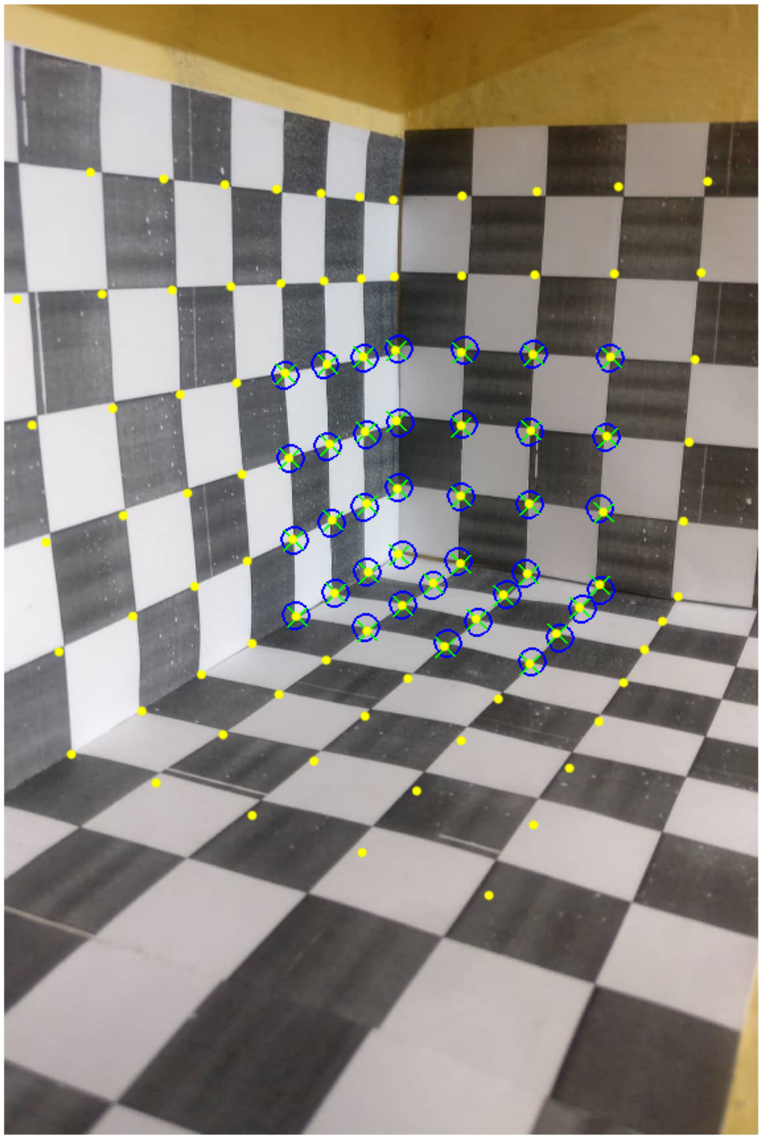
Question 1 : Here is the visualization of calculated points on the image:



Meaning of points
Measured 2D points - Blue
Calculated 2D points (2D and 3D measurements for estimating P) - Green
Calculated 2D points (2D and 3D measurements NOT used for estimating P) - Yellow

Reason for Normalizing Data

The normalization is basically a preconditioning to decrease CONDITION NUMBER of the matrix P (the larger the condition number, the nearer the matrix is to the singular matrix).

Putting it simply, the matrix P (un-normalized) consists of products of image coordinates which can have different scale.

The source and target coordinate data are usually noisy. Without normalization, the data from source can have multiple orders of magnitude larger variance than from target (or vice versa).

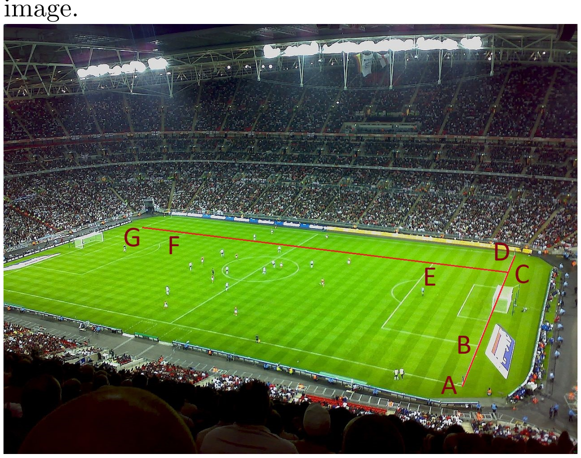
The homography estimation usually finds parameters in a least-squares sense - hence the best statistical estimate is found only if variances of the parameters are the same (or known beforehand, but it is more practical just to normalize the input). So normalization is essential not only for numerical stability, but also for more accurate estimation in presence of noise and faster solution (in case of iterative solver).

Question 2 : The four corners of box are used for calculating homography. These are:
Pixel coordinate - Real world coordinate
(1060,719) - (18,44)
(845, 677) - (0,44)
(959,534) - (0,0)
(1126,555) - (18,0)

Three corners of the field are used to calculate the length and breadth. Their pixel coordinates are (375,433), (1140,517) and (1024,808).

Length = 118.9513yd and Breadth = 74.4298yd.

Question 3 : We know cross-ratio remains invariant over projective transformations
Thus we will use cross-ratio to find dimensions of the football field in Wembley.
Firstly, for the breadth of the field, consider the points A, B, C and D in the image.



Pixel values are:
A - (1025, 808)
B - (1060, 720)
C - (1126, 555)
D - (1141, 518)

Distances in image domain are:
AC = 272.4 , AD = 312.3 , BC = 177.7 , BD = 217.6

Distances in real world are (assuming AB = CD = x):
AC = 44 + x , AD = 44 + 2x , BC = 44 , BD = 44 + x

Equating cross-ratio, we get:

$$\frac{\frac{272.4}{312.3}}{\frac{177.7}{217.5}} = \frac{\frac{44+x}{44+2x}}{\frac{44}{44+x}}$$

$$\implies x^2 - 5.9928x - 131.8416 = 0$$

$$\implies x = 14.86$$

$$\implies \text{Breadth} = \text{AD} = 44 + 2x = 73.72$$

Now, for the length of the field, consider the points C, E, F and G in the image.
Pixel values are:
C - (1126, 555)
E - (959, 534)

F - (407, 464)
G - (311, 455)

Distances in image domain are:
CF = 724.7 , CG = 821.1 , EF = 556.4 , EG = 652.8

Distances in real world are (assuming EF = y):
CF = 18 + y , CG = 36 + y , EF = y , EG = 18 + y

Equating cross-ratio, we get:

$$\frac{\frac{724.7}{821.1}}{\frac{556.4}{652.8}} = \frac{\frac{18+y}{36+y}}{\frac{y}{18+y}}$$

$$\implies 0.0355x^2 + 1.278x - 324 = 0$$

$$\implies x = 79.22$$

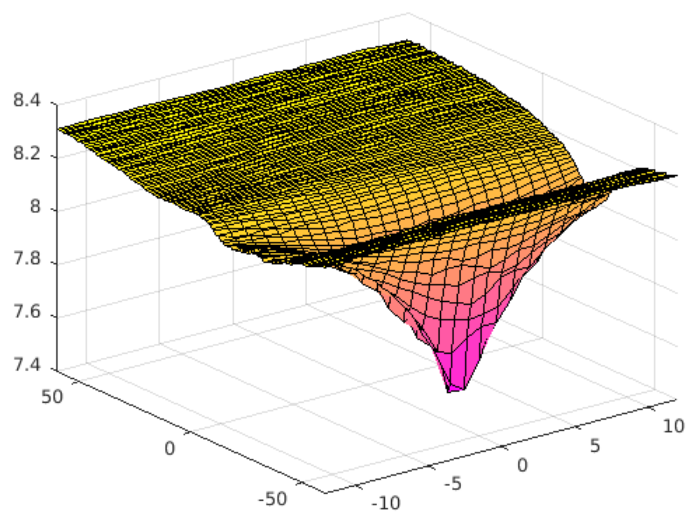
$$\implies \text{Length} = \text{CG} = 36 + y = 115.22$$

Therefore, length = 115.22yd and breadth = 73.72yd.

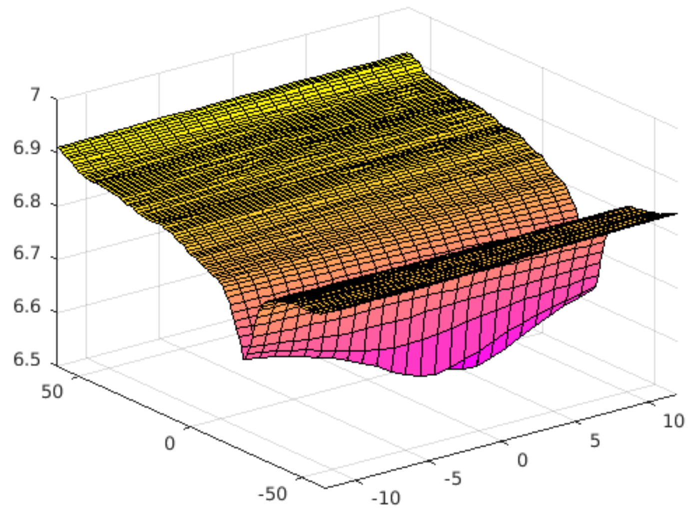
Question 5 : Quality of alignment is same for both pair of images. Output is exactly correct ie. t_x is predicted as 3 and θ is predicted as -23 or -24.

When the translation is very large, we do not get minimal joint entropy at correct value, instead it is obtained at boundary values of translation as there is minimum overlap between the two images. For exmaple, when original t_x is -100 and θ is 23.5, we get minimum joint entropy for t_x = -12 and $theta$ = -22.

Joint Entropy Plot for barbara.png



Joint Entropy Plot for flash.png



Joint Entropy Plot for undesirable barbara.png