Robusta

Stereo Camera Calibration

Mentor: Prof. Madhava Krishna
TA: Laksh Sir

Submitted by

Umang Sharma, 2021102028
Ashish Chokhani, 2021102016

5th December 2023

Overview

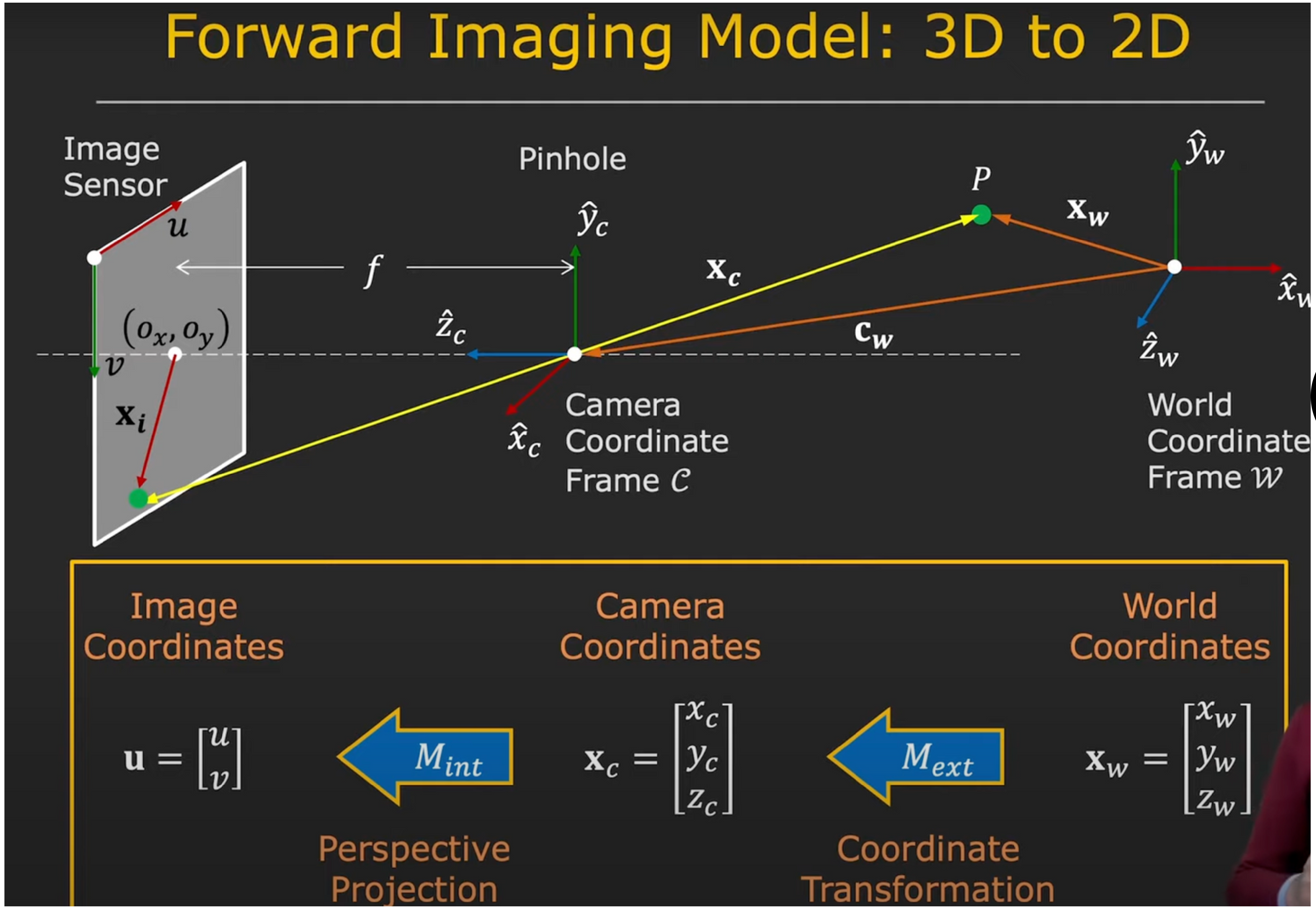
Introduction

Stereo camera callibration - The project involves using 2 real monocular cameras, setup our own stereo camera setup, implementing depth calculation from camera parameters.

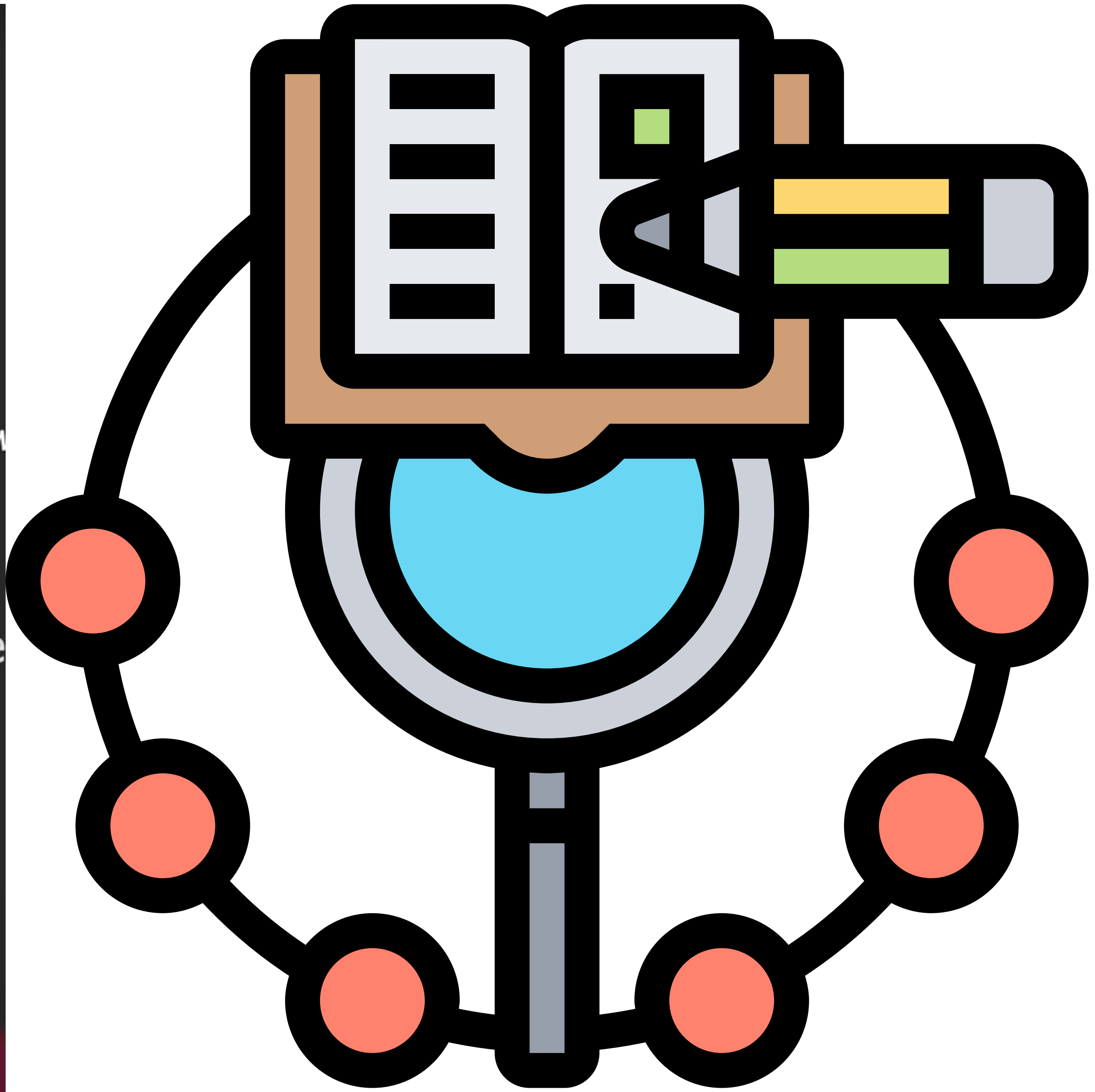
Report the effects of varying baseline.Also use this setup to map out the environment.

Add random noise to the images and try to reconstruct image.

Flow Chart



Forward Imaging Model: 3D to 2D
Image Sensor
Pinhole
u
f
$\hat{y}_c$
$\mathbf{x}_c$
P
$\mathbf{x}_w$
$\hat{y}_w$
$\hat{x}_w$
(o_x, o_y)
$\hat{z}_c$
$\mathbf{c}_w$
$\hat{z}_w$
v
$\mathbf{x}_i$
$\hat{x}_c$
Camera Coordinate Frame $\mathcal{C}$
World Coordinate Frame $\mathcal{W}$
Image Coordinates
Camera Coordinates
World Coordinates
$\mathbf{u} = \begin{bmatrix} u \\ v \end{bmatrix}$
M_{int}
$\mathbf{x}_c = \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix}$
M_{ext}
$\mathbf{x}_w = \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix}$
Perspective Projection
Coordinate Transformation



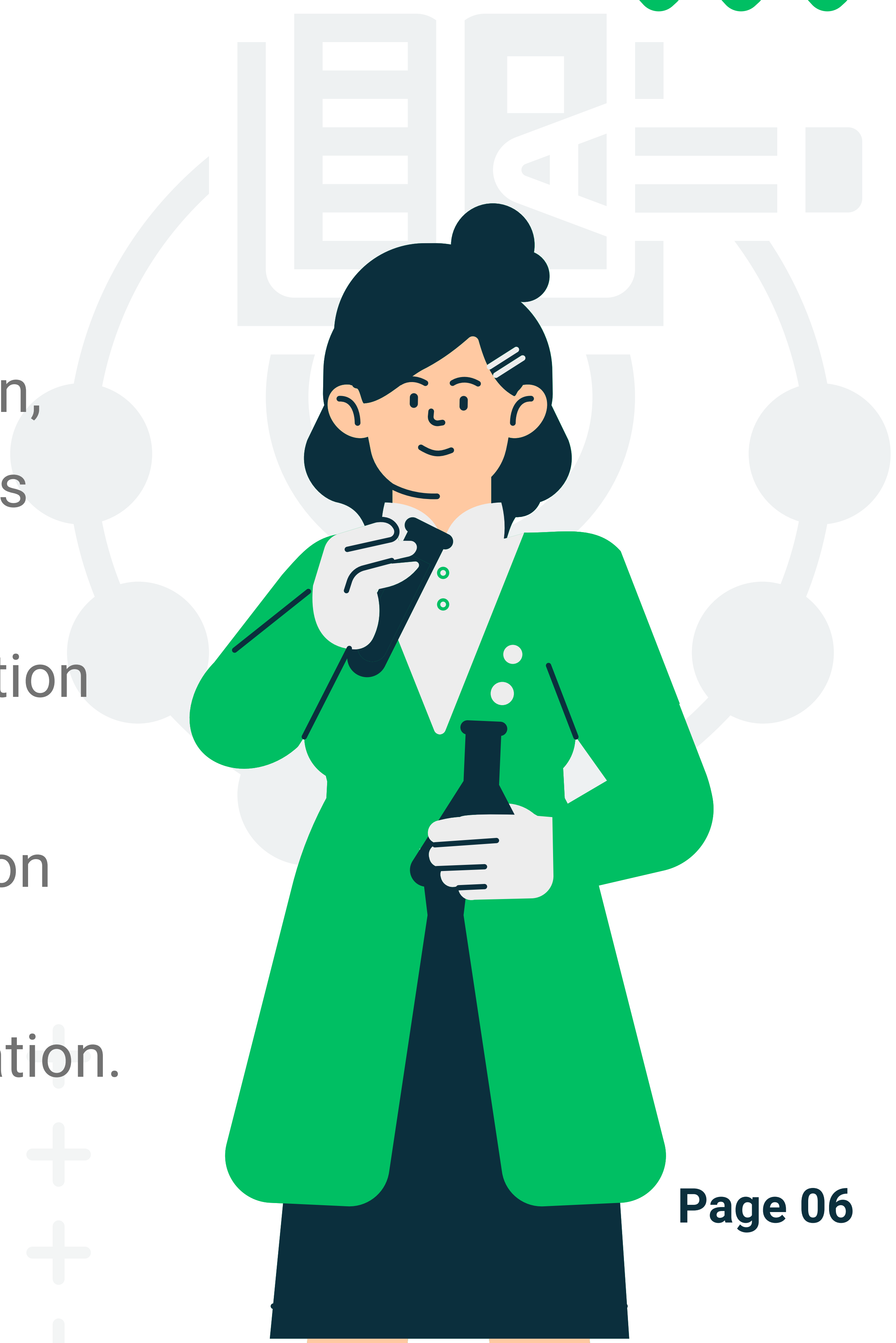
Work Progress to this point

- **Step 1: Stereo camera Setup**

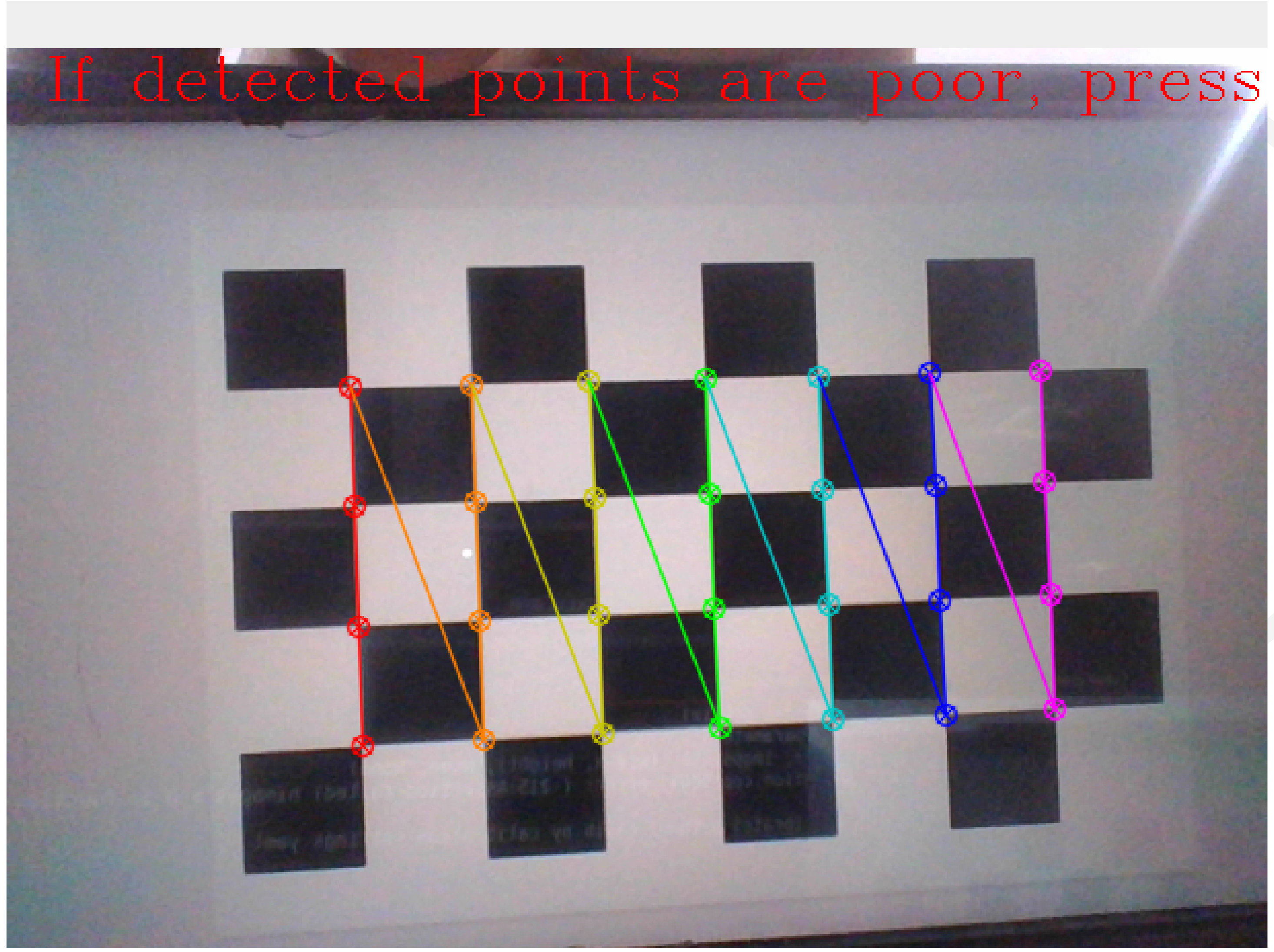
- Use information from two different views of the same object in order to reconstruct features visible in both views into 3D
- A setup with two monocular cameras and a varying baseline.

Work Progress to this point

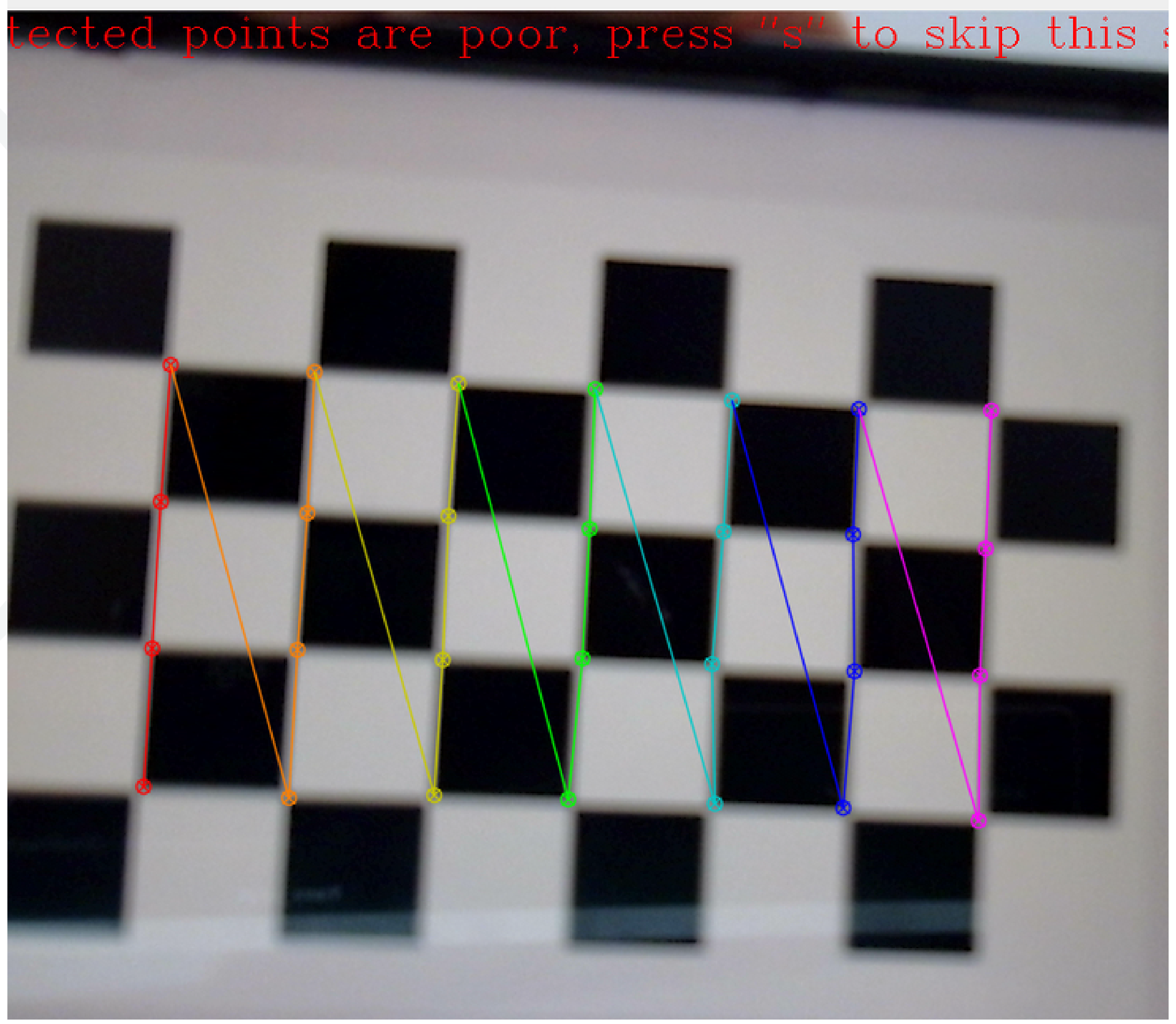
- ## Step 2: Saving Calibration Pattern Frames
- **Pattern Selection**: We typically choose a calibration pattern, typically a checkerboard or a grid of known dimensions. It's essential that the pattern is easily detectable in images.
- **Image Acquisition**: Capture multiple images of the calibration pattern from various angles and positions.
- **Data Storage**: Save these images for subsequent calibration steps. A comprehensive set of well-distributed calibration pattern frames is fundamental to achieving precise calibration.



Detection of Callibration pattern points



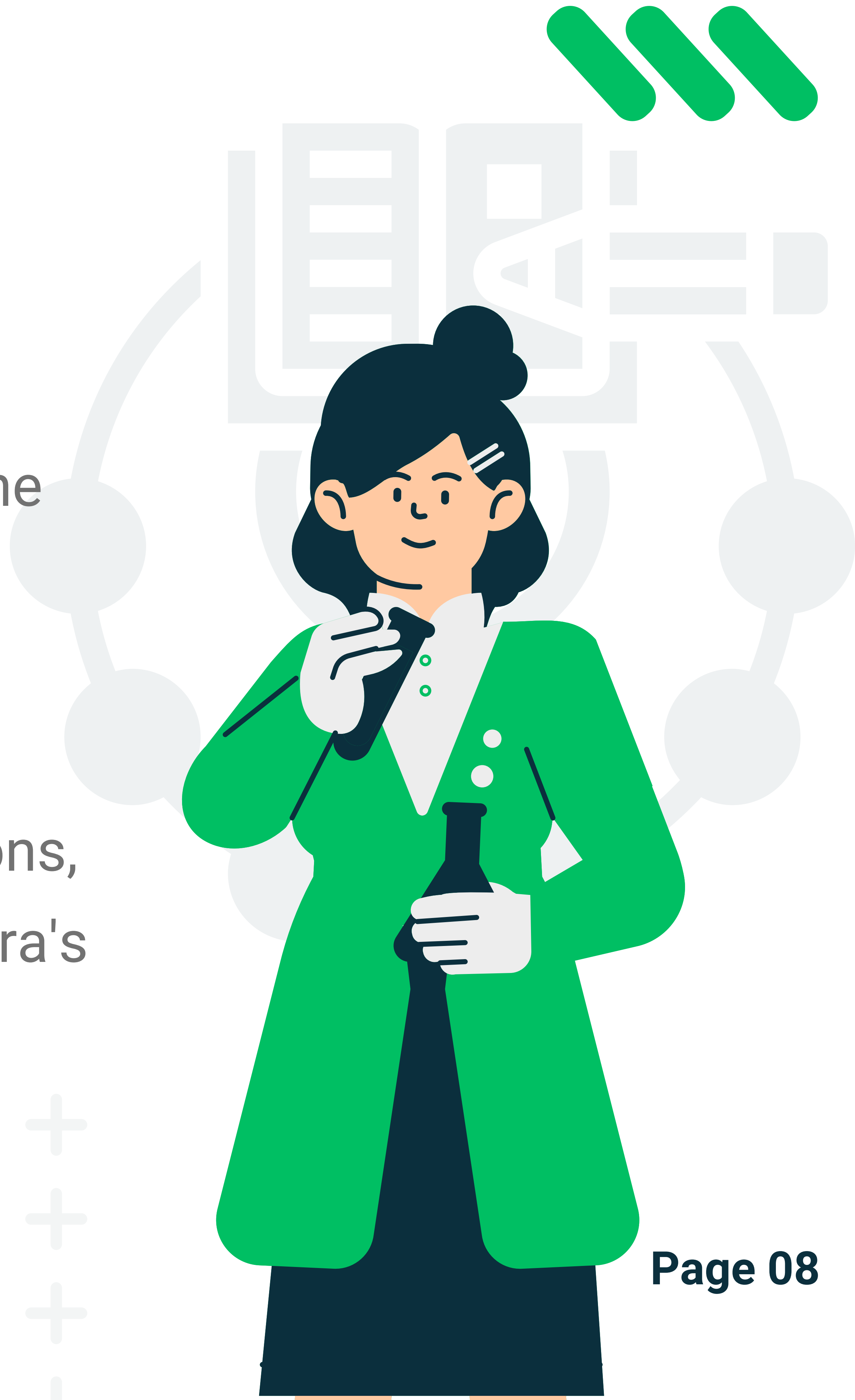
For 1st camera



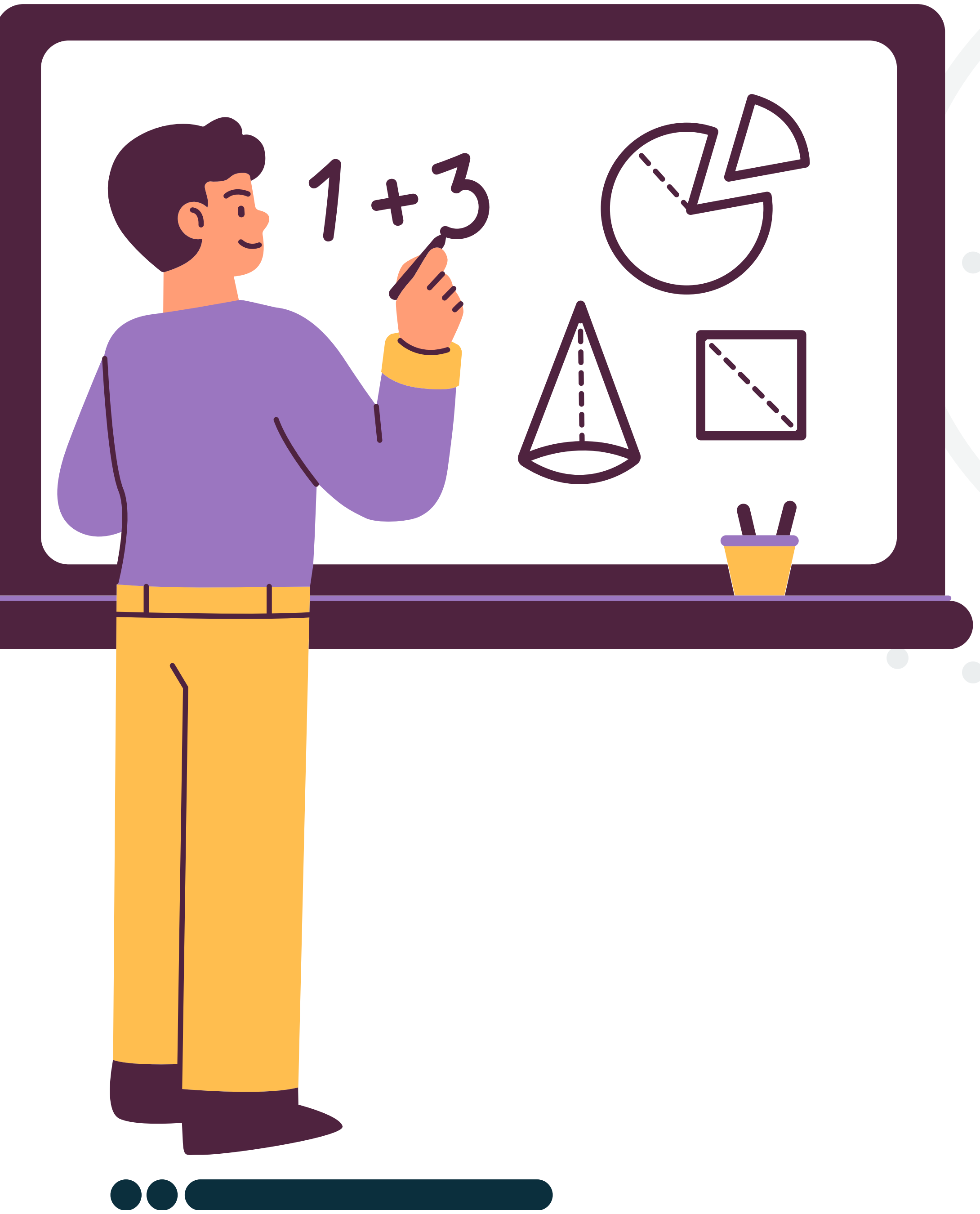
For 2nd camera

Work Progress to this point

- **Step 3: Intrinsic camera parameters**
- Intrinsic parameters characterize the internal aspects of the camera.
- Obtained through calibration procedures using known calibration patterns.
- These parameters are fundamental for correcting distortions, determining the field of view, and understanding the camera's internal geometry. They are crucial for precise 3D reconstructions.



Equations involved



Extracting Intrinsic/Extrinsic Parameters

We know that:

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & \boxed{p_{14}} \\ p_{21} & p_{22} & p_{23} & \boxed{p_{24}} \\ p_{31} & p_{32} & p_{33} & \boxed{p_{34}} \end{bmatrix} = \begin{bmatrix} f_x & 0 & o_x & 0 \\ 0 & f_y & o_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

That is:

$$\begin{bmatrix} p_{14} \\ p_{24} \\ p_{34} \end{bmatrix} = \begin{bmatrix} f_x & 0 & o_x \\ 0 & f_y & o_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} = K\mathbf{t}$$

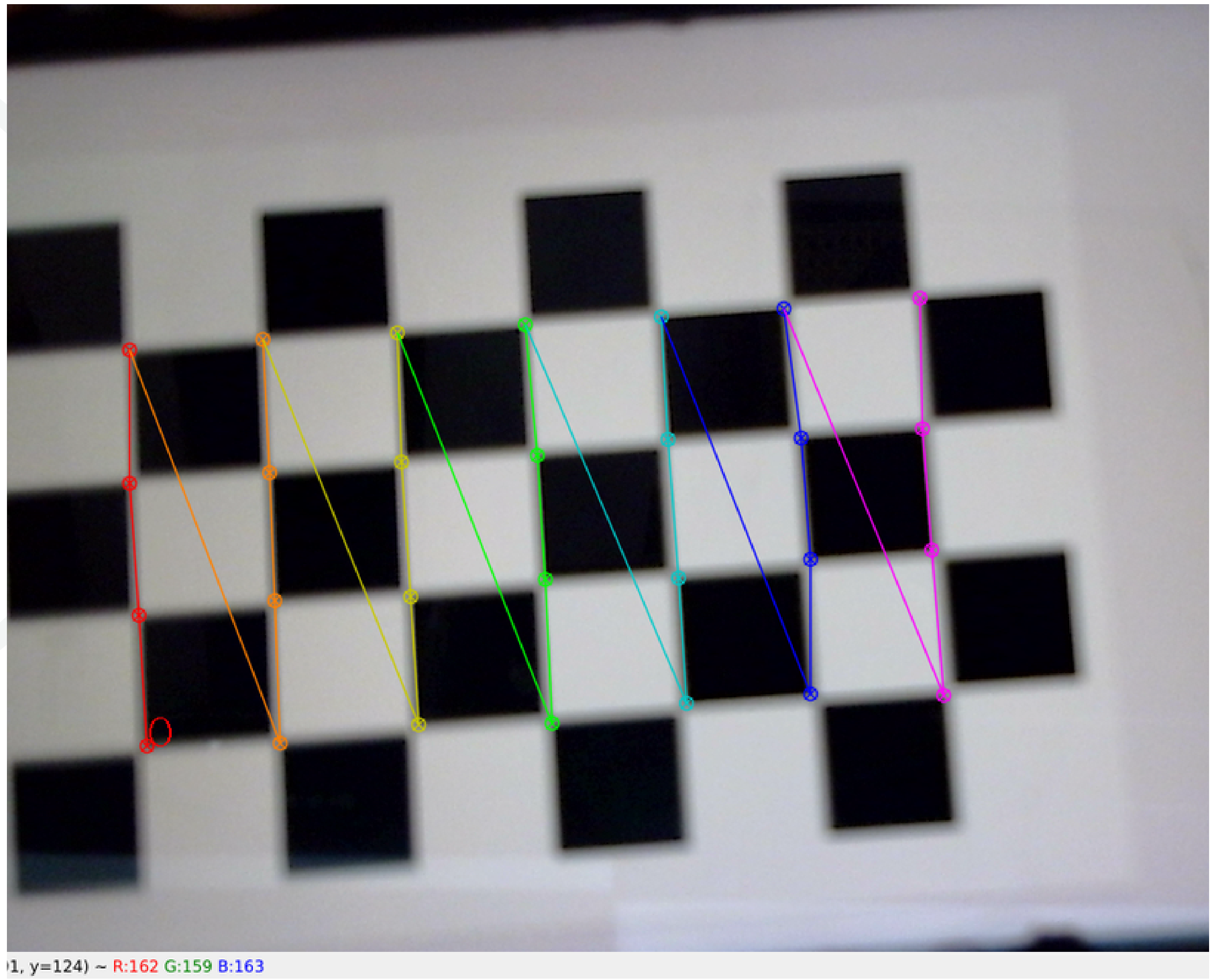
Therefore:

$$\boxed{\mathbf{t} = K^{-1} \begin{bmatrix} p_{14} \\ p_{24} \\ p_{34} \end{bmatrix}}$$

Work Progress to this point

- **Step 4:Camera0 to Camera1 Rotation and Translation**
- In this step, we focus on calculating the rotation matrix (R) and translation vector (T) that transform points from Camera0 to Camera1 coordinate space.
- R and T play a fundamental role in triangulating 3D points, depth calculation, and mapping objects in 3D space. Their precise determination is crucial for high-quality 3D reconstructions.

Getting World Coordinates



For 1st camera

For 2nd camera

Opencv assumes that the bottom left of the checkerboard pattern is the world coordinate.

Work Progress to this point

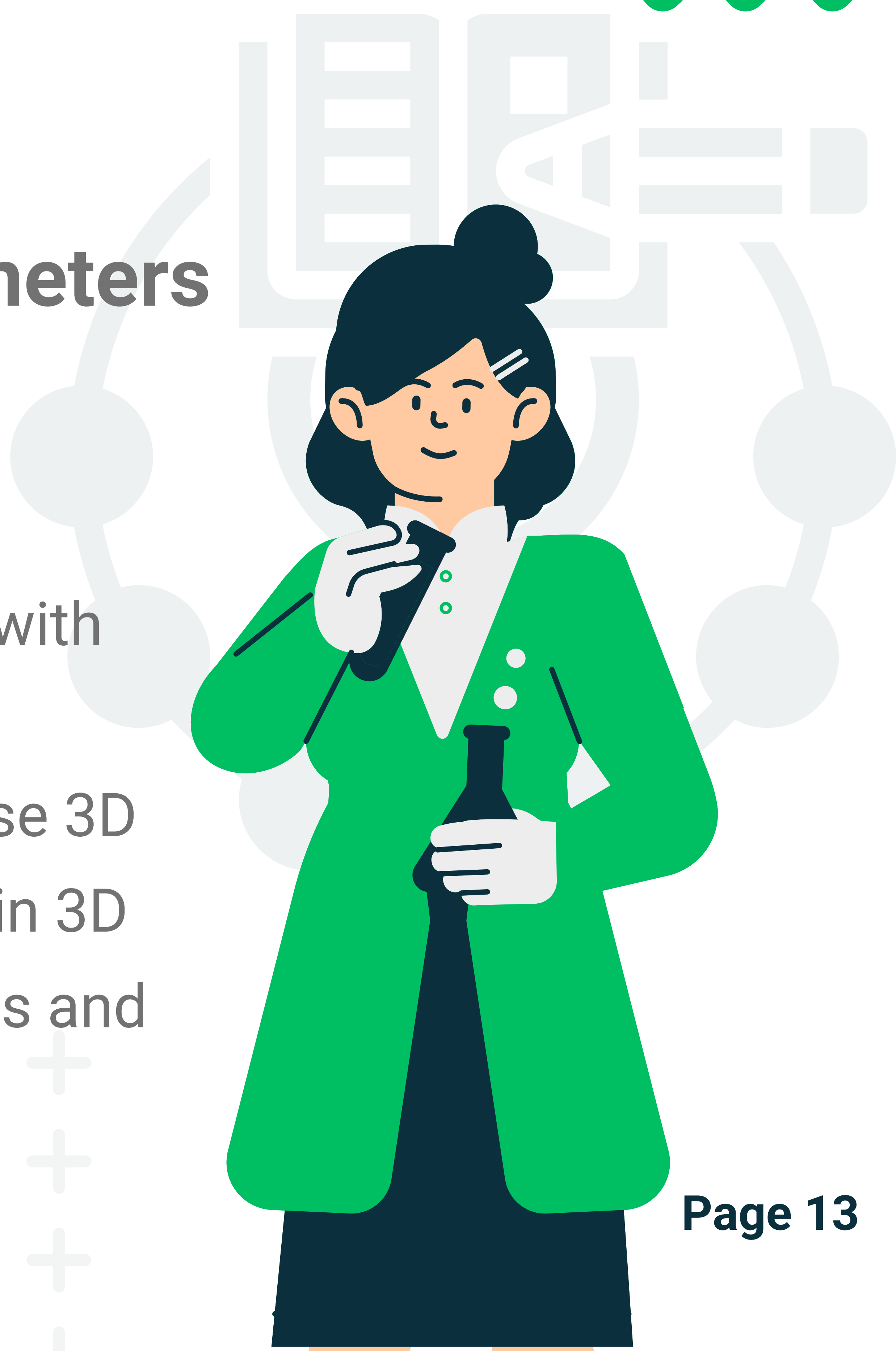
- **Step 5: Stereo Calibration Extrinsic Parameters**

- Extrinsic parameters establishes the relationship between the cameras and the 3D world.
- **World Space Origin:** Typically, we select Camera0's position as the world origin. This choice simplifies the calibration process and is commonly used in stereo vision.
- **World Origin Rotation**: The rotation matrix (R) is usually an identity matrix, indicating that the world's orientation matches Camera0's orientation.

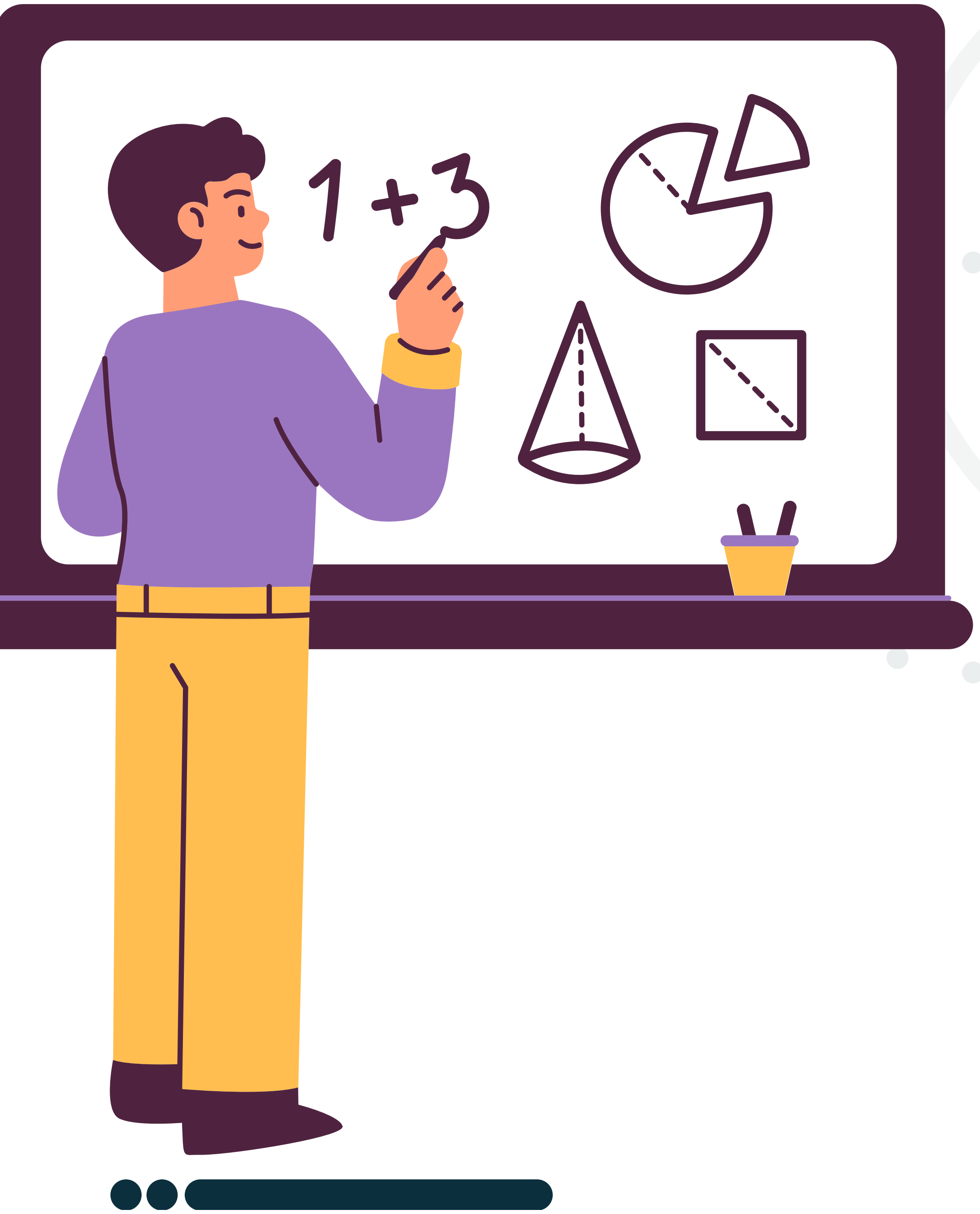
Work Progress to this point

- ## Step 5: Stereo Calibration Extrinsic Parameters

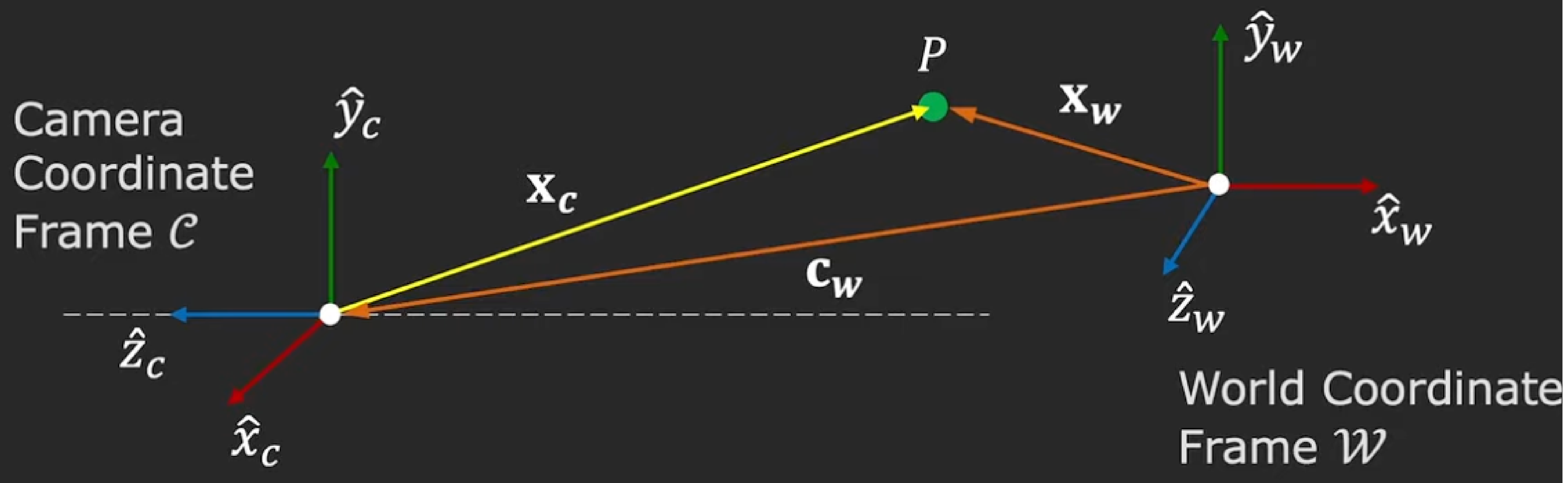
- **Practical Significance**: These parameters ensure that 3D triangulated points are in reference to a consistent world coordinate system. R and T essentially align the 3D world with the 3D coordinates in Camera0's view.
- **Application**: The extrinsic parameters are critical for precise 3D reconstructions, depth calculations, and mapping objects in 3D space. They establish the connection between the cameras and the real world.



Equations involved



Extrinsic Parameters



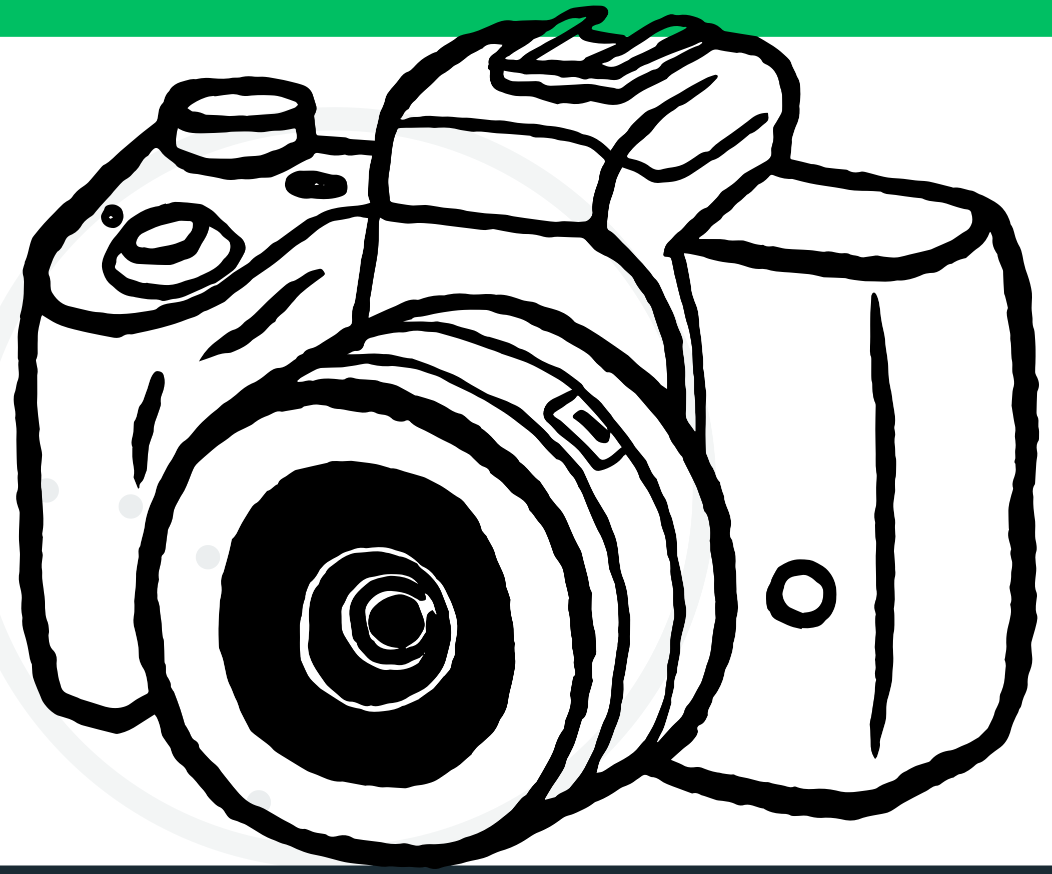
Position $\mathbf{c}_w$ and Orientation R of the camera in the world coordinate frame $\mathcal{W}$ are the camera's Extrinsic Parameters.

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}$$

→ Row 1: Direction of $\hat{x}_c$ in world coordinate frame

Obtained Intrinsic and Extrinsic parameters for baseline =6 cm

For 1st camera



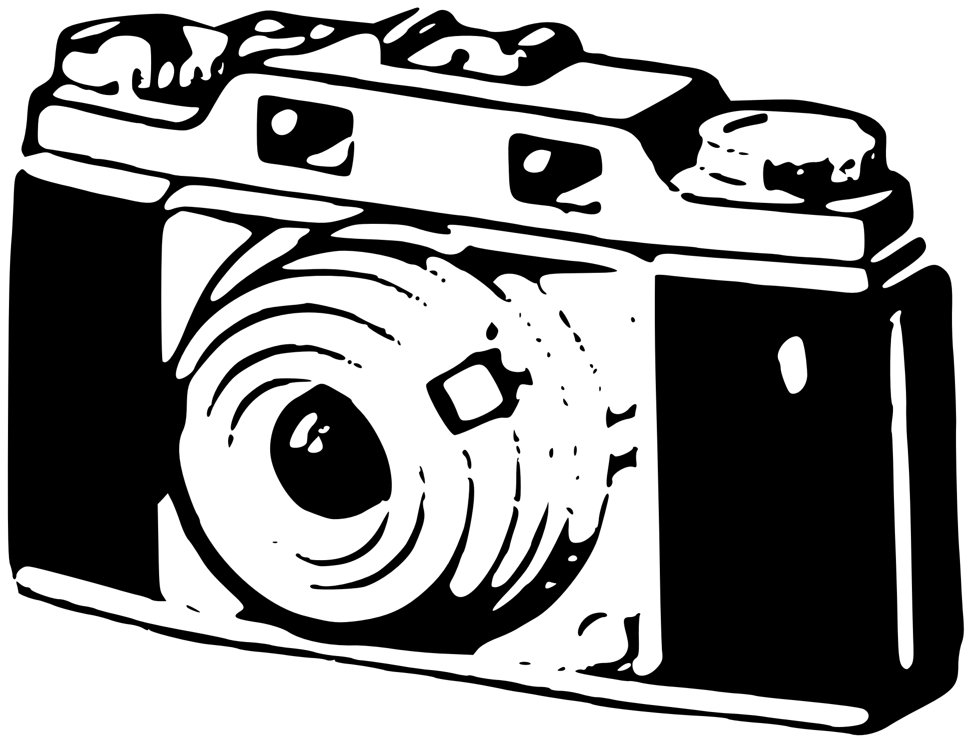
```
R:
1.0 0.0 0.0
0.0 1.0 0.0
0.0 0.0 1.0
T:
0.0
0.0
0.0
```

```
intrinsic:
338.18736797002754 0.0 334.96330674516423
0.0 346.18171193591695 193.33142720090092
0.0 0.0 1.0
distortion:
-0.040652002803869304 0.5728629939672975 0.0015908264192583171 -0.001895324038032022 -1.8007539122364322
```

Obtained Intrinsic and Extrinsic parameters for baseline =6 cm

```
intrinsic:
1255.7604351115792 0.0 651.3350560145929
0.0 1254.6820280856366 354.43815594046333
0.0 0.0 1.0
distortion:
0.40579841090871666 -2.701118134824421 -0.000587329901545849 0.011030965978108853 6.996600326659647
```

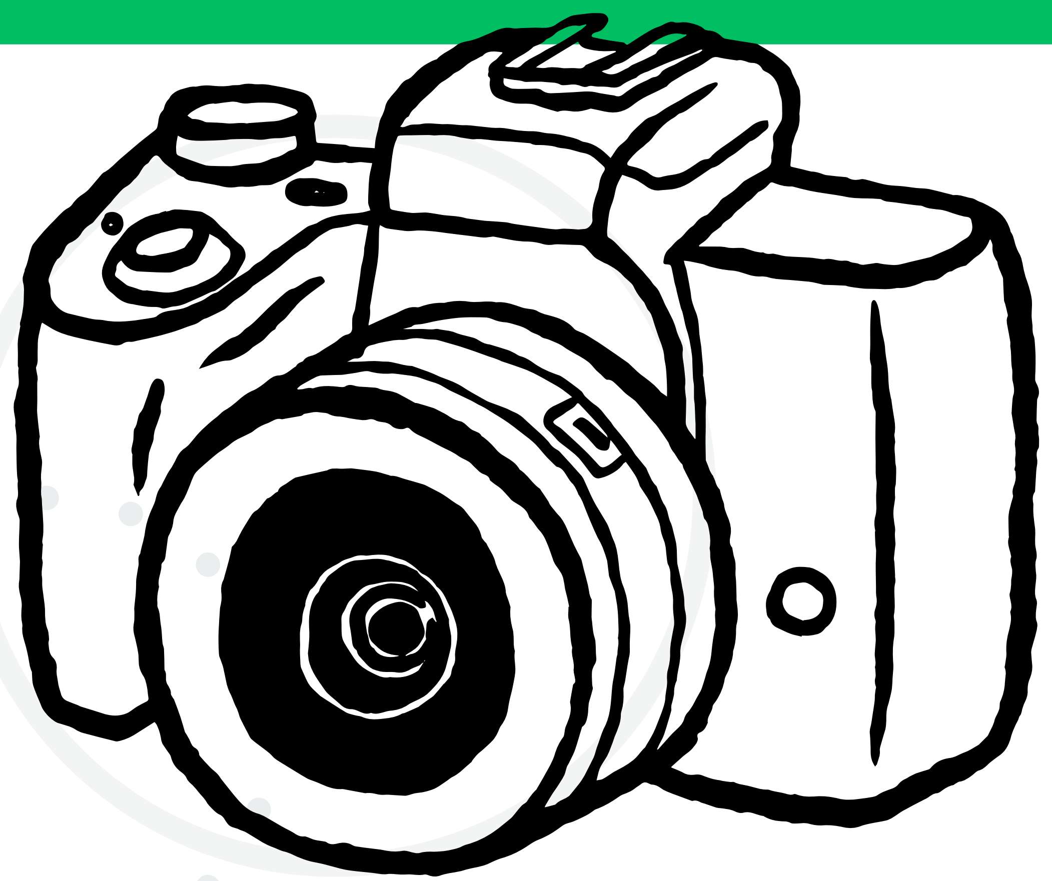
```
R:
0.9979743513545163 0.005634505223043135 -0.06336754997176308
-0.006215781239189237 0.9999403628232939 -0.008979691554044827
0.06331317479102541 0.009355380682299385 0.997949857833634
T:
-6.251265824874118
0.872655855609496
4.195869451396089
```



For 2nd camera

Obtained Intrinsic and Extrinsic parameters for baseline =8 cm

For 1st camera



```
R:
1.0 0.0 0.0
0.0 1.0 0.0
0.0 0.0 1.0
T:
0.0
```

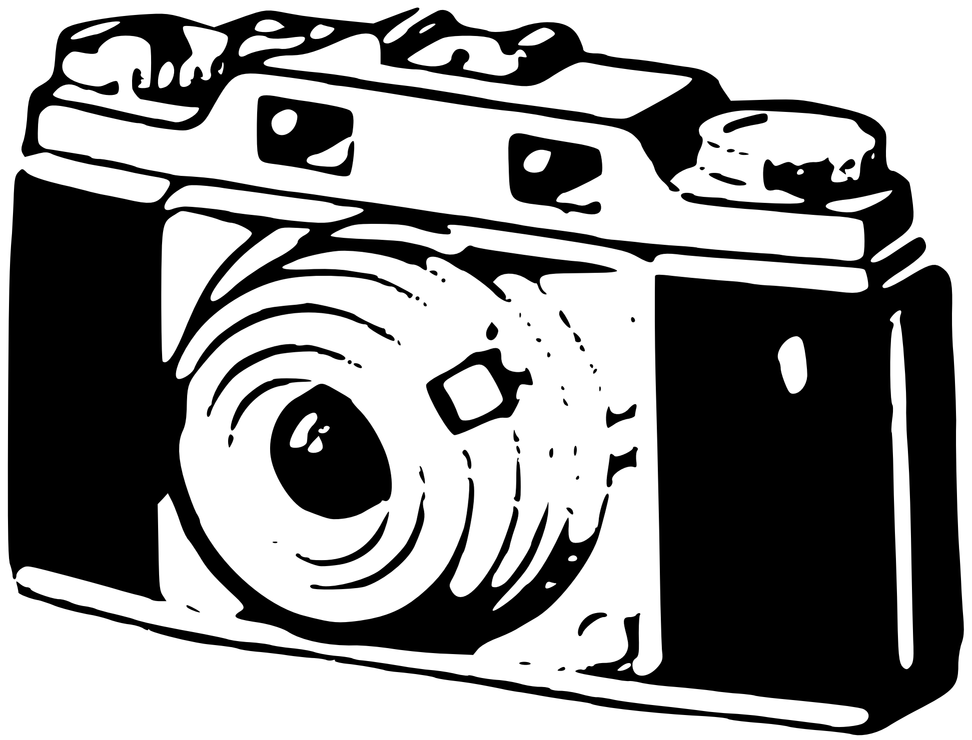
~/Downloads/parameter/camera_parameters_8

```
intrinsic:
560.7537014023897 0.0 312.3166764807988
0.0 554.7488275638356 269.3291186922631
0.0 0.0 1.0
distortion:
-0.010158646718115722 0.0755540524267572 0.0017592525884421231 -0.004479027715408595 -0.1635301764824323
```

Obtained Intrinsic and Extrinsic parameters for baseline =8 cm

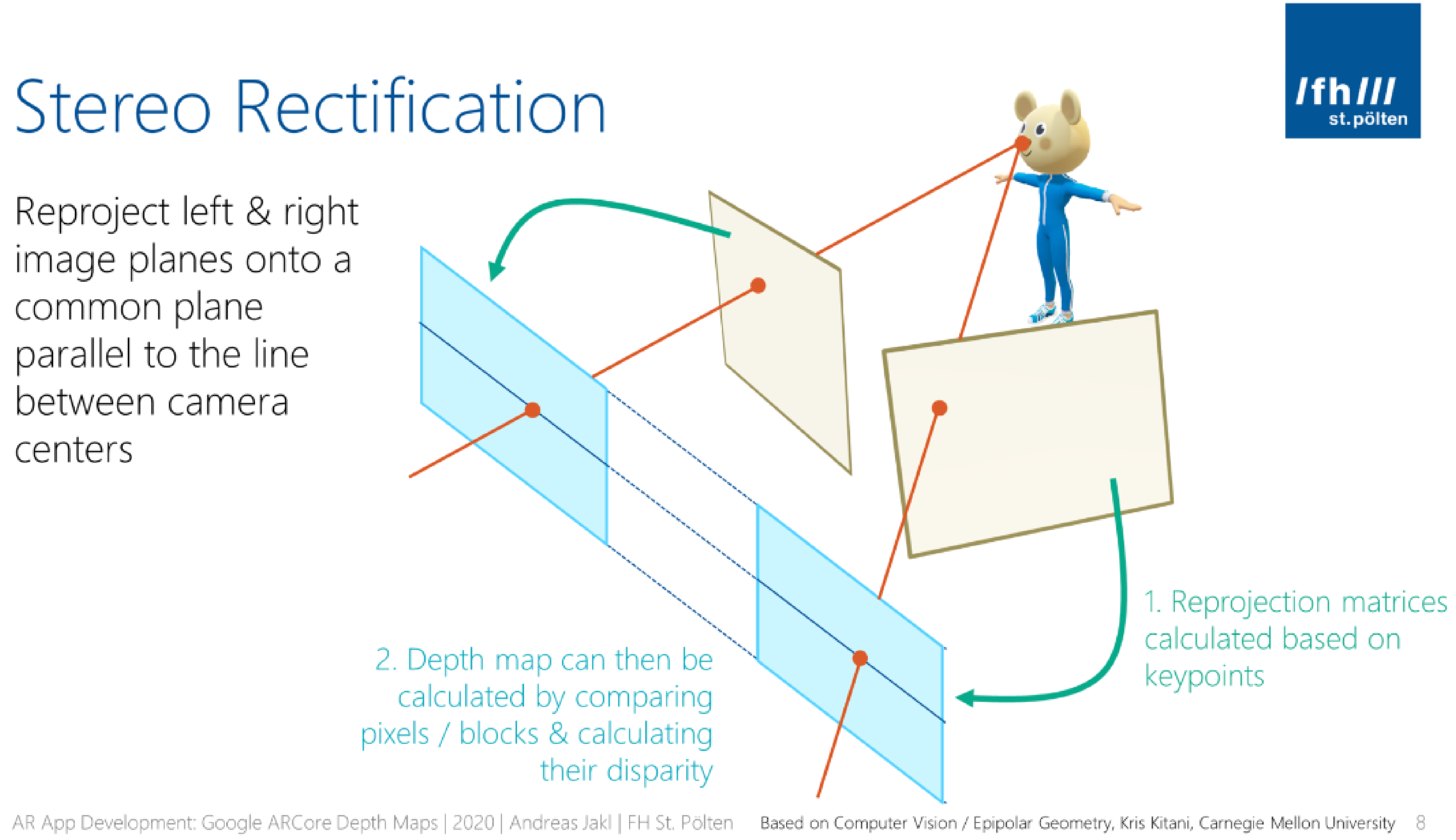
```
intrinsic:
2538.345110517018 0.0 593.415967453494
0.0 2622.6298869606103 353.39125400255335
0.0 0.0 1.0
distortion:
0.32825576472565876 -13.364673398281028 0.0026897348187760605 0.010557437751946178 169.39757819981855
```

```
R:
0.9992817306558963 0.004131338852181506 -0.03766901666682
0.0044071145369434465 0.9746243549709025 0.22380336020491393
0.03763774858855096 -0.22380862078311603 0.9739061048911976
T:
-7.6739166960028165
1.2520536463698175
14.607047944214173
```



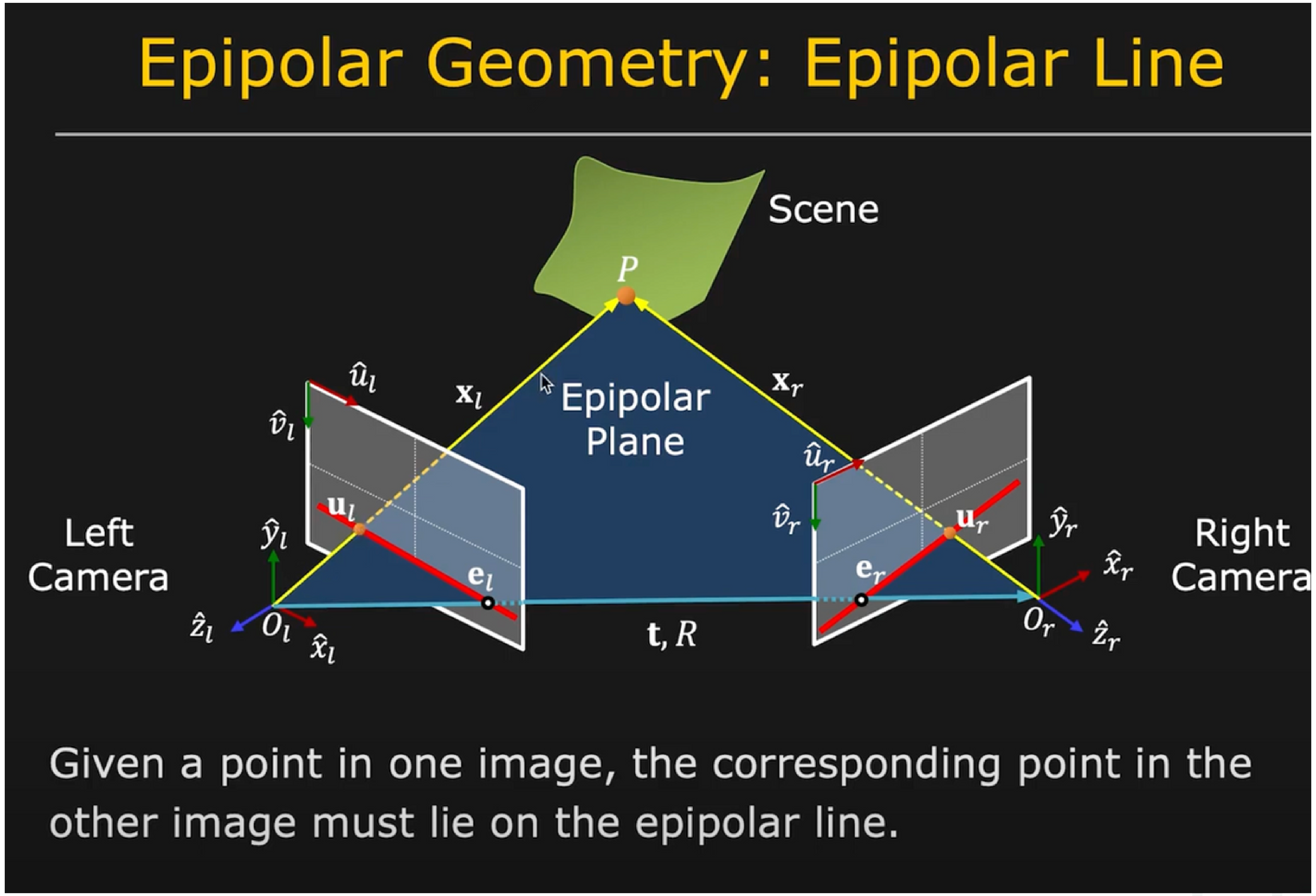
For 2nd camera

Stereo Rectification



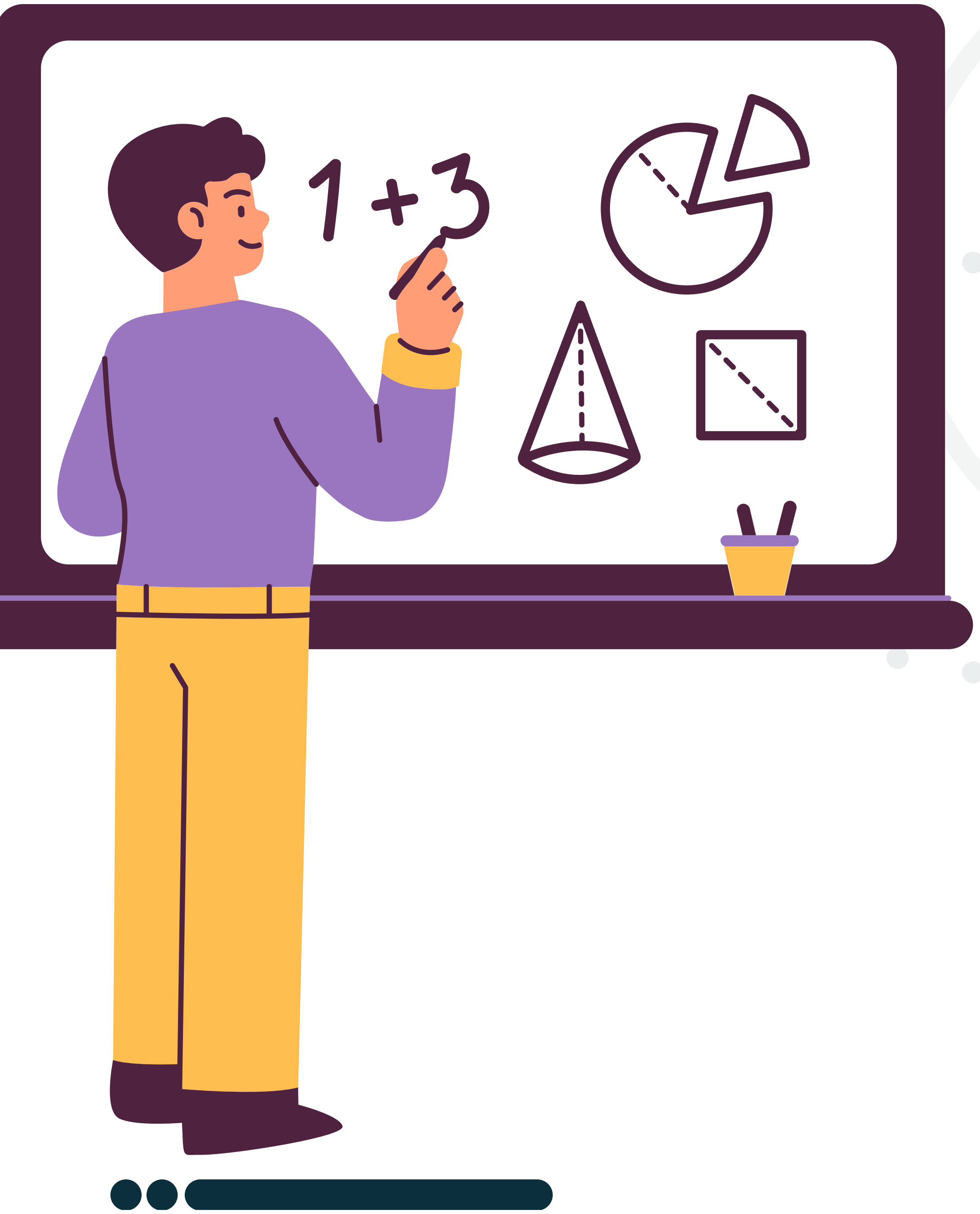
- Detect keypoints in each image.
- We then need the best keypoints where we are sure they are **matched in both images** to calculate reprojection matrices.
- Using these, we can **rectify the images** to a common image plane.
- Matching keypoints are on the same horizontal epipolar line in both images.
- This enables efficient pixel / block comparison to calculate the disparity map

Finding Correspondence



- Given a point in one image the corresponding point in the other image must lie on the epipolar line.
- Finding correspondence reduces to 1D search

Equations involved



Finding Epipolar Lines

Given: Fundamental matrix F and point on left image $(u_l,$

Find: Equation of Epipolar line in the right image

Epipolar Constraint Equation:

$$\begin{bmatrix} u_l & v_l & 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} = 0$$

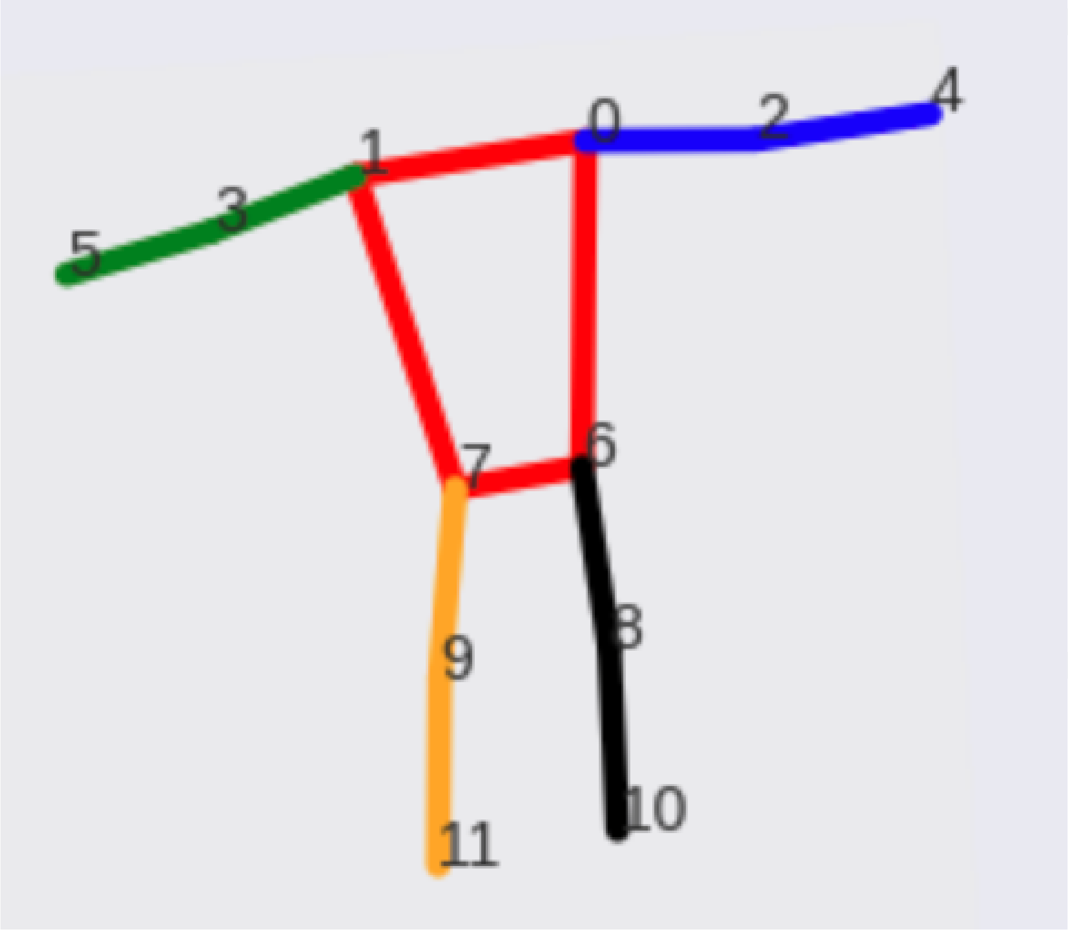
Expanding the matrix equation gives:

$$(f_{11}u_l + f_{21}v_l + f_{31})u_r + (f_{12}u_l + f_{22}v_l + f_{32})v_r + (f_{13}u_l + f_{23}v_l +$$

Equation for right epipolar line: $\boxed{a_l u_r + b_l v_r + c_l = 0}$

Similarly we can calculate epipolar line in left image for a point in right image.

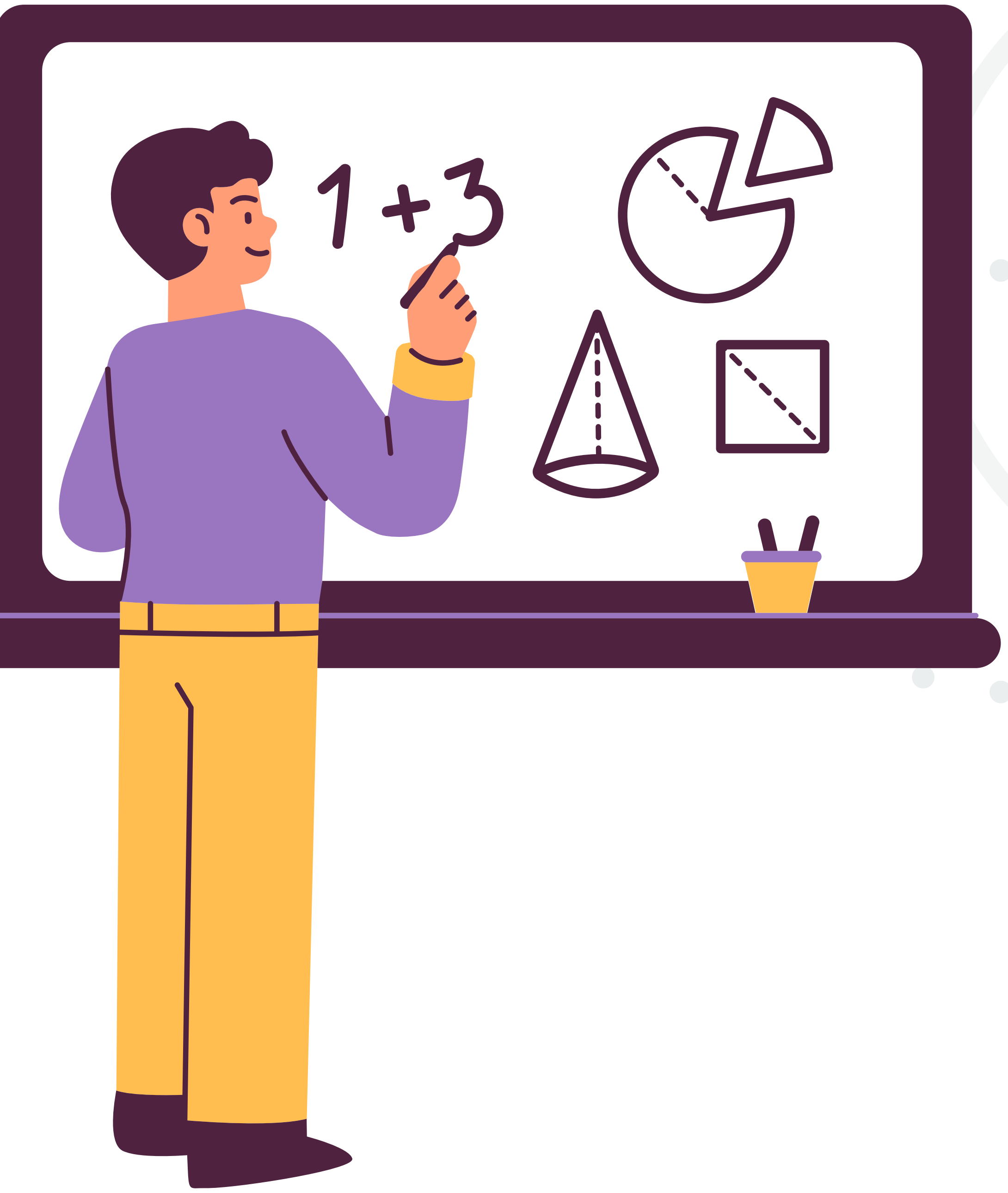
Real time 3D body pose estimation



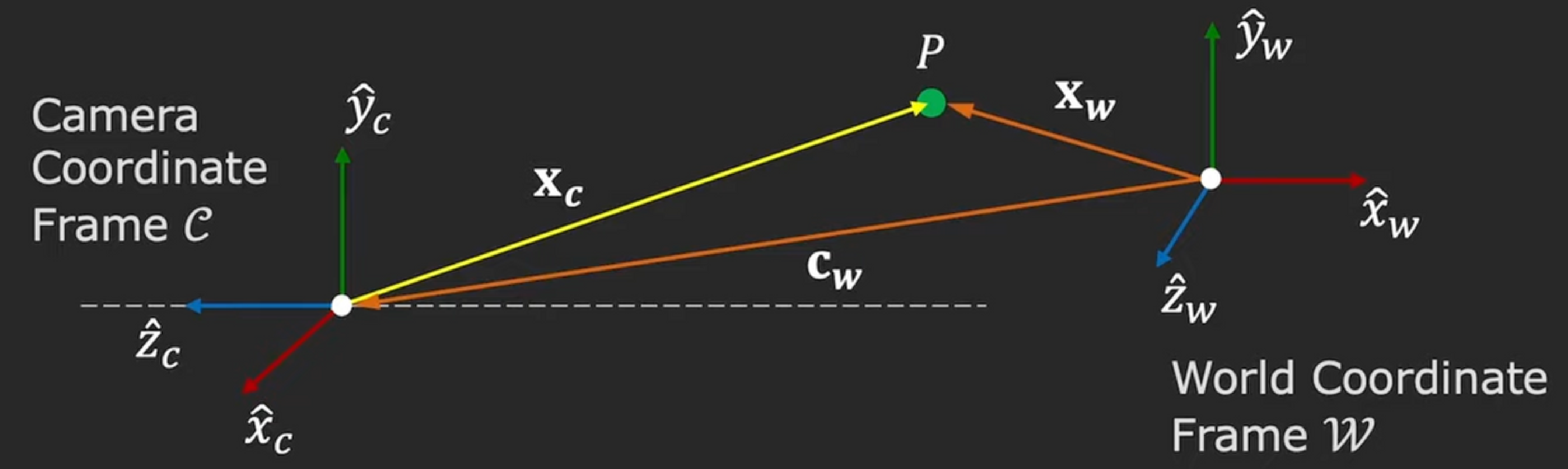
- After stereo setup and camera callibration, we use direct linear transform(DLT) to triangulate camera pixels to 3D coordinates.
- Once the keypoints are obtained from the estimator, we apply direct linear transform (DLT) through singular value decomposition (SVD).
- The 3D coordinate in each video frame is recorded
- The keypoints are indexed as shown in image.
- We obtain the projection matric by multiplying the camera matrix by the rotation and translation matrix.

3D body pose

Equations involved



World-to-Camera Transformation



Given the extrinsic parameters $(R, \mathbf{c}_w)$ of the camera, the camera-centric location of the point P in the world coordinate frame is:

$$\mathbf{x}_c = R(\mathbf{x}_w - \mathbf{c}_w) = R\mathbf{x}_w - R\mathbf{c}_w = R\mathbf{x}_w + \mathbf{t} \qquad \boxed{\mathbf{t} = -R\mathbf{c}_w}$$

$$\mathbf{x}_c = \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}$$

Equations involved

Projection Matrix P

Camera to Pixel

$$\begin{bmatrix}\tilde{u}\\ \tilde{v}\\ \tilde{w}\end{bmatrix} = \begin{bmatrix} f_x & 0 & o_x & 0\\ 0 & f_y & o_y & 0\\ 0 & 0 & 1 & 0\end{bmatrix}\begin{bmatrix}x_c\\ y_c\\ z_c\\ 1\end{bmatrix}$$

$$\tilde{\mathbf{u}} = M_{int}\,\tilde{\mathbf{x}}_c$$

World to Camera

$$\begin{bmatrix}x_c\\ y_c\\ z_c\\ 1\end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x\\ r_{21} & r_{22} & r_{23} & t_y\\ r_{31} & r_{32} & r_{33} & t_z\\ 0 & 0 & 0 & 1\end{bmatrix}\begin{bmatrix}x_w\\ y_w\\ z_w\\ 1\end{bmatrix}$$

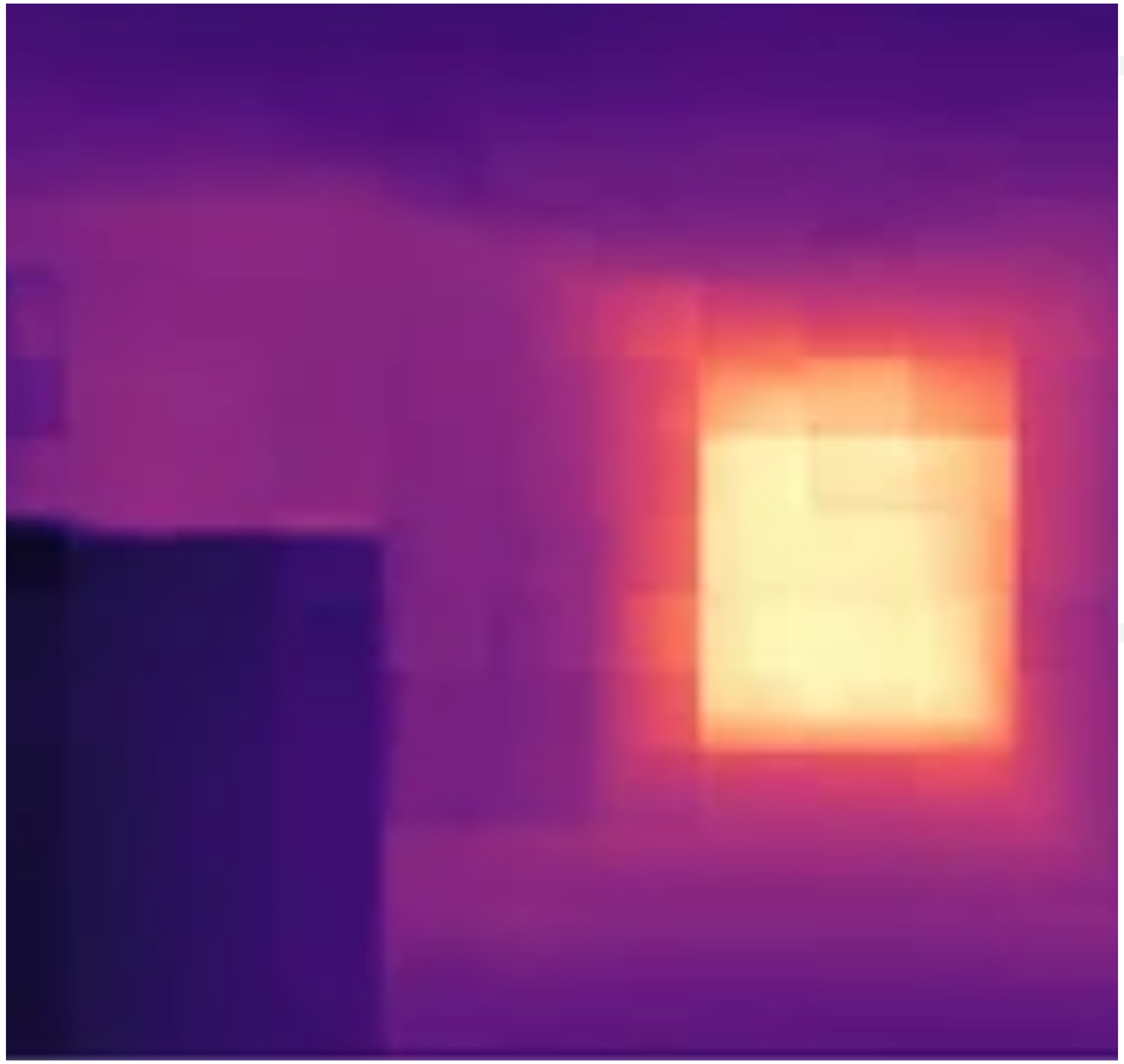
$$\tilde{\mathbf{x}}_c = M_{ext}\tilde{\mathbf{x}}_w$$

Combining the above two equations, we get the full projection matrix P:

$$\boxed{\tilde{\mathbf{u}} = M_{int}M_{ext}\,\tilde{\mathbf{x}}_w = P\,\tilde{\mathbf{x}}_w}$$

$$\begin{bmatrix}\tilde{u}\\ \tilde{v}\\ \tilde{w}\end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14}\\ p_{21} & p_{22} & p_{23} & p_{24}\\ p_{31} & p_{32} & p_{33} & p_{34}\end{bmatrix}\begin{bmatrix}x_w\\ y_w\\ z_w\\ 1\end{bmatrix}$$

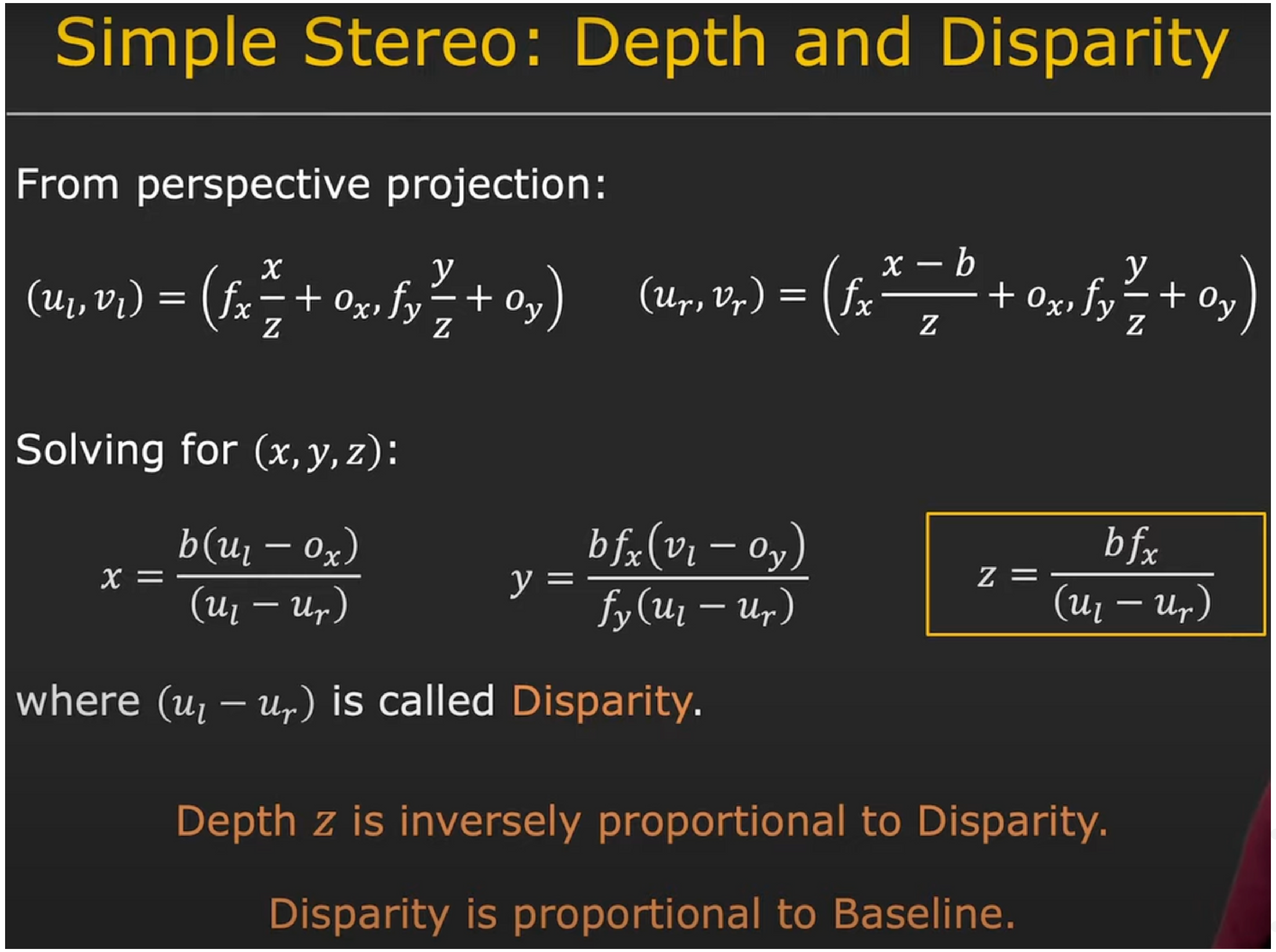
Depth Maps



- Now that the cameras are fully calibrated and rectified, they can be used to generate depth maps.
- Then, in a capture loop undistort the images using remap(),
- Convert them to grayscale with cvtColor(),
- Compute the depth map with a StereoBM object.
- When previewing the depth map, we need to scale it down to a visible range before showing it:

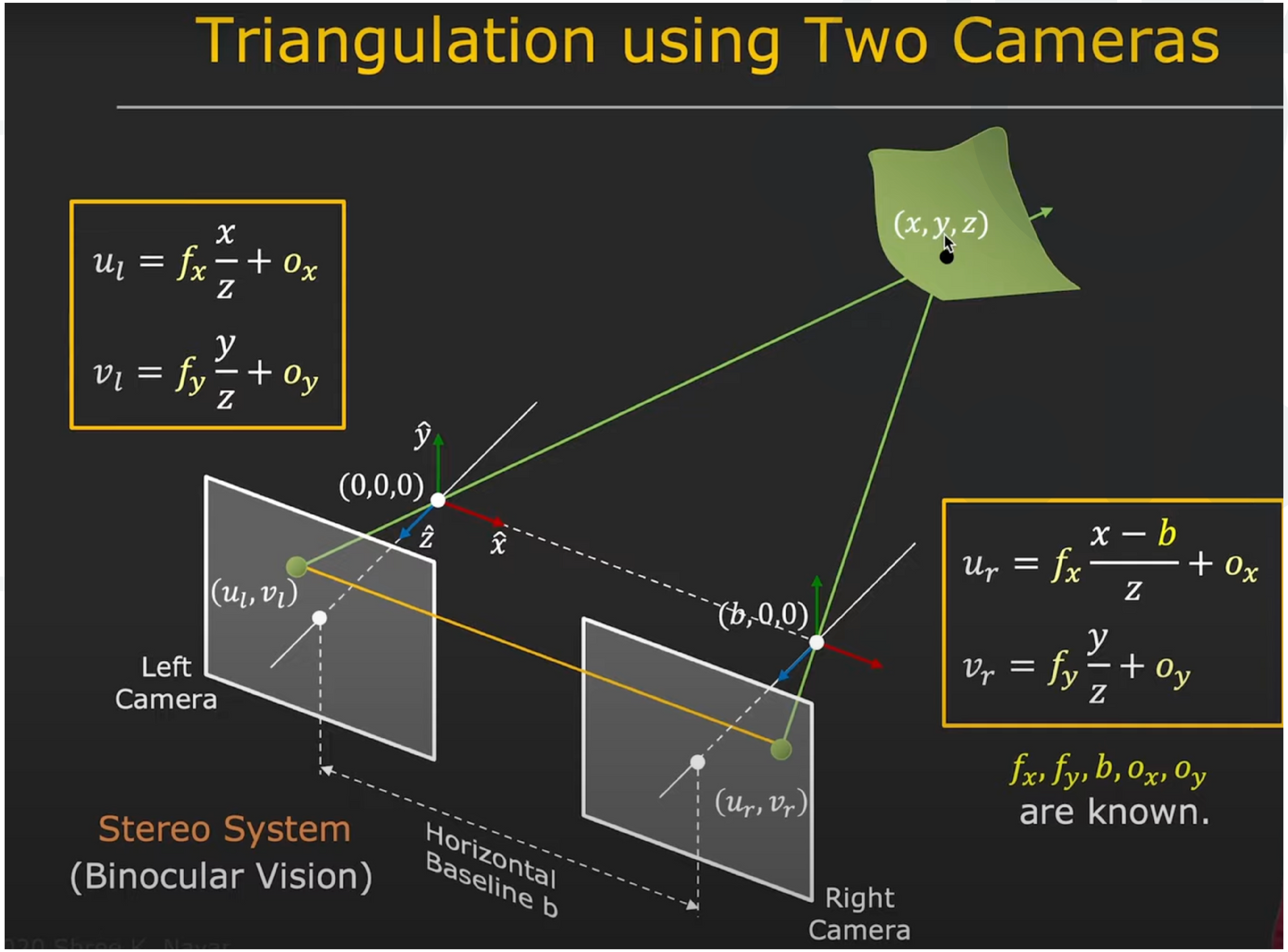
Depth Maps

Effects of Varying baseline

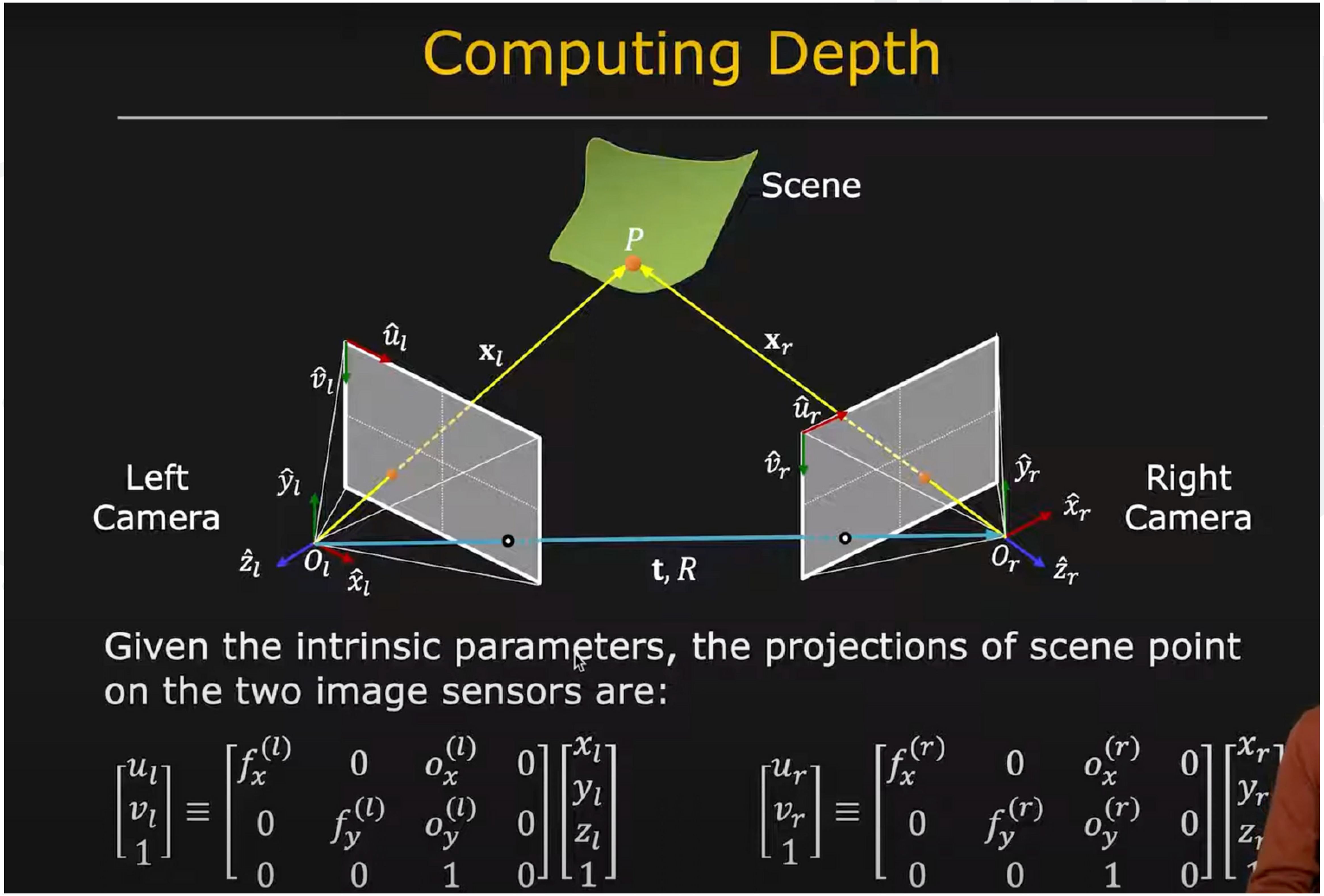


- Z is inversely proportional to baseline,
- Hence, the smaller the depth, the greater the parallax, and the closer the object has greater parallax.
- This is why the closer objects in the disparity map are darker.

Equations involved



Equations involved



Equations involved

Computing Depth: Least Squares Solution

$$\underbrace{\begin{bmatrix} u_r m_{31} - m_{11} & u_r m_{32} - m_{12} & u_r m_{33} - m_{13} \\ v_r m_{31} - m_{21} & v_r m_{32} - m_{22} & v_r m_{33} - m_{23} \\ u_l p_{31} - p_{11} & u_l p_{32} - p_{12} & u_l p_{33} - p_{13} \\ v_l p_{31} - p_{21} & v_l p_{32} - p_{22} & v_l p_{33} - p_{23} \end{bmatrix}}_{\substack{A_{4\times3} \\ \text{(Known)}}} \underbrace{\begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix}}_{\substack{\mathbf{x}_r \\ \text{(Unknown)}}} = \underbrace{\begin{bmatrix} m_{14} - m_{34} \\ m_{24} - m_{34} \\ p_{14} - p_{34} \\ p_{24} - p_{34} \end{bmatrix}}_{\substack{\mathbf{b}_{4\times1} \\ \text{(Known)}}}$$

Find least squares solution using pseudo-inverse:

$$A\mathbf{x}_r = \mathbf{b}$$

$$A^T A\mathbf{x}_r = A^T\mathbf{b}$$

$$\boxed{\mathbf{x}_r = (A^T A)^{-1} A^T\mathbf{b}}$$

Calculating depth

- Decompose the projection matrices into the camera intrinsic matrix K, and extrinsics R, t.
- Get the focal length f from the K matrix
- Compute the baseline b using corresponding values from the translation vectors t
- Compute depth map of the image using the following formula and the calculated disparity map d

Depth_Calculation

Equations involved

Simple Stereo: Depth and Disparity

From perspective projection:

$$(u_l, v_l) = \left(f_x \frac{x}{z} + o_x, f_y \frac{y}{z} + o_y\right) \qquad (u_r, v_r) = \left(f_x \frac{x-b}{z} + o_x, f_y \frac{y}{z} + o_y\right)$$

Solving for (x, y, z):

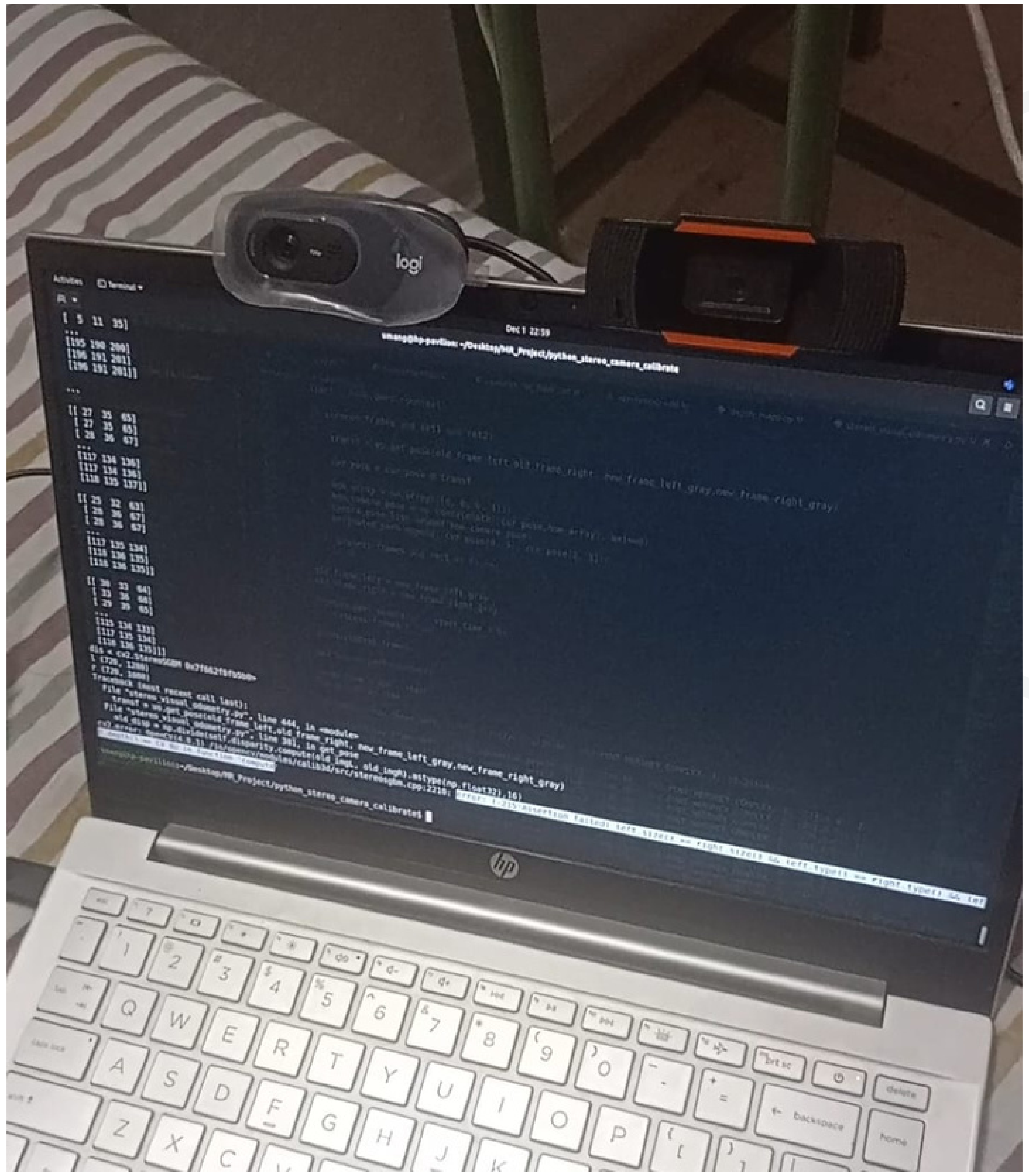
$$x = \frac{b(u_l - o_x)}{(u_l - u_r)} \qquad y = \frac{bf_x(v_l - o_y)}{f_y(u_l - u_r)} \qquad \boxed{z = \frac{bf_x}{(u_l - u_r)}}$$

where $(u_l - u_r)$ is called Disparity.

Depth z is inversely proportional to Disparity.

Disparity is proportional to Baseline.

Environment Map



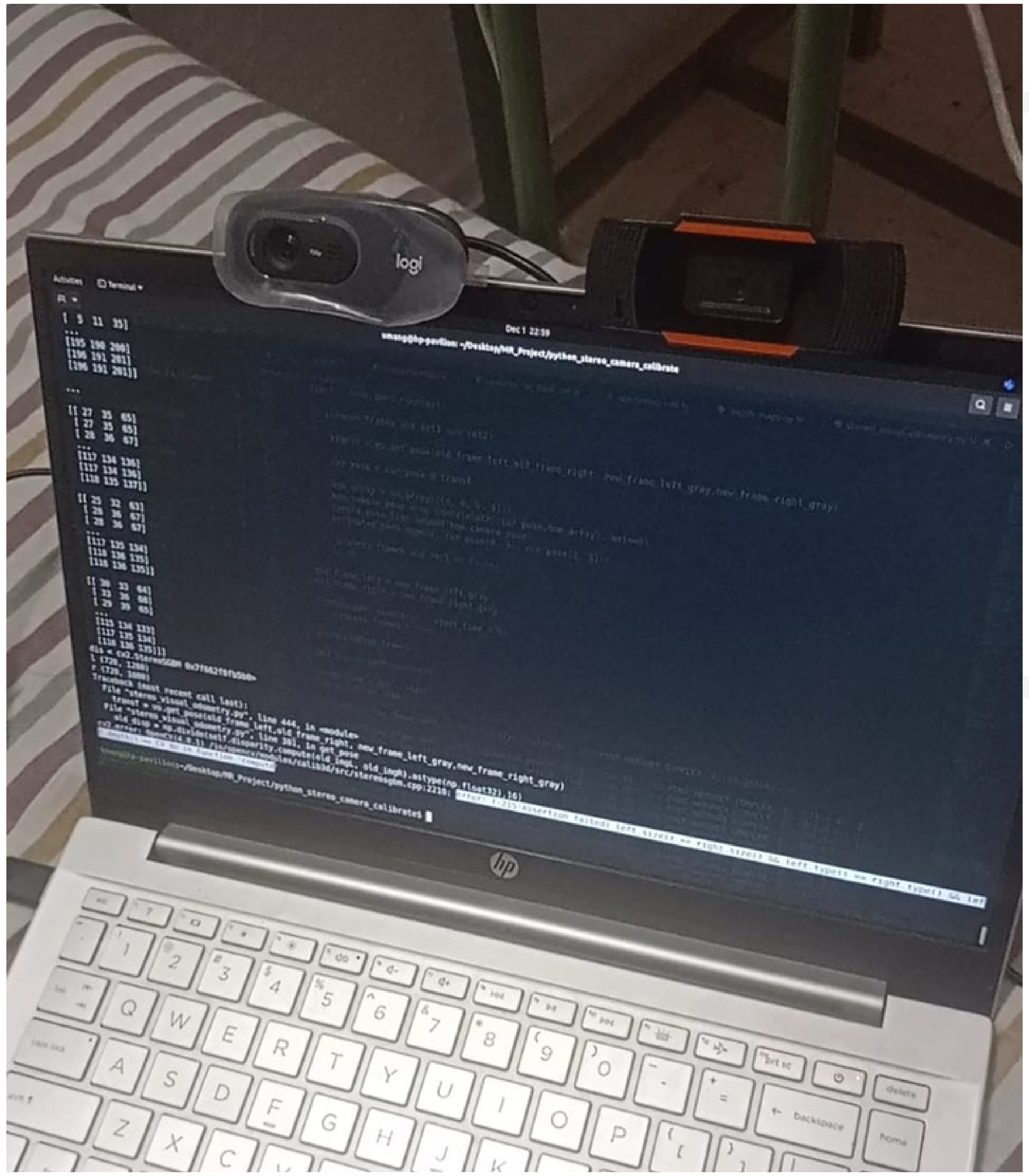
- We use StereoSGBM algorithm from OpenCV is used to calculate the disparity maps between left and right stereo images.
- Fast feature detection (**cv2.FastFeatureDetector_create()**) is used to find keypoints in the images.
- The code performs triangulation of feature points to obtain 3D points. And then we try to visualize

Map the environment

Equations involved

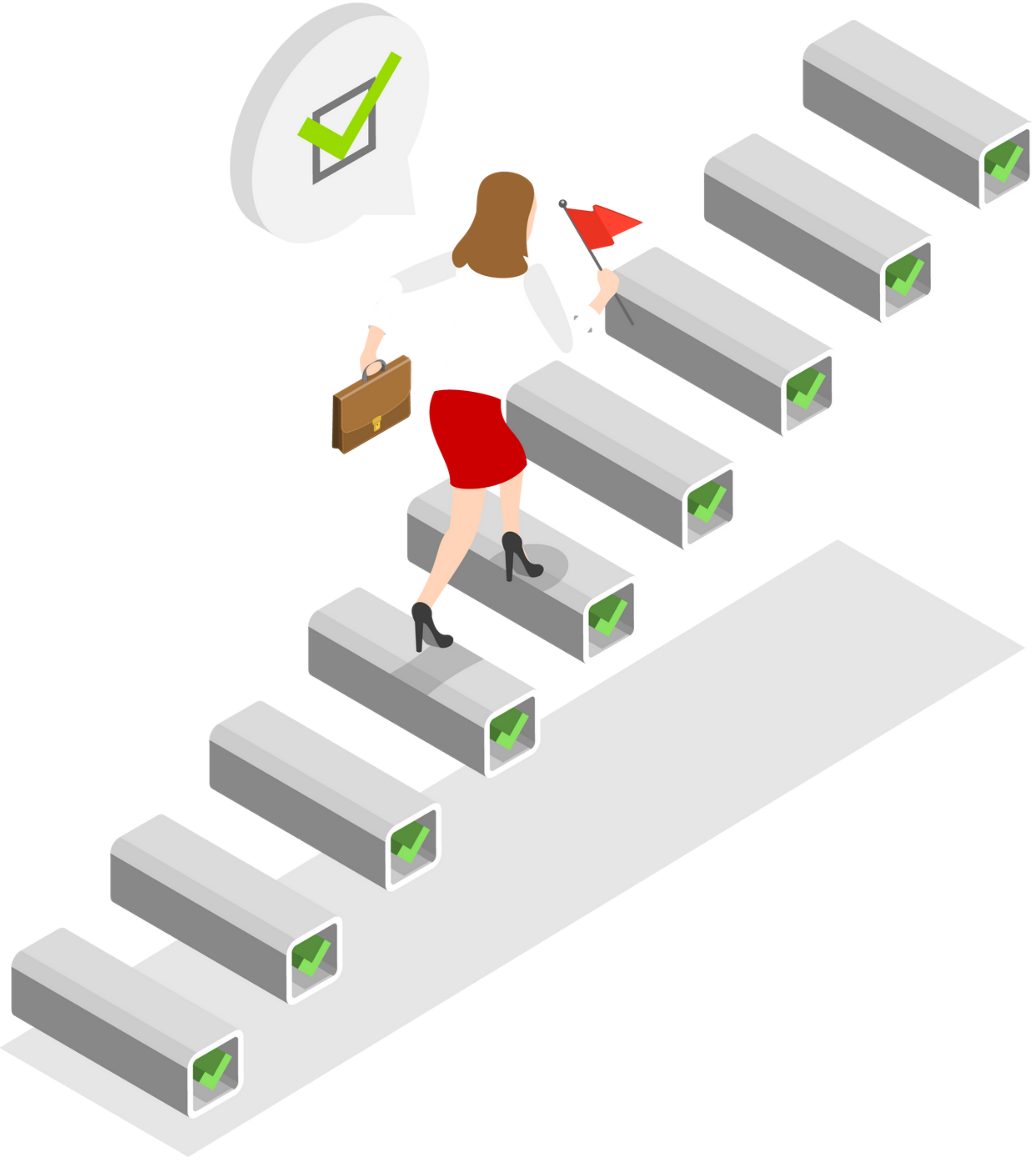
$$E = \sum_i w_i \left[\left(Rp_i + T - q_i \right) \cdot n_i \right]^2 + \left[w_{odo} \cdot \left(y - \hat{y} \right) \right]^2 .$$

Denoising with Deep Learning



- Image Enhancement i.e. removing random noise from the image is highly desired in computer vision.
- Autoencoder is an appropriate consideration specifically due to its application in Denoising
- Has great potential in the feature extraction and data component understanding
- Data is corrupted in some manner through the addition of random noise, and the model is trained to predict the original uncorrupted data.

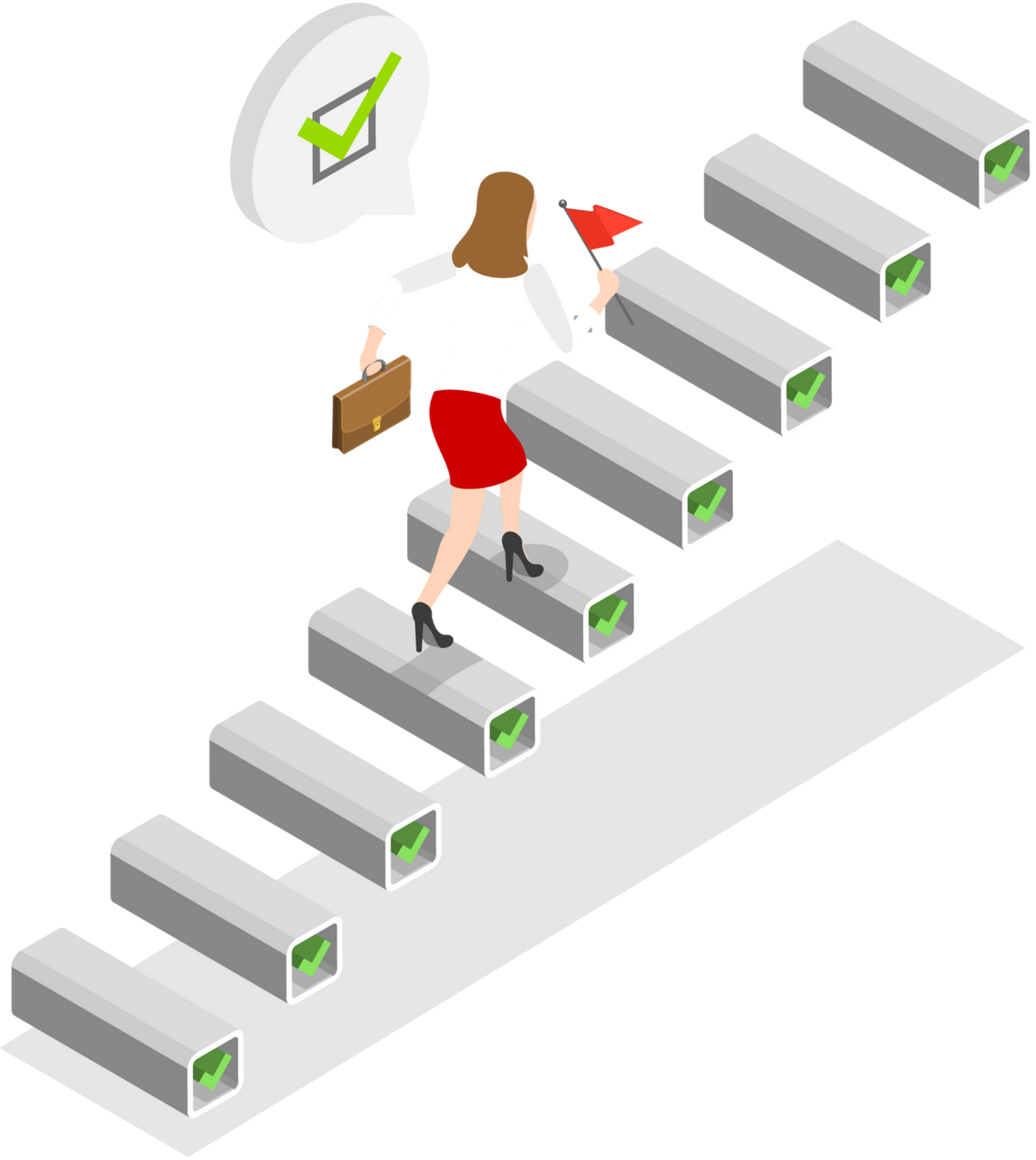
Challenges faced



- **Data Quality**: Ensuring that the captured calibration pattern frames are of high quality was challenging. Factors like lighting conditions and focus impacted the accuracy of the calibration.

- **Pattern Detection**: Detecting and accurately identifying calibration patterns in images was bit challenging.

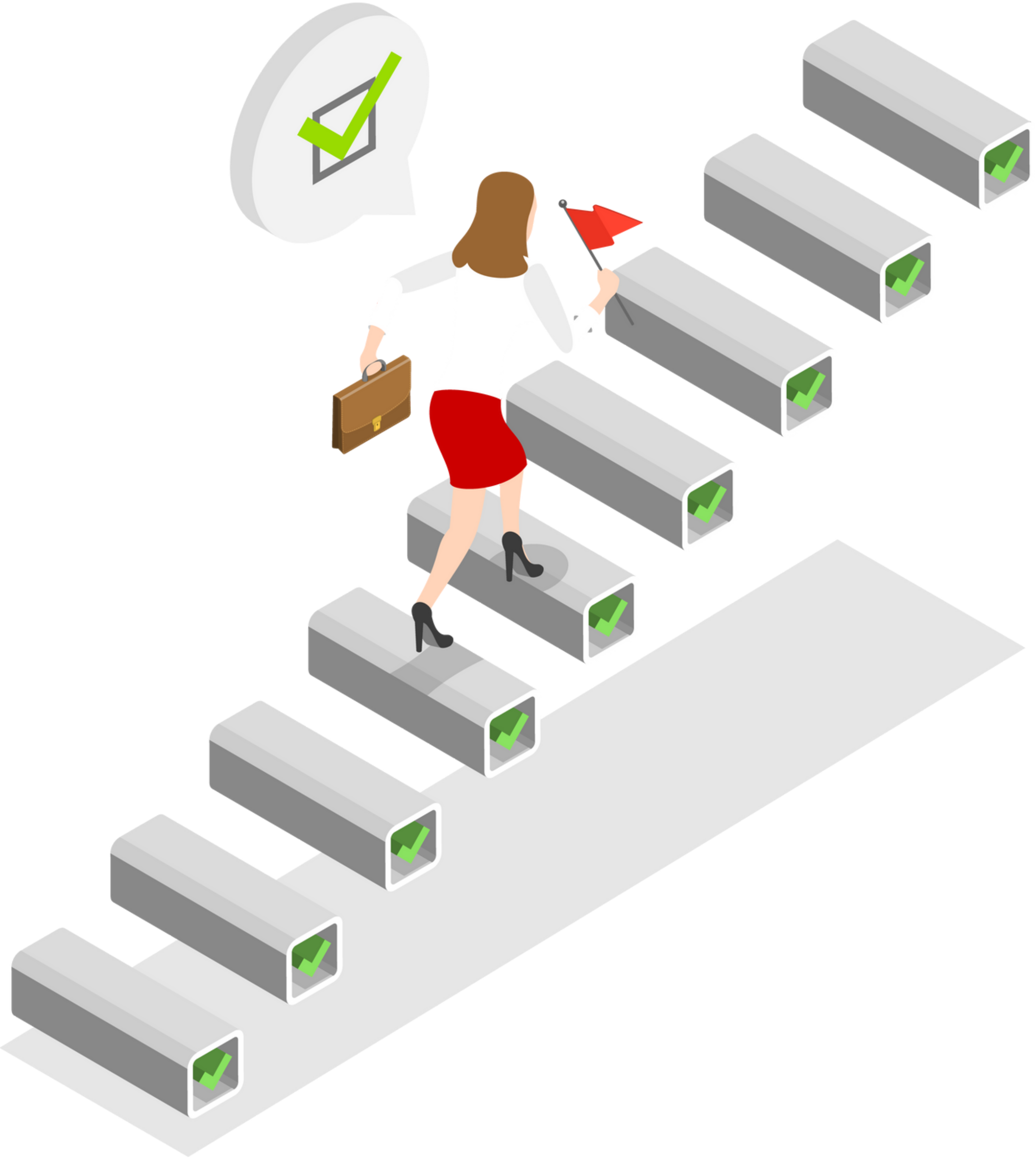
Challenges faced



- **Pattern Positioning**: Maintaining precise and consistent positioning of the calibration pattern across multiple images was essential. Misalignment introduced errors in calibration.

- **Camera Synchronization**: Ensuring that both cameras are synchronized in terms of exposure, triggering, and image capture was challenging, as unsynchronized cameras could lead to errors in depth calculations.

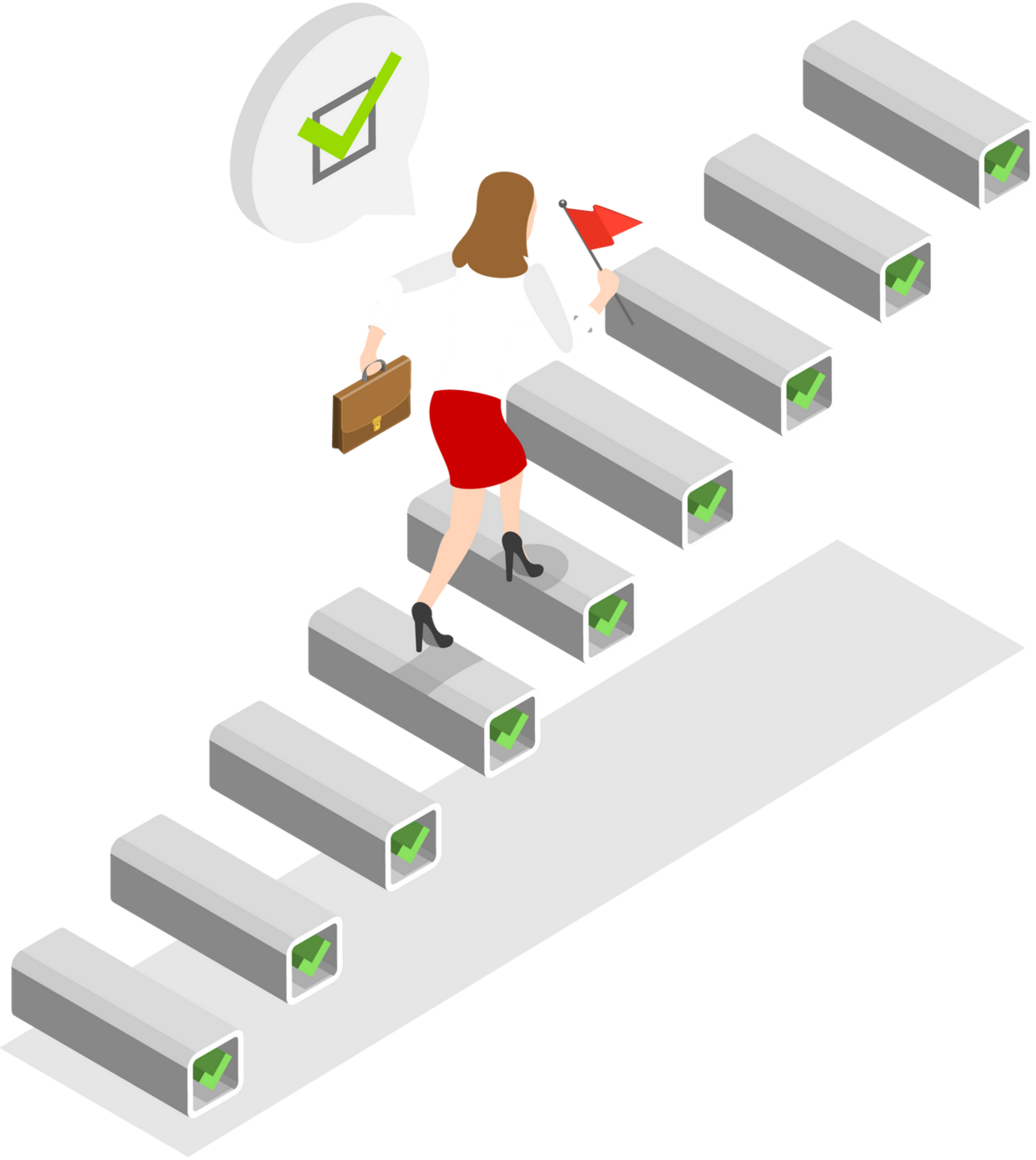
Challenges faced



- **Pattern Positioning**: Maintaining precise and consistent positioning of the calibration pattern across multiple images was essential. Misalignment introduced errors in calibration.

- **Camera Synchronization**: Ensuring that both cameras are synchronized in terms of exposure, triggering, and image capture was challenging, as unsynchronized cameras could lead to errors in depth calculations.

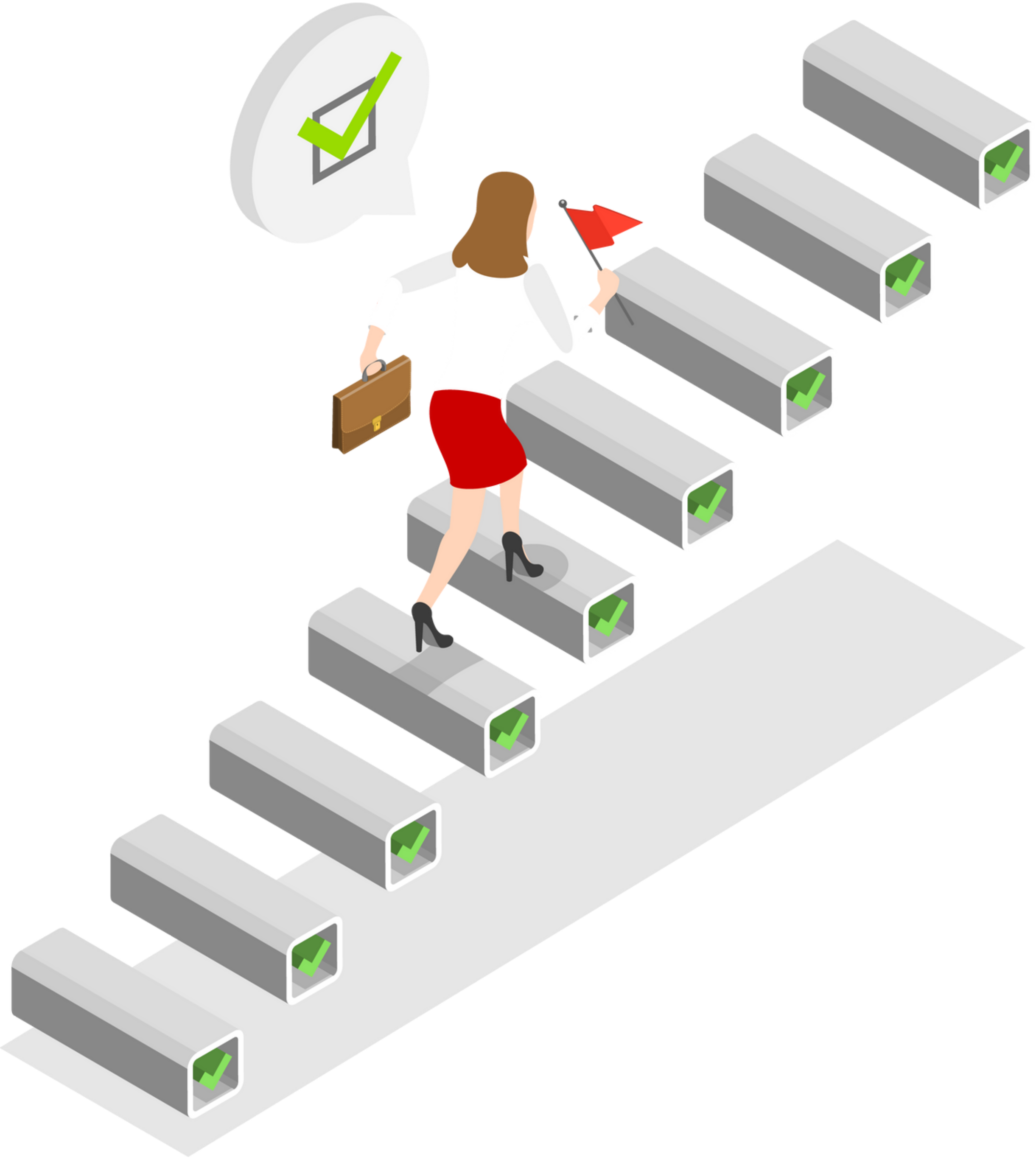
Challenges faced



- **Baseline Effect**: Varying the baseline affected the accuracy of depth perception. Larger baselines improved depth resolution but introduced challenges related to stereo correspondence.

- **Mapping the Environment**: Creating an accurate 3D map of the environment requires robust handling of the stereo disparity information. Due to different resolution of images, the task could not be achieved completely.

Challenges faced



- **Image noise:** Adding random noise to the images introduces challenges in reconstructing the images accurately.

- **Calibration Stability**: Ensure that the stereo camera setup remains stable during the capture process. Any movement or misalignment compromised the calibration and subsequent depth calculations.

Conclusion

- Stereo camera calibration is a critical process in computer vision, enabling accurate 3D reconstructions.
- Understanding the effects of varying baseline and addressing potential challenges are essential for successful calibration and depth calculation.
- Future work in 3D mapping and image enhancement holds promise for exciting applications in various fields.

Questions?

References And Literature Survey

01 Coding Over Sets for DNA Storage, Andreas Lenz , Student Member, IEEE, Paul H. Siegel , Life Fellow, IEEE, Antonia Wachter-Zeh , Member, IEEE, and Eitan Yaakobi , Senior Member, IEEE

Camera resectioning

Camera resectioning is the process of estimating the parameters of a pinhole camera model approximating the camera that produced a given photograph or video; it determines which incoming light ray is associated with each pixel on the resulting image. Basically, the process determines the...

Wikipedia / Nov 10

Get 3D coordinates from 2D image pixel if extrinsic and intrinsic parameters are known

I am doing camera calibration from tsai algo. I got intrensic and extrinsic matrix, but how can I reconstruct the 3D coordinates...

Stack Overflow

International Institute
Of Information
Technology Hyderabad

Team: Robusta

THANKS

FOR YOUR ATTENTION

Mentor: Prof. Madhava Krishna
TA: Laksh Sir

Submitted by

Umang Sharma, 2021102028
Ashish Chokhani, 2021102016

5th December 2023