Preliminaries

- **Perspective projection**: $P = (X, Y, Z)^T$ projects to $p = (x, y, z)^T = \frac{f}{Z}P$

- **(Scaled) Orthographic projection**: $P = (X, Y, Z)^T$ projects to $(x, y)^T = s(X, Y)^T$, $s > 0$ is some scalar factor (usually set to $s = \frac{1}{Z_0}$).

- **Rigid transformation**: $Q = RP + t$, $R$ is $3 \times 3$ satisfying $RR^T = I$, $t = (t_x, t_y, t_z)^T \in \mathbb{R}^3$.

- **Essential matrix**: Here is an algebraic way to derive the Essential matrix that relates two images. $p$ and $q$ are two corresponding image points (perspective projection of $P$ and $Q$, which represent the same scene points in two coordinate frames associated with the camera in each image).

  $Q = RP + t$

  implies

  $t \times Q = t \times (RP + t) = (t \times RP) + (t \times t) = t \times RP$

  Take scalar product of $Q$ with both sides:

  $Q^T(t \times Q) = Q^T(t \times RP)$

  But $Q$ is orthogonal to $t \times Q$, so

  $Q^T(t \times RP) = 0$

  or

  $Q^T[t]_\times RP = 0$

  Define $E = [t]_\times R$ we obtain

  $Q^T EP = 0$

  Finally, $q \propto Q$ and $p \propto P$, thus

  $q^T Ep = 0$

  (Verify this by plugging $p = \alpha P$ and $q = \beta(RP + t)$.)
Questions

Question 1:
A calibration pattern is a set of 3D points at known positions. Such a pattern is used to calibrate a stereo rig (determine the transformation between the two cameras).

Suppose we take two images of a calibration pattern with the camera at unknown positions and orientations (so a point $P$ from the pattern is expressed as $R_1P + t_1$ in the coordinate system of image 1; and $P ightarrow R_2P + t_2$ in image 2). To calibrate the cameras in this case we need a more general expression for the essential matrix.

Express the essential matrix between the two images in terms of $R_1$, $t_1$, $R_2$ and $t_2$.

Question 2:
Let $P_i = (X_i, Y_i, Z_i)^T$, $i = 1, \ldots, N$ be a set of 3D points in space with image projections $p_i = (x_i, y_i)^T = \frac{1}{Z_i}P_i$. Assuming a rigid camera motion $P'_i = RP_i + t$ where $R$ is the rotational component and $t$ is the translational component of camera motion, with projections $p'_i = \frac{1}{Z'_i}P'_i$.

Show that in the case of pure rotation ($t = 0$), it is not possible to recover the structure of the scene (the depth $Z_i$) given any number of matching pairs $p_i, p'_i$.

Show the existence of a mapping from $p_i$ to $p'_i$ that does not include 3D structure (i.e., depth $Z_i$). Explain why the existence of such mapping proves the claim.

Question 3:
A camera is imaging an object at two time instances. Point correspondences are given across the two images. Can depth be recovered if between the two images:
1. There is only camera rotation?
2. There is only an object rotation (about its center of mass)?

In case where depth cannot be recovered show explicitly that it can be eliminated from the equations relating the two images.

Question 4:
1. Show that the right epipole (the epipole in the first image) is given by $v = \alpha R^T t$ for some $\alpha \neq 0$. Hint: To show this show that $Ev = 0$.

Explain why this implies that $v$ is the epipole (the intersection of all epipolar lines).

2. Derive an expression for the left epipole (in other words, find $u$ such that $u^T E = 0$). Hint: invert the rigid transformation that relates $P$ and $Q$ and use the expression derived for the right epipole.

Question 5:
Let $H$ be a non trivial homography ($H \neq \alpha I$). A point is called fixed if it does not change location following the application of $H$. A line is called fixed if each point on the line is mapped by $H$ to some point along the same line. A line is called pointwise fixed if each point along the line is fixed.

1. Derive an expression for a fixed point; what should such a point satisfy? How many isolated points can be fixed under $H$?
2. In the case of maximal number of isolated fixed points, identify the fixed lines.
3. How many lines can be pointwise fixed?
Question 6:

Two images of the same scene are related by general rotation and translation. Corresponding points $p_i$ and $p'_i$ in the two images satisfy the fundamental relation $p'_i^T F p_i = 0$. Suppose that all the points $P_i$ lie on some plane, and so corresponding points in the two images are related by a homography $p'_i \cong H p_i$ (i.e., $p'_i = \alpha_i H p_i$ for some scalar $\alpha_i \neq 0$). SHOW that in this case there exist infinitely many solutions to the fundamental matrix $F$.

Hint: A matrix $M$ is called skew-symmetric if it has the form:

$$ M = \begin{pmatrix} 0 & m_{12} & m_{13} \\ -m_{12} & 0 & m_{23} \\ -m_{13} & -m_{23} & 0 \end{pmatrix} $$

(a) SHOW that $H^T F$ is a skew-symmetric matrix. (You can use the fact that a matrix $M$ is skew-symmetric iff $\forall x : x^T M x = 0$.)

(b) SHOW that this implies the following constraint on $F$:

$$ H^T F + F^T H = 0 $$

(c) SHOW that this provides only 6 linearly independent equations for $F$ (instead of 8).

(d) If the points $P_i$ come from more than just one plane in the scene, how many planes are needed in order to uniquely recover $F$?

Question 7:

In class we saw that the disparity map can be computed by minimizing some energy function.

1. Assume that in addition to the pair of stereo images, you also have a laser scan of the scene: for each pixel $p$ it gives a rough estimation of the depth, $Z_p$, at the pixel. WRITE a new energy function that incorporates this additional information from the laser scan.

   You may assume that that the focal length ($f$) and the distance between the two cameras ($b$) are known, and disparity $p = \frac{fb}{Z_p}$.

2. Depth discontinuities are more likely to happen along edges in the image (e.g., pixels with strong intensity gradient). WRITE a new energy function that incorporates this knowledge.