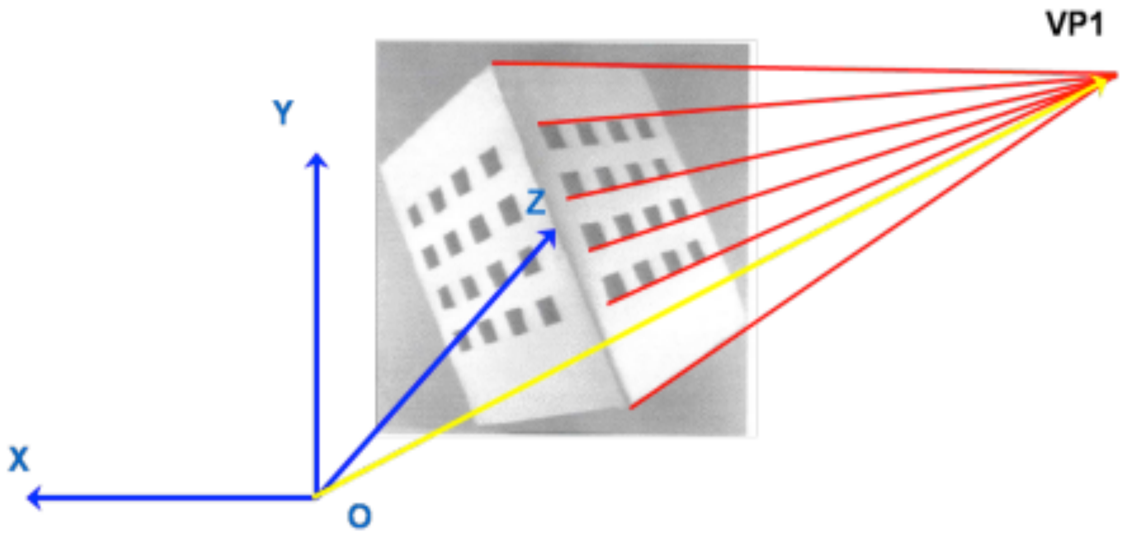
ASSIGNMENT – 3

CSc I6716
Computer Vision

Name – Jatin Jain
EMPLID # 24145250
PhD (Electrical Engineering)

Q.1 (Camera Models- 20 points) Prove that the vector from the viewpoint of a pinhole camera to the vanishing point (which is a point on the image plane) of a set of 3D parallel lines in space is parallel to the direction of that set parallel lines. Please show steps of your proof.
*Hint**:** You can either use geometric reasoning or algebraic calculation.*
If you choose to use geometric reasoning, you can use the fact that the projection of a 3D line in space is the intersection of its "interpretation plane" with the image plane. Here the interpretation plane (IP) is a plane passing through the 3D line and the center of projection (viewpoint) of the camera. Also, the interpretation planes of two parallel lines intersect in a line passing through the viewpoint, and the intersection line is parallel to the parallel lines.
If you select to use algebraic calculation, you may use the parametric representation of a 3D line: P = P0 +tV, where P= $(X,Y,Z)^T$ is any point on the line (here ^{T} denote for transpose), P0 = $(X0,Y0,Z0)^T$ is a given fixed point on the line, vector V = $(a,b,c)^T$ represents the direction of the line, and t is the scalar parameter that controls the distance (with sign) between P and P0.

Ans. In pinhole camera the focal length varies by given ratio of the image dimension to the distance of the pinhole from the image. Due to perspective, all parallel lines in 3D space appear to meet in a point on the image - the vanishing point, which is the common intersection of all the image lines



<u>Geometric Reasoning</u>
The interpretation plane (IP) is a plane passing through the 3D line and the center of projection (viewpoint) of the camera. Also, the interpretation planes of two parallel lines intersect in a line passing through the viewpoint, and the intersection line is parallel to the parallel lines.

The vector x is also parallel to the parallel lines since vanishing point and O are on the vanishing point. So, Vector OX is also parallel.

Proof: Vanishing Point V and the O (Viewing Point) should be on interpretation plane. So, the vector OX should also be on interpretation plane. So, the vector and the parallel lines cannot be skew lines.
If not parallel, then it will have an intersection point with the ground plane. Hence parallel lines form an intersection point.

ALGEBRAIC REASONING

3D line: P = P0 +tV
where P= $(X,Y,Z)^T$
P0 = $(X0,Y0,Z0)^T$
vector V = $(a,b,c)^T$
t is the scalar parameter

Using x = f(X/Z) the vanishing point of it's image.
So here X = P

$$\lim_{t\to\pm\infty} \mathrm{p} = f\left(\frac{\mathrm{P0}\ +\mathrm{tV}}{\mathrm{Pz}\ +\mathrm{tVz}}\right)$$

$$\mathrm{F}\left(\frac{\mathrm{P}}{\mathrm{Pz}}\right) = \begin{pmatrix} Px/Pz \\ Py/Pz \\ 1 \end{pmatrix}$$

Q.2 (Camera Models- 20 points) Show that relation between any image point (xim, yim)^T (in the form of $(x1,x2,x3)^T$ in projective space) of a planar surface in 3D space and its corresponding point $(Xw, Yw, Zw)^T$ on the plane in 3D space can be represented by a 3×3 matrix. You should start from the general form of the camera model $(x1,x2,x3)^T = M_{int}M_{ext}(Xw, Yw, Zw, 1)^T$, where the image center (ox, oy), the focal length f, the scaling factors(sx and sy), the rotation matrix R and the translation vector T are all unknown. Note that in the course slides and the lecture notes, I used a simplified model of the perspective project by assuming ox and oy are known and sx = sy =1, and only discussed the special cases of a plane. So you cannot directly copy those equations I used. Instead you should use the general form of the projective matrix, and the general form of a plane $n_x X_w + n_y Y_w + n_z Z_w = d$.

Ans $n_x X_w + n_y Y_w + n_z Z_w = d$
$n^T P_w = d$
where n = parameters of the plane
P_W= 3D point on the plane = $(Xw, Yw, Zw)^T$
Zw can be considered as a function of Xw and Yw
Thus Zw = 0 and on ground plane Pw = $(Xw, Yw, 0, 1)^T$
3D point (Xw, Yw, Zw) is imaged as a 2D point (Xw, Yw)

The relation between image plane to camera

$$(X, Y) = \left(\frac{fx}{Z}, \frac{fy}{Z}\right)$$

The relation b/w image plane + image frame:

$$(X, Y) = \left(-(x_{im} - O_x) s_x, -(y_{im} - O_y) s_y\right)$$

Thus, the below equation:

$$\left(\frac{fx}{Z}, \frac{fy}{Z}\right) = \left[-(x_{im} - O_x) S_x, -(y_{im} - O_y) s_y\right]$$

$$\begin{bmatrix} x_{im} \\ y_{im} \end{bmatrix} = \left\{ \begin{matrix} \left[-\frac{fx}{S_x Z} + O_x\right] \\ \left[-\frac{fY}{S_y Z} + O_y\right] \end{matrix} \right]$$

$$= \left\{ \begin{matrix} \left(\frac{-f_x X + O_x Z}{Z}\right) \\ \left(\frac{-f_y Y + O_y Z}{Z}\right) \end{matrix} \right\}$$

$$\frac{f}{S_x} = f_x, \quad \frac{f}{S_y} = f_y.$$

Camera Model Matrix (Image frame)

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} = \begin{pmatrix} -f_x & O_x & O_x \\ O_x & -f_y & O_y \\ O_x & O_y & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

World Coordinate System to Camera Coordinate System

$$P = RP_w + T$$

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \begin{pmatrix} X_w \\ Y_w \\ Z_w \end{pmatrix} + \begin{pmatrix} T_x \\ T_y \\ T_z \end{pmatrix}$$

In camera model, we convert it to 3x4 matrix.

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & T_x \\ r_{21} & r_{22} & r_{23} & T_y \\ r_{31} & r_{32} & r_{33} & T_z \end{pmatrix} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix}$$

Combining the image matrix to camera matrix & the camera matrix to world matrix

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} = \begin{pmatrix} -f_x & 0 & O_x \\ 0 & -f_y & O_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & r_{13} & T_x \\ r_{21} & r_{22} & r_{23} & T_y \\ r_{31} & r_{32} & r_{33} & T_z \end{pmatrix} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} x_{im} \\ y_{im} \end{pmatrix} = \left\{ \begin{array}{c} \dfrac{(-f_x r_{11} + O_x r_{31}) X_w + (-f_x r_{12} + O_x r_{32}) Y_w + (-f_x r_{13} + O_x r_{33}) Z_w - f_x T_x + O_x T_z}{r_{31} X_w + r_{32} Y_w + r_{33} Z_w + T_z} \\[2em] \dfrac{(-f_y r_{21} + O_x r_{31}) X_w + (-f_y r_{12} + O_y r_{32}) Y_w + (-f_y r_{13} + O_y Y_{33}) Z - f_y T_y + O_y T_z}{r_{31} X_w + r_{32} Y_w + r_{33} Z_w + T_z} \end{array} \right\}$$

Assuming given equation below is -(1)

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{bmatrix} -f r_{11} & -f r_{12} & -f T_x \\ -f r_{21} & -f r_{22} & -f T_y \\ r_{31} & r_{32} & T_z \end{bmatrix} \begin{pmatrix} X_w \\ Y_w \\ 1 \end{pmatrix}$$

$n_z = 1$, so the plane equation will be.

$$Z_w = d - n_x X_w - n_y Y_w$$

Now using the above equation, we get Mas-

$$M = \begin{bmatrix} -f r_{11} & -f r_{12} & -f r_{13} & -f T_x \\ -f r_{21} & -f r_{22} & -f r_{23} & -f T_y \\ r_{31} & r_{32} & r_{33} & T_z \end{bmatrix}$$

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{bmatrix} -f(r_{11} - n_x r_{13}) & -f(r_{12} - n_y r_{13}) & -f(d r_{13} + T_x) \\ -f(r_{21} - n_x r_{23}) & -f(r_{22} - n_y r_{23}) & -f(d r_{23} + T_y) \\ (r_{31} - n_x r_{33}) & (r_{32} - n_y r_{33}) & d r_{33} + T_z \end{bmatrix} \begin{pmatrix} X_w \\ Y_w \\ 1 \end{pmatrix}$$

$$M_{ext} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & T_x \\ r_{21} & r_{22} & r_{23} & T_y \\ r_{31} & r_{32} & r_{33} & T_z \end{bmatrix}$$

$$= \begin{bmatrix} R_1^T & T_x \\ R_2^T & T_y \\ R_3^T & T_z \end{bmatrix}$$

M_{ext} includes all the extrinsic parameters.

M_{int} = all intrinsic parameters.

$$M_{in} = \begin{bmatrix} -f_x & 0 & O_x \\ 0 & -f_y & O_y \\ 0 & 0 & 1 \end{bmatrix}$$

Combining M_{ext} & M_{int} in eq-(1)

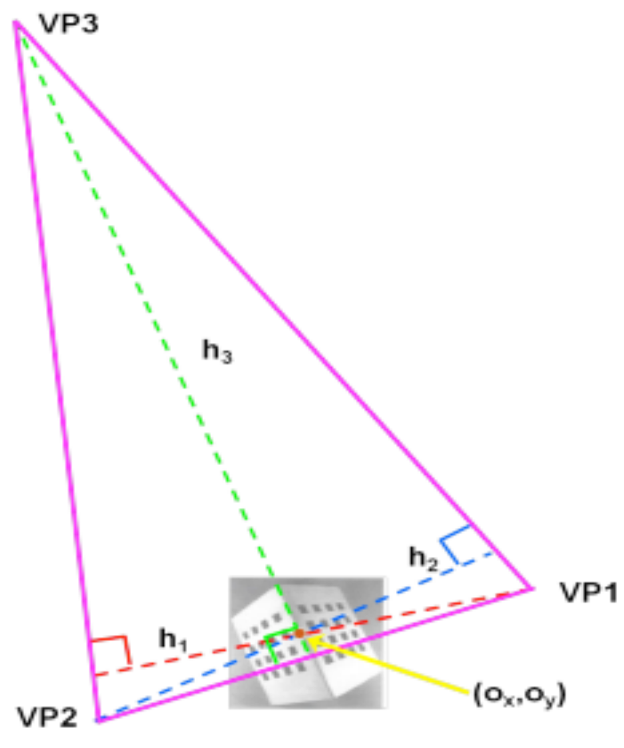
$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = M_{int} \, M_{ext} \begin{pmatrix} x_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix}$$

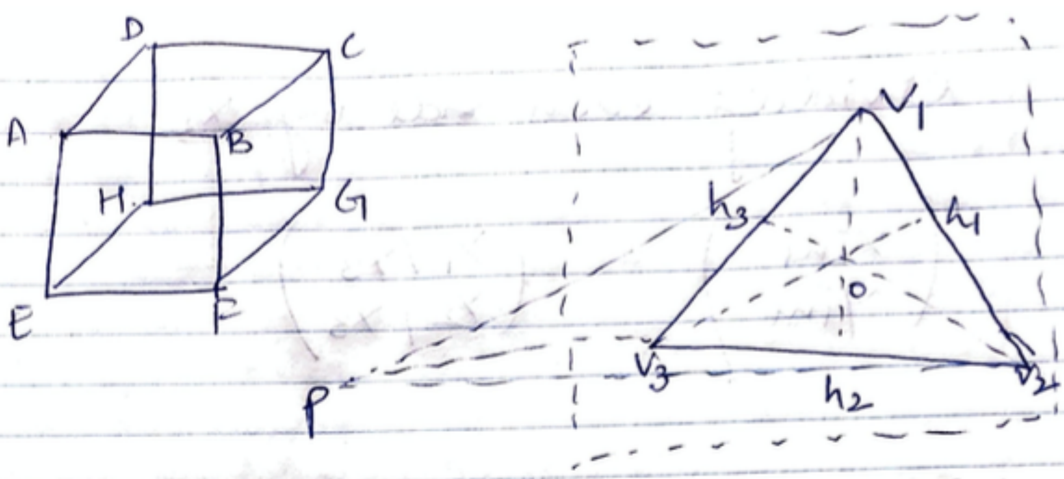
Dividing each side by X_3

$$\begin{pmatrix} X_{im} \\ Y_{im} \end{pmatrix} = \begin{pmatrix} X_1 / X_3 \\ X_2 / X_3 \end{pmatrix}$$

3. (Calibration- 20 points) Prove the Orthocenter Theorem by geometric arguments: Let T be the triangle on the image plane defined by the three vanishing points of three mutually orthogonal sets of parallel lines in space. Then the image center is the orthocenter of the triangle T (i.e., the common intersection of the three altitudes. Note that you are asked to prove the Orthocenter Theorem rather than that the orthocenter itself as the common interaction of the three altitudes, which you can use as a fact.
(1) Basic proof: use the result of Question 1, assuming the aspect ratio of the camera is 1. (10 points)

Ans. The center of projection of the camera in 3D space is O (O_x, O_y). Three mutually orthogonal sets of parallel lines: L1, L2, and L3. Assume there is a triangle formed by lines V1, V2, V3, and then they are also V1, V2, and V3 are the three vanishing points.





P is the center of projection

$AB \parallel CD \parallel EF$, $AD \parallel BC \parallel FG$, $AE \parallel BF \parallel CG$

$AD \perp AB \perp AE$, $BC \perp CD \perp CG$, $BE \perp EF \perp CG$

So V_1, is the vanishing pt of AE, BF, CG.

V_2 is the vanishing pt of AD, BC, FG

V_3 is the vanishing pt

So V_1, V_2, V_3 forms a triangle on T. and O is the centre on T. hence $OP \perp T$.

So $PV_1 \parallel AE \parallel BF \parallel CG$.
$PV_2 \parallel AD \parallel BC \parallel FG$.
$PV_3 \parallel AB \parallel EF \parallel DC$

hence $PV_1 \perp V_1V_2$, $PV_2 \perp V_1V_3$,

$PV_3 \perp V_1V_2$

V_1h_2, V_2h_3, V_3h_1 are altitudes of T.

$V_1h_2 \perp V_2V_3$, $V_2h_3 \perp V_1V_3$, $V_3h_1 \perp V_2V_1$

O is the projection of P in the image plane.

O is the intersection pt of V_1h_2, V_2h_3, V_3h_1.

(2) If you do not know the focal length of the camera, can you still find the image center using the Orthocenter Theorem? Can you further estimate the focal length? For both questions, please ***show why (and then how) or why not.*** *(5 points)*

Ans. Yes, if we do not know the focal length of the camera we can still find the image center length using orthocenter theorem, But finding focal length from the image is difficult, since vanishing point is intersection of the two lines in the image and not related to focal length.

(3) If you do not know the aspect ratio of the camera, can you still find the image center using the Orthocenter Theorem? ***Show why or why not.*** *(5 points)*

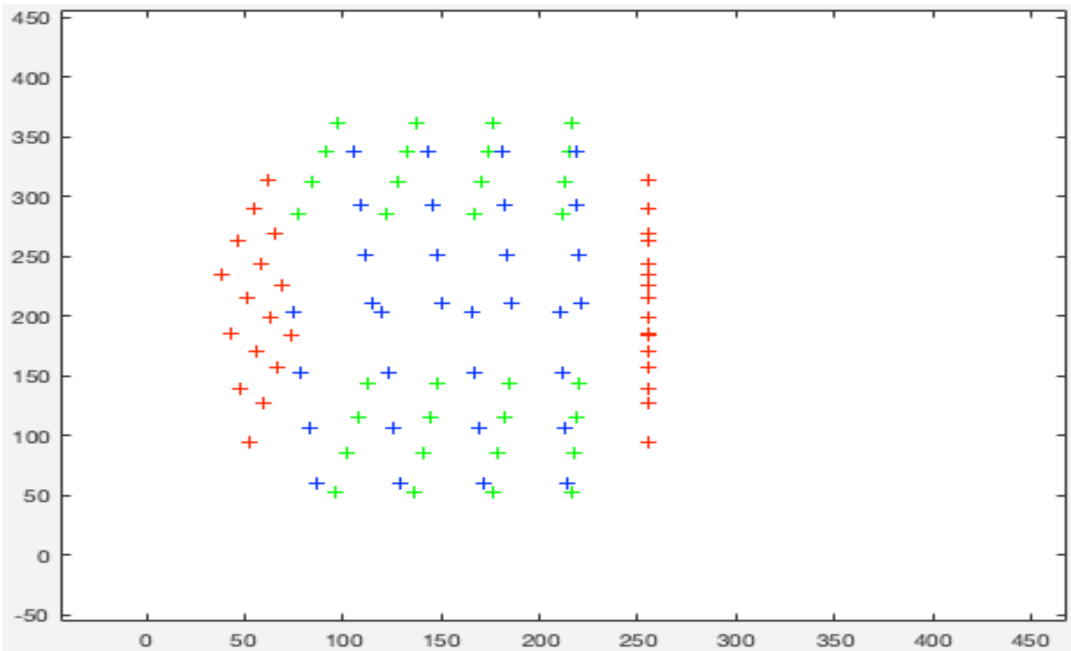
Ans. No if we do not know the aspect ratio and the focal length of the camera, we cannot find the image center using orthocenter theorem, since vanishing point is not unique and we can't get the unique orthocenter if we do not know the aspect ratio and focal length.

Q4. Calibration Programming Exercises (40 points): Implement the direct parameter calibration method in order to (1) learn how to use SVD to solve systems of linear equations; (2) understand the physical constraints of the camera parameters; and (3) understand important issues related to calibration, such as calibration pattern design, point localization accuracy and robustness of the algorithms. Since calibrating a real camera involves lots of work in calibration pattern design, image processing and error controls as well as solving the equations, we will mainly use simulated data to understand the algorithms. As a by-product we will also learn how to generate 2D images from 3D models using a "virtual" pinhole camera. The calibration procedure has the following three steps:

1. *Calibration pattern "design". Generate data of a "virtual" 3D cube similar to the one shown in* ***here*** *of the lecture notes in camera calibration. For example, you can hypothesize a 1x1x1 m^3 cube and pick up the coordinates of 3-D points on corners of each black square in your world coordinate system. Make sure that your data is sufficient for the following calibration procedures. In order to check*

the correctness of your data, draw your cube (with the control points marked) using Matlab (or whatever tools you are selecting). I have provided a piece of **starting code** in Matlab for you to use.

Ans. A virtual 3D cube is generated and as a result, a graph is generated. On the graph, at each surface there are 4x4 points. Those points are used as variables for equation. Xw=0, Yw=0, Xw=1, Yw=1, Zw=1, Zw=0. So Pw is a 1x1x1 m3 cube, 96 points are generated and each surface have 16 points. Total 96 points for calibration



2. *A "virtual" camera and its images. Design a "virtual" camera with known intrinsic parameters including focal length f, image center (o_x, o_y) and pixel size (s_x, s_y). As an example, you can assume that the focal length is f = 16 mm, the image frame size is 512*512 (pixels) with (o_x,o_y) = (256, 256), and the size of the image sensor inside your camera is 8.8 mm *6.6 mm (so the pixel size is (s_x,s_y) = (8.8/512, 6.6/512)). Capture an image of your "virtual" calibration cube with your virtual camera in a given pose (i.e., R and T). For example, you can take the picture of the cube 4 meters away and with a tilt angle of 30 degree. Use three rotation angles alpha, beta, gamma to generate the rotation matrix R (refer to the lecture notes in camera model – please check the correctness of the R equation especially for signs). You may need to try different pose in order to have a suitable image of your calibration target. In your report, please clearly list the parameters you used for generating your calibration image. They are called the "ground truth" of the parameters.*

Ans. Intrinsic parameters given focal length f=16mm,

image center (ox, oy) = (256, 256), and pixel size (sx, sy). Using three rotation angles alpha, beta, gamma to generate the rotation matrix R, I will tilt it to 30 degree to have better projection.
Sx = 0.0088/512
Sy = 0.0066/512
Fx = f/s_x
= 930.90
Similarly, F_y = 1241.2

T = $\begin{bmatrix} 0 & 0 & 4 \\ Tx & Ty & Tz \end{bmatrix}$

$$\alpha = 30°$$
$$\beta = 0\ °$$
$$\gamma = 0°$$

R = Ra*Rb*Rr

Matrix = $\begin{bmatrix} 1.0 & 0 & 0 \\ 0 & 0.86 & -0.5 \\ 0 & 0.5 & 0.8 \end{bmatrix}$

3. *Direction calibration method: Estimate the intrinsic (f_x, f_y, aspect ratio a, image center (o_x,o_y)) and extrinsic (R, T and further alpha, beta, gamma) parameters. Use SVD to solve the homogeneous linear system and the least square problem, and to enforce the orthogonality constraint on the estimate of R. You are asked to do the following:*

i. Use the accurately simulated data (both 3D world coordinates and 2D image coordinates) to the algorithms, and compare the results with the "ground truth" data (which are given in step (a) and step (b)). Remember you are practicing a camera calibration, so you should pretend you know nothing about the camera parameters (i.e. you cannot use the ground truth data in your calibration process). However, in the direct calibration method, you could use the knowledge of the image center (in the homogeneous system to find extrinsic parameters) and the aspect ratio (in the Orthocenter theorem method to find image center).

Ans. The image center Ox = 256; Oy = 256, every point in 3D have corresponding point image, first we find the matrix A where A=UDVT If the matrix A is a real matrix, then U and V are also real, Calculate Tz, Fx, Fy, forming 8 cols and 96 rows, which 8 unknows v=(v1…..v8)T 7 of them are independent parameters and 8th row is zero.

$$x_i X_i v_1 + x_i Y_i v_2 + x_i Z_i v_3 + x_i v_4 - y_i X_i v_5 - y_i Y_i v_6 - y_i Z_i v_7 - y_i v_8 = 0$$
$$(v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8) = (r_{21}, r_{22}, r_{23}, T_y, \alpha r_{11}, \alpha r_{12}, \alpha r_{13}, \alpha T_x)$$

A matrix [96x8]

V= [1 x8]
α = sqrt ($v_5^{-2} + v_6^{-2} + v_7^{-2}$)
α = 1.6

A matrix =

-217.8480	-43.5696	-43.5696	-217.8480	21.2634	4.2527	4.2527	21.2634
-209.3619	-41.8724	-83.7448	-209.3619	-7.4798	-1.4960	-2.9919	-7.4798
-201.5123	-40.3025	-120.9074	-201.5123	-34.0676	-6.8135	-20.4406	-34.0676
-194.2299	-38.8460	-155.3839	-194.2299	-58.7338	-11.7468	-46.9870	-58.7338
-212.8666	-85.1466	-42.5733	-212.8666	69.9366	27.9747	13.9873	69.9366
-204.7570	-81.9028	-81.9028	-204.7570	39.9713	15.9885	15.9885	39.9713
-197.2426	-78.8970	-118.3455	-197.2426	12.2054	4.8822	7.3232	12.2054
-190.2602	-76.1041	-152.2082	-190.2602	-13.5947	-5.4379	-10.8757	-13.5947
-208.1078	-124.8647	-41.6216	-208.1078	116.4336	69.8602	23.2867	116.4336

SVD: [U,V,S]=svd(A)
V:

0.4618	-0.3469	0.3377	-0.0749	0.3104	-0.0129	0.6710	-0.0000
0.3374	0.0467	-0.0292	-0.4083	-0.4777	0.0892	-0.0163	0.6928
0.2443	-0.3095	-0.6967	-0.2771	-0.0977	0.3332	0.0432	-0.4000
0.5662	-0.3648	0.1026	0.3206	0.0738	-0.1695	-0.6316	0.0000
-0.2268	-0.2603	-0.4307	0.2868	0.4865	0.1192	0.0476	0.6000
-0.3162	-0.3905	0.4293	-0.4108	0.1704	0.5159	-0.3152	0.0000
-0.0990	-0.2486	0.1312	0.6121	-0.5509	0.4369	0.2052	0.0000
-0.3626	-0.6055	-0.0470	-0.1497	-0.3021	-0.6172	0.0714	-0.0000

v= [-0.0000 0.6928 -0.4000 0.0000 0.6000 0.0000 0.0000 -0.0000]

Calculating Aspect ratio

Aspect Ratio	Ground Truth	Estimated	Difference
Ar	0.75	0.7544	0.00

Calculate Fx, Fy

	Ground Truth	Estimated	Difference
Fx	930.9	930.9	0
Fy	1241.2	1241.2	0

Calculate R & T
$R_3^T = R_1^T + R_2^T$
Comparing Ground Truth and Estimated R is identical to the virtual Camera

Calculate T = [0,0,4]

	Ground Truth	Estimated	Difference
Tx	0	6.7009	6.7009
Ty	0	1.8468	1.8468
Tz	4	4	0

ii. *Show experimentally whether the unknown aspect ratio matters in estimating the image center, and how the initial estimation of image center affects the estimating of the remaining parameters. Give a solution to solve the problems if any and implement it.*

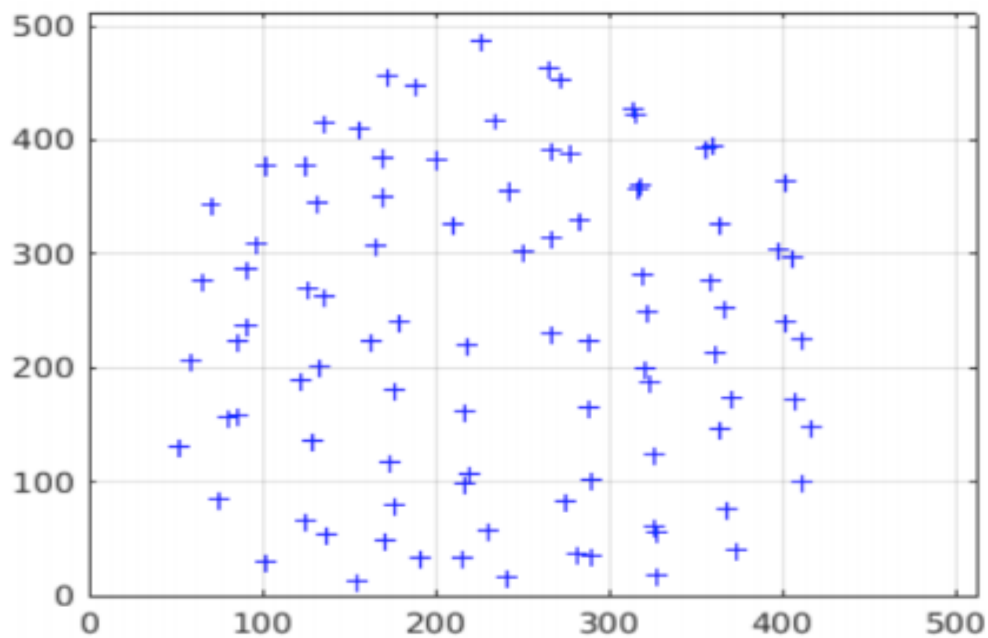
Ans. If the ratio of an image is different from the ratio of screen, we may not see the whole image. Images won't fit properly if the screen is narrower than the image. We know the first 2 rows of R and first two components from the T. By this 3rd row will calculate from the R1 and R2. Next is to Determine the sign s positive or negative, assuming s as positive, test sign of s use the image point. We got X = -1 and x = 1 so s is negative in this case. if we do not know the aspect ratio of the camera, we can't find the image center using Orthocenter Theorem. Because the vanishing point is not unique, and we can't get the unique orthocenter if we do not know the aspect ratio.

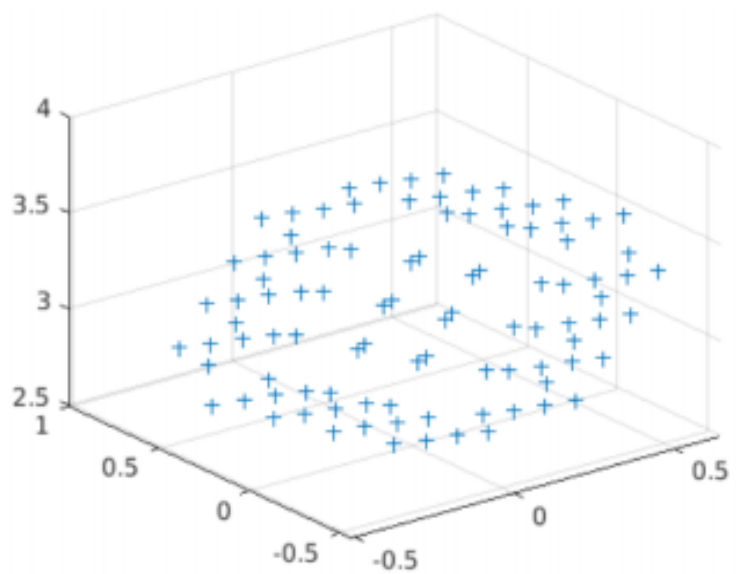
iii. Accuracy Issues. Add in some random noises to the simulated data and run the calibration algorithms again. See how the "design tolerance" of the calibration target and the localization errors of 2D image points affect the calibration accuracy. For example, you can add 0.5 mm random error to 3D points and 1.0 pixel random error to 2D points. Also analyze how sensitive of the Orthocenter method is to the extrinsic parameters in imaging the three sets of the orthogonal parallel lines. **(* extra points:10**)

Ans. Adding some noise, 0.1 mm random error to 3D points and 0.5pixel random error to 2D points. Once noise is added we can Then we can constraints by using SVD.

	Ground Truth	Noise	Difference
AR	0.75	0.7491	0.0009
Fx	930.9	926.9095	3.9905
Fy	1241.2	1242.7	1.5
Tx	6.7009	-2.6265e-0.4	0.00093
Ty	1.8468	-6.6564e0.4	0.00055
Tz	4	3.9453	0.05

As per the result – Aspect ratio, Fx, Fy, Tx, Ty, Tz are all changed by adding the noise. We can conclude that orthocenter method is very sensitive to the extrinsic perimeters.


0
100
200
300
400
500
0
100
200
300
400
500


4
3.5
3
2.5
1
0.5
0
-0.5
-0.5
0
0.5