Assignment on MVG

Assignment on MVG

- If you have duplicated this, please add your name here during submission.

> MY NAME & DESIGNATION: Vikrant Dewangan, CSD, UG2k18

Transformations & Single View Geometry

Contact Rahul.

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/61b7ea04-3d14-43b5-8260-52829eccce9a/MVG_Assignment_1.pdf

`Homework: TODO0: 40 Marks`

Toggle to see the questions.

- ▼ `Homework: TODO1 - 20 marks` 👅 - ***Single View Geometry***
  - ▼ Does the scale of the world points play any role in camera calibration (DLT) (as in measuring the points in metres vs cms vs kms)? If so, why or why not?

    Answer:No, the scale of the world points do not play any role. When converting from one form of units to another form of units, the scale would appear in both numerator and denominator of x and y.

$$x_i = \frac{u_i}{w_i} = \frac{A^T X_i}{C^T X_i}$$
$$y_i = \frac{v_i}{w_i} = \frac{B^T X_i}{C^T X_i}$$

▼ Why does DLT fail when all the points are on a single plane? Mathematical show this.

Answer DLT fails for coplanar points as in case of coplanar points, we don't have the right number of equations to solve and get the P matrix. The rank of M reduces.

$$M = \begin{bmatrix} \dots \\ a_{x_i}^T \\ a_{y_i}^T \\ \dots \end{bmatrix}$$
$$= \begin{bmatrix} -X_i & -Y_i & -Z_i & -1 & 0 & 0 & 0 & 0 & x_iX_i & x_iY_i & x_iZ_i & x_i \\ 0 & 0 & 0 & 0 & -X_i & -Y_i & -Z_i & -1 & y_iX_i & y_iY_i & y_iZ_i & y_i \end{bmatrix}$$

Now imagine, that Z_i is zero and we are lying instead on the x-y plane. We would get -

$$M = \begin{bmatrix} \dots \\ a_{x_i}^T \\ a_{y_i}^T \\ \dots \end{bmatrix}$$
$$= \begin{bmatrix} -X_i & -Y_i & 0 & -1 & 0 & 0 & 0 & 0 & x_iX_i & x_iY_i & 0 & x_i \\ 0 & 0 & 0 & 0 & -X_i & -Y_i & 0 & -1 & y_iX_i & y_iY_i & 0 & y_i \end{bmatrix}$$

Thus, M has no longer the rank of 11.

▼ Toggle to see:

While solving for the optimization problem for DLT we obtain that P vector should be an eigen vector of $M^T M$. Show that the last eigen vector gives the lowest value for the square error. (Hint: substitute $M^T M$ with UDV^T)

Answer:We have

$$\Omega = v_i^T(s_i^2 v_i v_i^T)v_i = s_i^2 v_i^T v_i v_i^T v_i = s_i^2$$

Thus, we have the squared error depending upon the singular value si. Choosing the lowest eigen value will give the lowest singular value and thus, the squared error will be the least.

Epipolar Geometry

Contact Shubodh.

▼ `Homework: TODO2 - 10 Marks` ***- Fundamental Matrix***

▼ What are the no of degrees of freedom of F? Elaborate.

Answer:We have, F being the matrix that relates one image to another in terms of **pixel coordinates**.

$$x^T F x = 0$$

where

$$\mathbf{x} = (x, y, 1)^{\mathrm{T}} \text{ and } \mathbf{x}' = (x', y', 1)^{\mathbf{T}}$$

Thus, F is a 3 x 3 matrix, then it can be written as -

$$x'xf_{11} + x'yf_{12} + x'f_{13} + y'xf_{21} + y'yf_{22} + y'f_{23} + xf_{31} + yf_{32} + f_{33} = 0$$

Say, currently having 9 degrees of freedom

$$f = (f_{11}, f_{12}, f_{13}, f_{21}, f_{22}, f_{23}, f_{31}, f_{32}, f_{33})$$

Apart from this, we have that since we are in homologous coordinates, we have effectively only 8 degrees left (for example, we remove last component to be multiplied)

$$f = (f'_{11}, f'_{12}, f'_{13}, f'_{21}, f'_{22}, f'_{23}, f'_{31}, f'_{32}, 1)$$

Where

$$f'_{ij} = \frac{f_{ij}}{f_{33}}$$

And now since the rank of F is 2, we have that it's determinant is 0. And thus, working out the equation, we have -

$$\begin{bmatrix} f'_{11} & f'_{12} & f'_{13} \\ f'_{21} & f'_{22} & f'_{23} \\ f'_{31} & f'_{32} & 1 \end{bmatrix}$$

$$f'_{11} \cdot f'_{22} + f'_{12} \cdot f'_{23} \cdot f'_{31} + f'_{13} \cdot f'_{21} \cdot f'_{32} - f'_{11} \cdot f'_{23} \cdot f'_{32} - f'_{12} \cdot f'_{21} - f'_{13} \cdot f'_{22} \cdot f'_{31} = 0$$

Now here, we have lost a degree of freedom, since their summation is zero and any one can be written as linear combination of another. Thus, we have only **7** degrees of freedom left.

▼ `Homework: TODO3 - 5 Marks` - ***Singularity Cases in 8 point algorithm***

One singularity case when the above 8 point algorithm fails is under pure rotation (another case is if all the corresponding points are points on a plane (in the world)).

- 2a: Answer - Elaborate for pure rotation case.

  For pure rotation, we have that camera is rotated about an axis.

- `BONUS 5 Marks` 2b: Answer - Elaborate for world points planar case.

  The world planar case occurs when all the given n points are in a plane altogether.

- `Homework: TODO4 - 5 Marks` - ***RANSAC***

  - Explain why RANSAC is necessary for accurate F estimation.

    Answer:RANSAC is helpful in general, for removing outliers which dont have any influence on the inliers. RANSAC chooses randomly sampled points, with many iterations taking place.RANSAC provides robust estimation of camera matrices better than using just 8 point algorithm which is highly unstable.

  - Explain briefly how you'd apply RANSAC algorithm to computation of F (using normalized 8 point algorithm)?

    Answer:

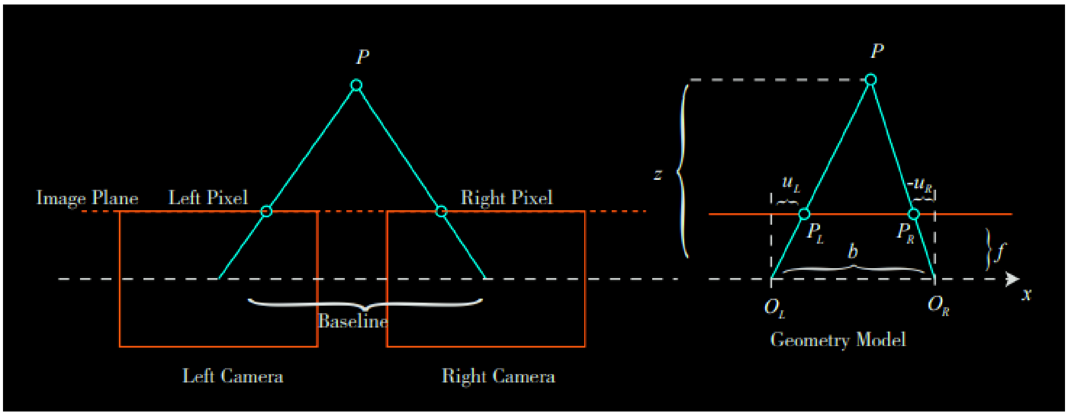
```
for N iterations:
  Select a min subset k from data-set s randomly
  Compute F using the normalized 8 point algorithm from the subset k
  Classifiy (xi,xi') as outlier/inlier based on scoring function and a computation method
Select the best solution having the highest score
```

- `Homework: TODO5 - 15 Marks` - ***Stereo***

  1. Deriving depth equation

     - a. Derive the depth equation step by step with proper explanation.

       Answer:



       We have, from similarity of triangles.

$$(\triangle PP_LP_R \sim \triangle PO_LO_R)$$

We drop an altitude PH, PI from P to the image plane and camera planes

$$(\triangle PP_LH \sim \triangle PO_LI)$$

$$\frac{PH}{PI} = \frac{PP_L}{PO_L} = \frac{P_LP_R}{O_LO_R}$$

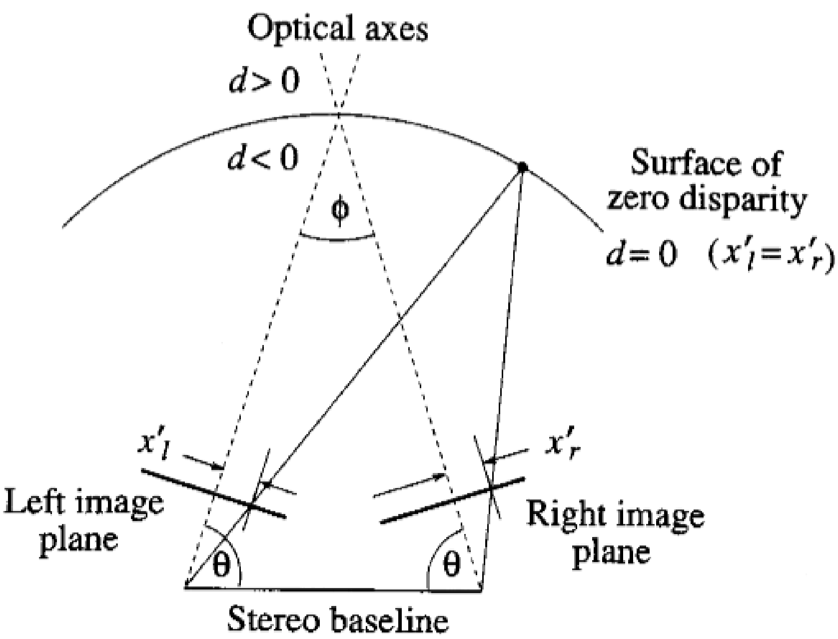
$$\frac{z-f}{z} = \frac{b-u_L+u_R}{b}$$

$$1-\frac{f}{z} = 1-\frac{d}{b}$$

$$\therefore z = \frac{fb}{d}$$

▼ b. Can you give a real world example where "d" is zero?

Answer: "d" refers to the disparity amongst the left and the right perspectives. It is 0 along a surface -



Thus, in real world, we can say that if object is spherical surface, it gives 0 disparity.

2. There are many ways we humans reason about depth. One most common cue seems to be the same principle used in stereo.

▼ a. Elaborate on this statement.

Answer: Just like stereo camera, humans are gifted with two sets of eyes which form images in their retinal cavity. The brain then combines them together using the same principle in stereo to give us an idea of the depth.

▼ b. Also, if that is really the case, why do you think we do pretty good in sensing depth with 1 eye closed (or people born with 1 eye)?

Answer:That is because, there are many monocular cues which help us in sensing the depth. The following are some examples -

- **Texture Gradient** : This is important as the texture dilutes as we look into the distance. For example the trees on far side of mountain may not be visible compared to the near side.
- **Relative Size** : Human brain is taught the sizes of things, like the height of a building, the length of the road by signs etc. Thus we don't exactly compute the depth from absolute level.
- **Parallax** : Objects that are closer seem to zoom by faster than do objects in the distance. When you're riding in a car, for example, the nearby telephone poles rush by much faster than the trees in the distance. This visual clue allows you to perceive the fast moving objects in the foreground as closer than the slower moving objects off in the distance.

3. What happens if the left and right camera in the above image aren't facing in the same direction (but with some common overlap region of the world)? In other words, they have a rotational transformation in addition to existing transnational transformation. Put your answer under "Answer: 3a" section. After writing down your answer, google "stereo rectification" and explain what you understand under "Answer: 3b" section.

▼ Answer: 3a

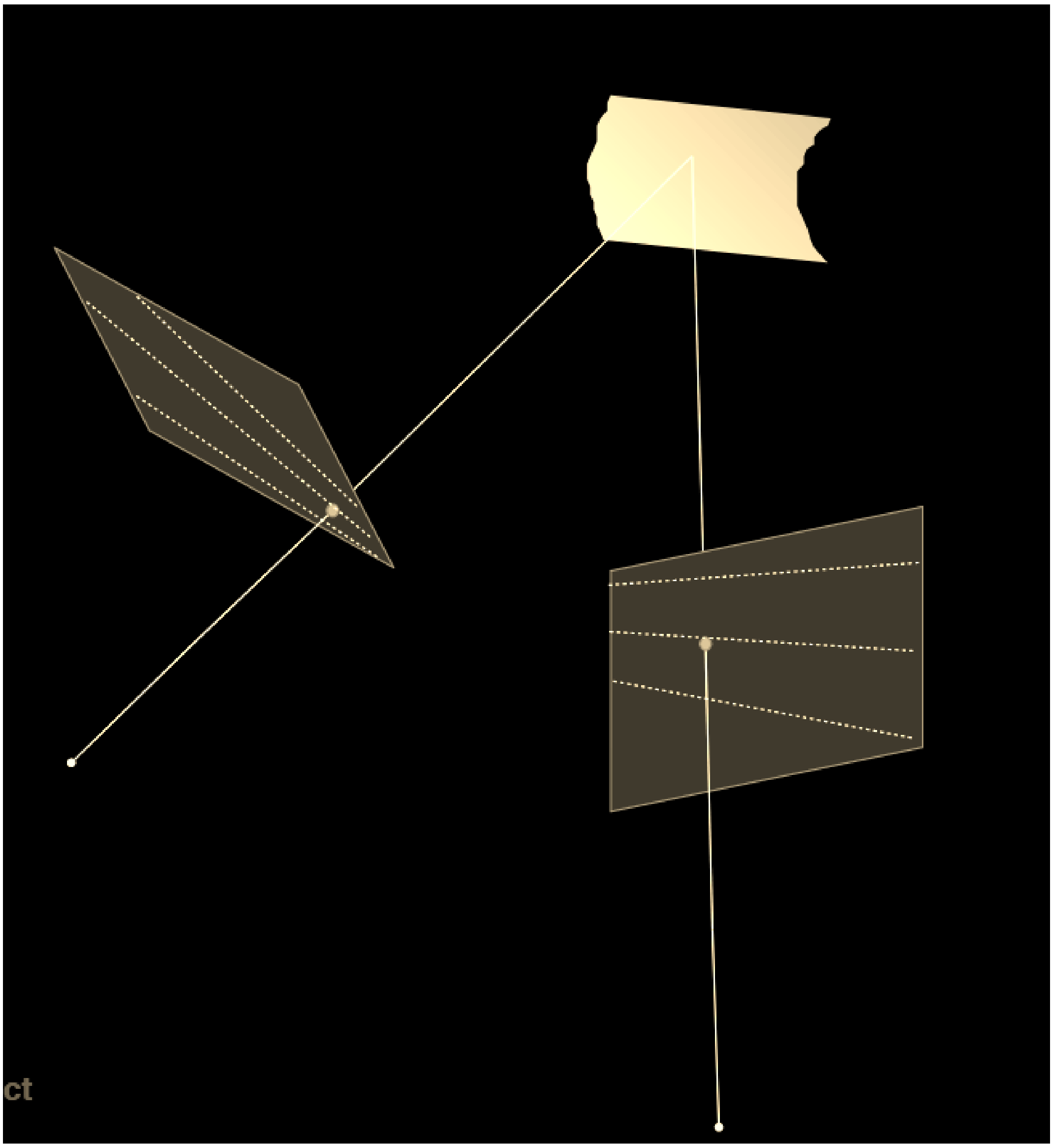
In the case, they have a rotation and translation transformation, then the camera frame of one would be needed to be converted to another by multiplying it with the rotation matrix.

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

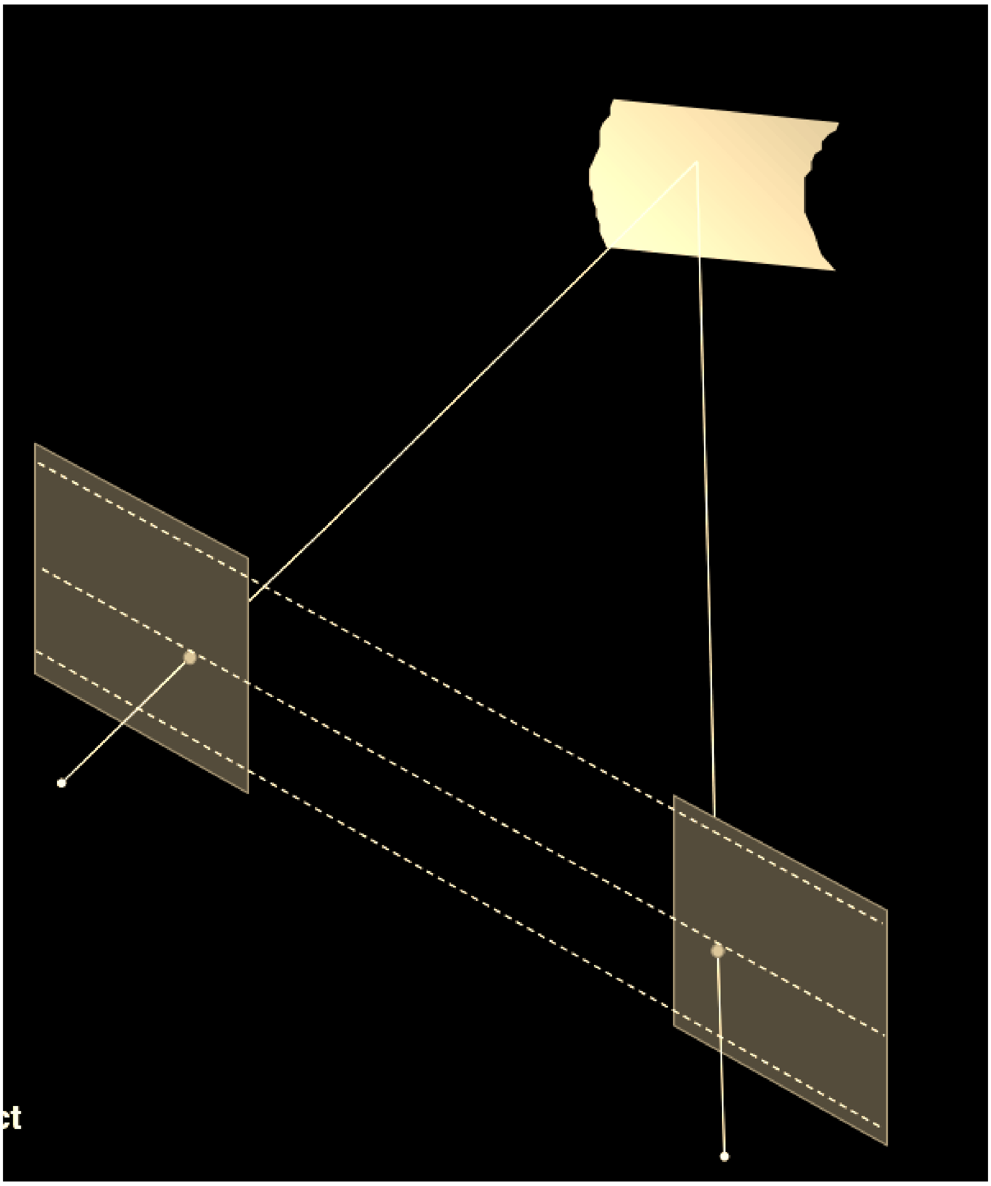
▼ Answer: 3b

Stereo rectification (or Image rectification) is the process of making the epipolar lines horizontal. It is done by reprojecting image plane onto a common plane parallel to the line between camera centers.

From this



To this

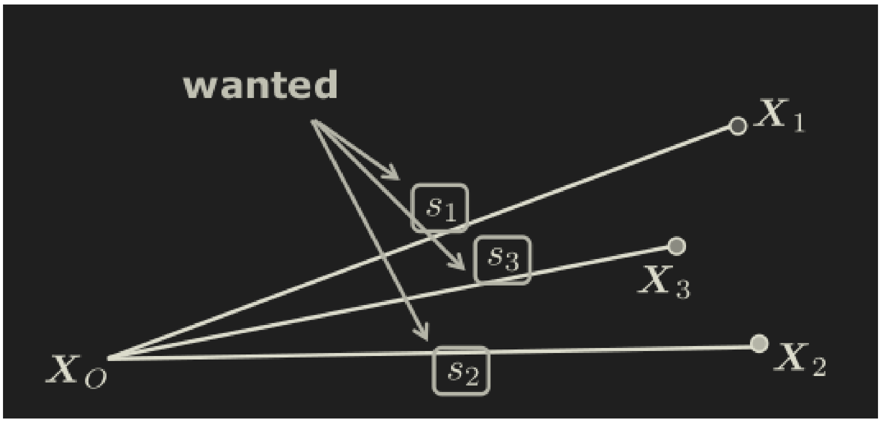


▼ `Homework: TODO6 - 5 Marks` - ***PnP***

Explain the working principle of the P3P algorithm?

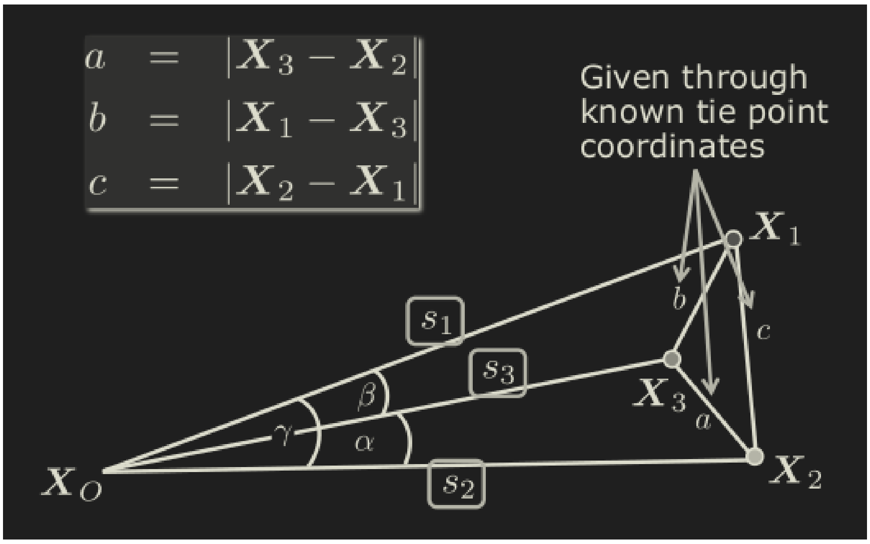
The PnP (Points and Perspectives) problem that given some N points and their coordinates in the 3D world as well as their 2D projections, we have to estimate the calibrated camera pose (6D of freedom - 3 Translational + 3 Rotation).

Say, we have XO and need to find the length of projection rays s1, s2 and s3 -



We would have, say if

$$
\begin{aligned}
a &= |X_3 - X_2| \\
b &= |X_1 - X_3| \\
c &= |X_2 - X_1|
\end{aligned}
$$



From the cosine rule, we have -

$$
\begin{aligned}
a^2 &= s_2^2 + s_3^2 - 2s_2s_3\cos\alpha \\
b^2 &= s_1^2 + s_3^2 - 2s_1s_3\cos\beta \\
c^2 &= s_1^2 + s_2^2 - 2s_1s_2\cos\gamma
\end{aligned}
$$

$$\text{We have:} \quad a^2 = s_2^2 + s_3^2 - 2s_2s_3\cos\alpha$$

$$\text{Define:} \quad u = \tfrac{s_2}{s_1} \quad v = \tfrac{s_3}{s_1}$$

$$
\begin{aligned}
s_1^2 &= \frac{a^2}{u^2 + v^2 - 2uv\cos\alpha} \\
&= \frac{b^2}{1 + v^2 - 2v\cos\beta} \\
&= \frac{c^2}{1 + u^2 - 2u\cos\gamma}
\end{aligned}
$$

We substitute u in the equation of a 4 degree polynomial -

$$A_4v^4 + A_3v^3 + A_2v^2 + A_1v + A_0 = 0$$

$$
\begin{aligned}
A_4 = &\left(\tfrac{a^2-c^2}{b^2} - 1\right)^2 - \tfrac{4c^2}{b^2}\cos^2\alpha \\
A_3 = &4\left[\frac{a^2-c^2}{b^2}\left(1 - \frac{a^2-c^2}{b^2}\right)\cos\beta\right. \\
&\left.- \left(1 - \frac{a^2+c^2}{b^2}\right)\cos\alpha\cos\gamma + 2\frac{c^2}{b^2}\cos^2\alpha\cos\beta\right]
\end{aligned}
$$

$$
\begin{aligned}
A_1 = &4\left[-\left(\frac{a^2-c^2}{b^2}\right)\left(1 + \frac{a^2-c^2}{b^2}\right)\cos\beta\right. \\
&+ \frac{2a^2}{b^2}\cos^2\gamma\cos\beta \\
&\left.- \left(1 - \left(\frac{a^2+c^2}{b^2}\right)\right)\cos\alpha\cos\gamma\right]
\end{aligned}
$$

$$A_0 = \left(1 + \frac{a^2-c^2}{b^2}\right)^2 - \frac{4a^2}{b^2}\cos^2\gamma$$

Instructions

- The assignment is for 100 marks. Coding question is named as TODO0 (Coding assignment - PDF embed at the top of this page) and the theory questions are named from TODO1 - TODO6. So 7 questions in total.
- Deadline is 14/06.

How to submit the assignment?

Create a Notion account and duplicate this page to your workspace. You will have to share the link of this duplicated page with us to see your answers. **Enable comment access so that we can share feedback. (Share → Show link options → Allow comments)**

▼ For theory questions, you can do one of the following things:

▼ Write the whole individual answer on paper and take a photo of it and insert it under each respective "Answer:" field.

You can either

1. use Notion Android app to insert images one by one or
2. take pictures of all the answers at once, send all the images to your laptop and then drag and drop a particular image under a respective answer.

▼ (Preferable for those with bad handwriting) Type out the answer and use Notion LaTeX blocks. Note that Notion only supports display math blocks.

1. As for inline math, you will have to do jugaad (such as render it somewhere and then take a screenshot and paste it).
2. If you want to draw something, you might have to take the 1st approach or use some software like TikZ and paste screenshots.

▼ For coding questions

If your preferred language is python then code in jupyter notebook and display all your outputs in the notebook. Jupyter notebook and Matlab files should be mailed to Rahul Sajnani.

Email: rahul.sajnani@research.iiit.ac.in