COMPUTER VISION ASSIGNMENT 3

JUSTIN DE WITT* (21663904)

18 August 2022

1 QUESTION 1

1.1 Estimating The Camera Matrix

To estimate the camera matrix, we need to identify points on an image where the real-world position of each point in determinable. In our input image, we are shown a connecting wall of Lego bricks, where each brick is of known dimension. Using the known Lego dimensions, we can determine the real-world position of any point on the wall, in relation to a chosen origin. Consider the following image:

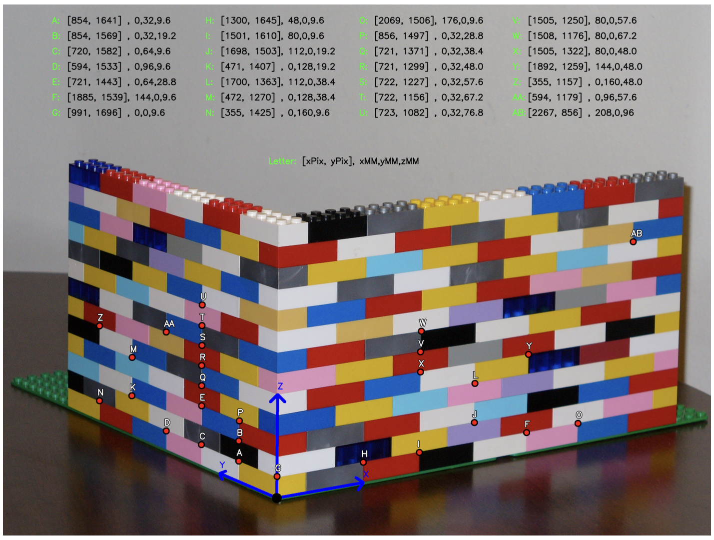


Figure 1: figure
Input Image With Determined Point Correspondencies

The dimension of each Lego brick is 32mm long, 16mm across, and 9.6mm tall. If we choose the origin to be the bottom corner where the adjacent walls connect, we can determine the position of each of the 28 points labeled in the image. Notice, each point correspondence consists of its location on the image, in terms of pixels, and its real-world location, in terms of millimeters.

From these positions we can determine an estimate for the camera matrix by solving the following system:

$$\begin{bmatrix} x_1 & y_1 & z_1 & 1 & 0 & 0 & 0 & 0 & -x_1'x_1 & -x_1'y_1 & -x_1'z_1 & -x_1' \\ 0 & 0 & 0 & 0 & x_1 & y_1 & z_1 & 1 & -y_1'x_1 & -y_1'y_1 & -y_1'z_1 & -y_1' \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n & y_n & z_n & 1 & 0 & 0 & 0 & 0 & -x_n'x_n & -x_n'y_n & -x_n'z_n & -x_n' \\ 0 & 0 & 0 & 0 & x_n & y_n & z_n & 1 & -y_n'x_n & -y_n'y_n & -y_n'z_n & -y_n' \end{bmatrix} \begin{bmatrix} p_{11} \\ p_{12} \\ p_{13} \\ p_{14} \\ \dots \\ p_{34} \end{bmatrix} = \begin{bmatrix} 0_1 \\ 0_2 \\ 0_3 \\ 0_4 \\ \dots \\ 0_{12} \end{bmatrix}$$

Notice that the matrix is not full-rank, which implies by the rank-nullity theorem that the dimension of the null space is non-zero. We can therefore solve this system using SVD, which yields $A = U\Sigma V^T$. Our vector P corresponds to the last column of V. The vector P is repacked into a 3x4 matrix, which is our projection matrix estimate. In our estimation, we obtain

$$P = \begin{bmatrix} -3.474e^{-3} & 1.836e^{-3} & 1.513e^{-6} & -4.879e^{-1} \\ 2.070e^{-4} & 3.755e^{-4} & 3.873e^{-3} & -8.729e^{-1} \\ -2.190e^{-7} & -3.147e^{-7} & 4.527e^{-8} & -4.926e^{-4} \end{bmatrix}$$

1.2 Correctness Proof

1.2.1 *Prove K is Upper-Triangular*

When performing QR-Decomposition, we obtain an orthogonal matrix $\hat{Q}$ and an upper triangular matrix $\hat{R}$. The transpose of an upper triangular matrix is lower triangular.

$$\begin{aligned} K &= W\hat{R}^T W \\ &= \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{R}^T_{11} & 0 & 0 \\ \hat{R}^T_{21} & \hat{R}^T_{22} & 0 \\ \hat{R}^T_{31} & \hat{R}^T_{32} & \hat{R}^T_{33} \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} \hat{R}^T_{31} & \hat{R}^T_{32} & \hat{R}^T_{33} \\ \hat{R}^T_{21} & \hat{R}^T_{22} & 0 \\ \hat{R}^T_{11} & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} \hat{R}^T_{33} & \hat{R}^T_{32} & \hat{R}^T_{31} \\ 0 & \hat{R}^T_{21} & \hat{R}^T_{22} \\ 0 & 0 & \hat{R}^T_{11} \end{bmatrix} \end{aligned}$$

1.2.2 *Prove R is Orthogonal*

Recall that QR decomposition yields an orthogonal matrix Q. The matrix W is orthogonal, because $WW^T = I \implies W^T = W^{-1}$. The product of two orthogonal matrices is an orthogonal matrix. Using these properties:

$$\begin{aligned} &R = W\hat{Q}^T \\ &\hat{Q}^T \text{ is orthogonal} \\ &\therefore W\hat{Q}^T \text{ is orthogonal because W is orthogonal} \end{aligned}$$

1.2.3 *Prove* $P = KR[I| - \hat{c}]$

From the previous two subsections we showed:

$$K = W\hat{R}^T W$$
$$R = W\hat{Q}^T$$

After substituting these into the expression we would like to prove:

$$
\begin{aligned}
&= (W\hat{R}^T W)(W\hat{Q}^T)[I| - \hat{\mathbf{C}}] \\
&= W\hat{R}^T I\hat{Q}^T[I| - \hat{\mathbf{C}}] \qquad \text{because WW=I} \\
&= W(\hat{R}^T\hat{Q}^T)[I| - \hat{\mathbf{C}}] \\
&= W(\hat{Q}\hat{R})^T[I| - \hat{\mathbf{C}}] \qquad \text{because } (AB)^T = B^T A^T \\
&= W((WP_{1:3})^T)^T[I| - \hat{\mathbf{C}}] \qquad \text{because } (\hat{Q}\hat{R}) = (WP_{1:3})^T \\
&= WWP_{1:3}[I| - \hat{\mathbf{C}}] \\
&= P_{1:3}[I|(P_{1:3}^{-1})\mathbf{P}_4] \qquad \text{because } \hat{\mathbf{C}} = -(P_{1:3}^{-1})\mathbf{P}_4 \\
&= [P_{1:3}|P_{1:3}^{-1}\mathbf{P}_4] \\
&= [P_{1:3}|\mathbf{P}_4] \\
&= P
\end{aligned}
$$

1.2.4 *Decomposition Uniqueness*

QR decomposition is not unique in general; There are proofs which demonstrate that the decomposition is unique if the matrix is both full rank, and contains only positive entries along its diagonal. In our calculations, the projection matrices does not exclusively contain positive entries along its main diagonal. This suggests the QR decomposition is not unique. In my opinion we would like to have a unique decomposition because the camera should have a "unique" orientation and position at the time the photo was taken, described by the rotation matrix R and the position vector $\hat{\mathbf{C}}$. I hypothesise that the decomposition is unique up until a constant. Uniqueness is important in this scenario because we obtain both the intrinsic and extrinsic properties of the camera based on the dissection of the projection matrix. We assume the camera has a fixed orientation and position in space at the time the photo was taken, and the only way that the position, orientation, and intrinsic properties of the camera can be fixed, is if the decomposition is unique.

1.3 Decomposing The Camera Matrix into Intrinsic and Extrinsic Components

The first three columns of the projection matrix is somewhat similar to a "change of basis" matrix, which translates a point written in the real-world coordinate system (mm), to a point in the image coordinate system (pixels). The projection matrix can be decomposed into the product of two matrices, one which translates a real-world location to a location in the camera's internal coordinate system (M_{ext}), and one which translates a point from the camera's internal coordinate system to image coordinates (M_{int}). To clarify:

$$
\mathbf{x}_{pix} = \begin{bmatrix} \alpha_x & s & X_0 & 0 \\ 0 & \alpha_y & Y_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}
\begin{bmatrix} r_{11} & r_{12} & r_{13} & c_x \\ r_{21} & r_{22} & r_{23} & c_y \\ r_{31} & r_{32} & r_{33} & c_z \\ 0 & 0 & 0 & 1 \end{bmatrix}
\begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} = P\mathbf{x}_w
$$

Where the left matrix is the intrinsic matrix, and the right is the extrinsic matrix. In the intrinsic matrix α_x and α_y specify the cameras pixel density, X_0 and Y_0 are the image origin coordinates, and s is a skewness parameter. The extrinsic matrix contains the rotation matrix r, and the camera's real-world position **c**. The upper-triangular matrix, k, corresponding to $M_{int_{1:3}}$, is referred to as the *calibration matrix*. The *rotation matrix*, r, corresponding to $M_{ext_{1:3,1:3}}$ is an orthonormal matrix. We can decompose a matrix into the product of a diagonal matrix and an orthonormal matrix using QR decomposition. Because we know the first three columns of the projection matrix are a product of a diagonal matrix (k) and an orthonormal matrix (r), we use QR decomposition to determine these components. We solve for the cameras position vector **c** using the last column of the projection matrix.

$$\mathbf{c} = k^{-1} \begin{bmatrix} p_{14} \\ p_{24} \\ p_{34} \end{bmatrix}$$

The obtained decomposition was found to be:

$$K = \begin{bmatrix} 1.010e^4 & -2.570e^1 & 1.228e^3 \\ 0 & 1.009e^4 & 7.920e^1 \\ 0 & 0 & 1 \end{bmatrix} R = \begin{bmatrix} -0.822 & 0.570 & -0.011 \\ 0.058 & 0.103 & 0.993 \\ -0.567 & -0.815 & 0.117 \end{bmatrix} \mathbf{c} = \begin{bmatrix} -687.145 \\ -1034.971 \\ 362.482 \end{bmatrix}$$

2 QUESTION 2

2.1 Estimating the Camera Matrix

This photo contains an image of the same structure taken from another angle. The identified points are shown in the following image:

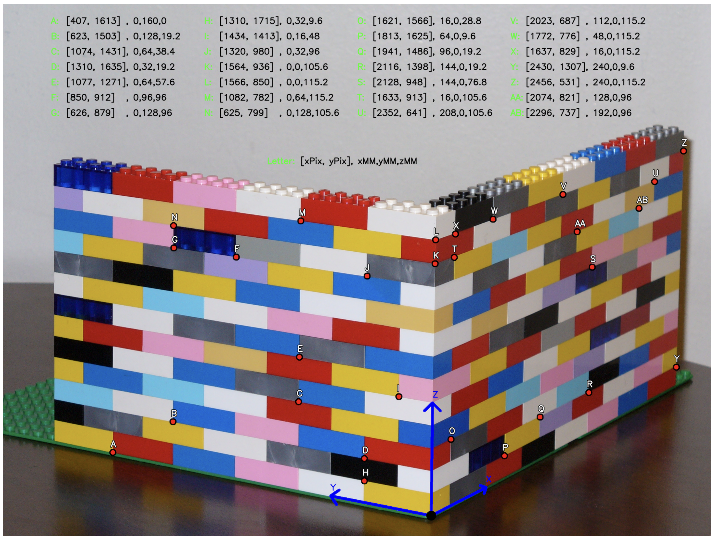


Figure 2: figure
Input Image With Determined Point Correspondencies

Using the methods described in Question 1, we obtain the following camera matrix P:

$$P = \begin{bmatrix} 2.289e^{-3} & -2.879e^{-3} & -2.903e^{-5} & 6.429e^{-1} \\ -3.754e^{-4} & -2.793e^{-4} & -3.632e^{-3} & 7.659e^{-1} \\ 3.161e^{-7} & 1.978e^{-7} & -6.171e^{-8} & 4.152e^{-4} \end{bmatrix}$$

which decomposes into the following calibration (K), and rotation matrices (R), along with the following position vector (**c**):

$$K = \begin{bmatrix} 9.670e^3 & -3.060e^0 & 1.091e^3 \\ 0 & 9.683e^3 & 3.517e^2 \\ 0 & 0 & 1 \end{bmatrix} R = \begin{bmatrix} 0.532 & -0.847 & 0.010 \\ -0.133 & -0.095 & -0.987 \\ 0.836 & 0.523 & -0.163 \end{bmatrix} \mathbf{c} = \begin{bmatrix} -923.999 \\ -514.657 \\ 345.939 \end{bmatrix}$$

We can easily determine the distance between the cameras at the time that each of these images was captured by simply subtracting the camera position vectors between both images, and calculating the norm of the difference vector.

$$\|\mathbf{c}_1 - \mathbf{c}_2\| = 571.927 \text{ mm}$$

We can also determine the change in angle between the camera positions at the time the images were taken by comparing the basis vectors of each camera's rotation matrix. Recall that the rotation matrix is orthonormal, implying all vectors are length 1, and all vectors are orthogonal. The direction of the principle axis ($\hat{\mathbf{z}}$) is the third row of the rotation matrix.

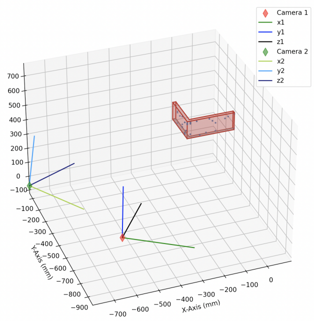
$$\hat{\mathbf{z}}_1 = \begin{bmatrix} -0.567 \\ -0.815 \\ 0.117 \end{bmatrix} \quad \hat{\mathbf{z}}_2 = \begin{bmatrix} 0.836 \\ 0.523 \\ -0.163 \end{bmatrix}$$

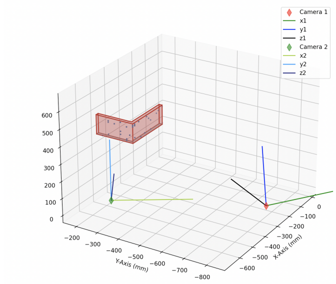
We can calculate the angle between these vectors using dot product, defined as:

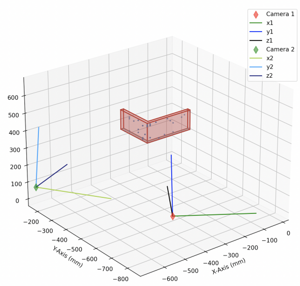
$$(\hat{\mathbf{z}}_1 \cdot \hat{\mathbf{z}}_2) = \|\hat{\mathbf{z}}_1\|\|\hat{\mathbf{z}}_2\| \cos\theta \implies \theta = 28.279^\circ$$

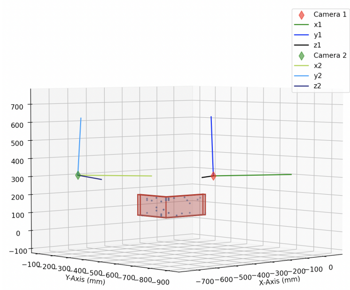
2.2 Modeling The Points and Cameras in 3D

As mentioned, the vector $\mathbf{c}_1$ and $\mathbf{c}_1$ contains the real-world location of cameras 1 and 2 respectively. We can determine the principle axis of the cameras by observing the corresponding rotation matrix, where $\hat{\mathbf{x}}_1$, $\hat{\mathbf{y}}_1$, $\hat{\mathbf{z}}_1$, $\hat{\mathbf{x}}_2$,$\hat{\mathbf{y}}_2$, and $\hat{\mathbf{z}}_2$ are the vectors associated with each axis. The "hat" denotes unit length, as the norm of each vector is 1mm. We can plot the real-world coordinates of the 28 identified points, along with the obtained axis and camera positions with the help of 3D-matplotlib. Consider the images:









Notice the cameras principle axis (z) point towards the image plane. It also appears as if the angle between cameras, and camera distance calculations from the previous section are accurate.

3 QUESTION 3

3.1 The Image of the World Coordinate System (Theory)

The columns of the projection matrix (p) can be used to determine the axis of the real-world coordinate system. To prove the statement in general, consider the ideal points in the x, y, and z directions:

$$X_{id} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} Y_{id} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} z_{id} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

Recall from Section 1; The projection matrix acts as a "change of basis" from real-world coordinates to image coordinates (pixels). That is- we can transform a point from the real-world coordinate system, to the image coordinate system, by simply multiplying the vector with the projection matrix P. Notice: if we perform this multiplication with the ideal points (points located at an infinite distance in the specified direction), we obtain the columns of the projection matrix.

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix}$$

$$\mathbf{p}_1 = P\mathbf{X}_{id} \quad \mathbf{p}_2 = P\mathbf{Y}_{id} \quad \mathbf{p}_3 = P\mathbf{Z}_{id}$$

To determine the location of the real-world origin, in terms of image coordinates, we can homogenise the real-world origin vector, and multiply with the projection matrix. This yields the final column of the projection matrix.

$$\text{Origin}_{img} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} p_{14} \\ p_{24} \\ p_{34} \\ p_{44} \end{bmatrix} = \mathbf{p}_4$$

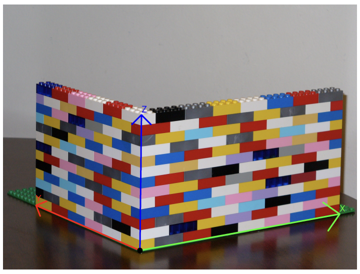
We can simply de-homogenise to obtain the origin location according to the image basis.

3.2 The Image of the World Coordinate System (Demonstration)

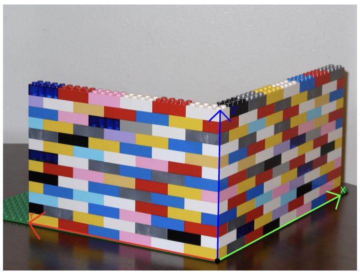
Notice, when we map a point from the real-world basis, to a point in the image basis, we need to homogenise the real-world coordinate. Consider mapping real-world the vector $[X, Y, Z]$ to its corresponding image location. The mapping is described by:

$$\begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} p_{11}X + p_{12}Y + p_{13}Z + p_{14} \\ p_{21}X + p_{22}Y + p_{23}Z + p_{24} \\ p_{31}X + p_{32}Y + p_{33}Z + p_{34} \end{bmatrix} = \begin{bmatrix} p_{11}X + p_{12}Y + p_{13}Z \\ p_{21}X + p_{22}Y + p_{23}Z \\ p_{31}X + p_{32}Y + p_{33}Z \end{bmatrix} + \begin{bmatrix} p_{14} \\ p_{24} \\ p_{34} \end{bmatrix}$$

This result demonstrates that the effect of adding the homogeneous component (the 4th dimension) adds the origin location to the mapped component. To summarize, the image location of the vector $[X, Y, Z]$ is calculated based in relation to the location of the homogeneous representation of the real-world origin. We can use this result to determine the image coordinates of the real-world X, Y and Z axis.
To determine each axis, we add the homogenised axis to the homogenised origin ($\mathbf{p_i} + \mathbf{p_4}$ for $i \in [1, 2, 3]$). This sum is a homogenised representation of the real-world axis. We can scale this sum term by any factor k, where the vanishing point is associated with $k = \infty$. The results are shown in the following images:



Camera 1 Perspective



Camera 2 Perspective

4 QUESTION 4

4.1 Epipolar Geometry

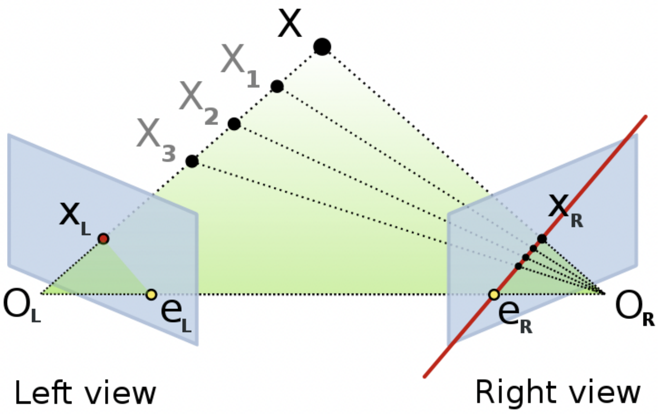
Throughout the previous questions we have described the movement of the camera between photos as “camera 1” and “camera 2”. The location of camera 2 in the image captured by camera 1 (and vice-versa) is of interest. This point is regarded as the *epipole*. It is determined that the two de-homogenised epipoles are:

$$\begin{bmatrix}-70225125\end{bmatrix} \quad \text{and} \quad \begin{bmatrix}-15784 - 730\end{bmatrix}$$

Notice that the locations of these points (in pixels) are not found within the boundaries of the images. We should regard the epipole as being the projection of the other camera’s centre onto the infinite extension of the image plane. The coordinates rest anywhere on this plane, which will likely be outside the borders of the image itself. If the principal axes of the two cameras are close to parallel, the epipoles move further away from the visible image content. In our findings we conclude the principal axes of the cameras deviate by approximately 28 degrees, which explains why the epipoles are outside the image borders.

4.2 Determining Point Pairs Between Images

From the projection matrices of each camera, we can determine (to an extent) where a point in one image resides in the other image. The location of a point in one image, is mapped to a line in the other image. The corresponding point is guaranteed to rest on the obtained line.



As illustrated in the figure, the left camera (O_L) observes a single real-world point **X**, and cannot distinguish between points which lie on the ray connecting the camera, and $\mathbf{X}$. The camera on the right can distinguish between the points $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_3$. If the points $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_3$ are projected onto the image plane of the right camera, we obtain a line. This suggests that each point observed in the real world by camera 1, can be mapped to a series of points (line) observed by the other camera. We demonstrate this result in the following images, where camera 1 identifies the corner points of our object, which correspond to lines in the image captured by camera 2:

4.3 The Fundamental Matrix

The fundamental matrix determined from the projection matrices of camera 1 and camera 2 is given by:

$$F = \begin{bmatrix} 1.555e^{-6} & -2.724e^{-5} & 4.646e^{-3} \\ 1.027e^{-4} & 1.361e^{-5} & 1.631e^{0} \\ 9.625e^{-2} & -1.914e^{0} & 1.211e^{2} \end{bmatrix}$$

If points from the photo taken by camera 2, written in image coordinates, are multiplied by the fundamental matrix F, the equation of a line written in $ax + by + c = 0$ form is obtained. The line is subsequently plotted on image 2. To obtain the corresponding line in image 1, we simply connect the point and the epipole of camera 1.

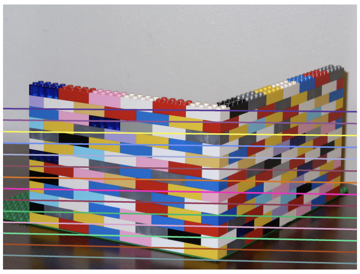


Figure 3: Image 1 Z-Axis Features

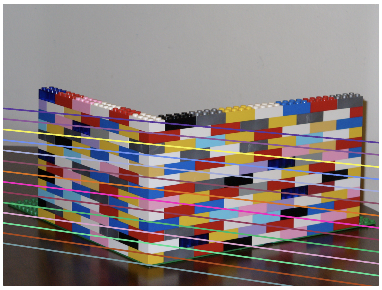


Figure 4: Image 2 Z-Axis Features