Computer Vision
Mid Sem - Solutions

1. (*20 points*) State True or False. If always true, justify your answer, else give a counterargument.

a) In order to find the vanishing plane, any three pairs of parallel lines will suffice.

b) Under an affine transformation, the ratio of lengths is preserved.

c) Using structure from motion techniques, we can make reliable length measurements in the scene.

d) The projective distortion in any 3D scene can be removed by mapping l_∞ to its canonical position.

e) Once you estimate the 3×3 homography matrix $\mathbf{H}$ between any pair of images of a generic 3D scene, you can map any point from one image to the other.

f) By simply using an image, it is impossible to measure the angle between two directions in the scene.

g) For estimating $\mathbf{H}_{2\times 2}$, a projective transformation in $\mathcal{P}^1$, we need at least two point correspondences.

h) A reflection matrix can be written as product of two rotation matrices.

i) A weak perspective projection leads to an affine distortion between the image and the scene points.

j) If the aspect ratio for a camera is unity, then the intrinsic matrix $\mathbf{K}$ applies a similarity transformation.

Solution:

a) **False.** All the three directions should be linearly independent.

b) **False.** Under an affine transformation, only the ratio of lengths along parallel or collinear lines are preserved, not in general.

c) **False.** The scale ambiguity cannot be resolved by standard SfM techniques.

d) **False.** If the scene contains only a plane, then the projective distortion can be *completely* removed by mapping l_∞ to $[0, 0, 1]^\top$. However, for a general 3D scene, the *plane* at infinity needs to be mapped to its canonical position.

e) **False.** 3×3 homographies can only map points from one plane to another.

f) **False.** Using vanishing points and intrinsic camera parameters, it is possible to compute the angle between directions corresponding to the vanishing points.

g) **False.** There are three degrees of freedom. Each point correspondence generates one constraints (1-D points) and therefore at least three point correspondences are required.

h) **False.** Since reflection matrices have a determinant smaller than zero, if $\mathbf{A}$ and $\mathbf{B}$ are rotation matrices, then $\det \mathbf{AB} = \det \mathbf{A} \det \mathbf{B} > 0$. Therefore a product of rotation matrices will always be a rotation.

i) **True.** Weak perspective assumes all 3-D points to be at the same depth. This results in uniform scaling that can be absorbed in to intrinsic matrix leading to an affine distortion.

j) **False.** This is true only if the skew factor is also zero.

2. (*20 points*) Show that for a linear system of equations $\mathbf{Ax} = \boldsymbol{b}$, minimizing the sum of squared errors is equivalent to maximizing the correlation between vectors $\mathbf{Ax}$ and $\boldsymbol{b}$. Derive the expression of the maximal correlation estimate of $\mathbf{x}$.

Solution: This question was not framed properly. Therefore everyone gets a score of 20 on this question. Sorry about that. The equivalence only holds true if there is a constraint on the norm of $\mathbf{x}$. This is an important aspect and since it was not stated in the question, we will not consider any answer wrong.

Solution: The least squared solution is obtained by minimizing the following function

$$
\begin{aligned}
&\min_{\mathbf{x}} \ ||\mathbf{Ax} - \boldsymbol{b}||_2^2 \\
\Rightarrow &\min_{\mathbf{x}} \ (\mathbf{Ax} - \boldsymbol{b})^\top (\mathbf{Ax} - \boldsymbol{b}) \\
\Rightarrow &\min_{\mathbf{x}} \ (\mathbf{x}^\top \mathbf{A}^\top \mathbf{Ax} + \boldsymbol{b}^\top \boldsymbol{b} - 2\boldsymbol{b}^\top \mathbf{Ax}) \\
\Rightarrow &\min_{\mathbf{x}} \ (\text{some positive number}) - 2\boldsymbol{b}^\top \mathbf{Ax} \\
\Rightarrow &\max_{\mathbf{x}} \ 2\boldsymbol{b}^\top \mathbf{Ax}
\end{aligned}
$$

The remainder of the derivation follows as done in class (using first derivatives and equating it to zero). The final solution turns out to be $\mathbf{x}^* = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \boldsymbol{b}$.

3. (*20 points*) Consider a vector $(2, 5, 1)^\top$, which is rotated by $\pi/2$ about the Y-axis, followed by a rotation about X-axis by $-\pi/2$ and finally translated by $(-1, 3, 2)^\top$. Find the new coordinates of this vector. What is the coordinate transformation matrix in this case? Where does the origin of the initial frame of reference get mapped to?

Solution:

$$
\begin{aligned}
\mathbf{R}_y &= \begin{bmatrix} \cos\frac{\pi}{2} & 0 & \sin\frac{\pi}{2} \\ 0 & 1 & 0 \\ -\sin\frac{\pi}{2} & 0 & \cos\frac{\pi}{2} \end{bmatrix} \\
\mathbf{R}_x &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\frac{-\pi}{2} & -\sin\frac{-\pi}{2} \\ 0 & \sin\frac{-\pi}{2} & \cos\frac{-\pi}{2} \end{bmatrix} \\
\mathbf{t} &= \begin{bmatrix} -1 \\ 3 \\ 2 \end{bmatrix}
\end{aligned}
$$

The transformation to be applied is:

$$
\begin{aligned}
\mathbf{T} &= \begin{bmatrix} \mathbf{R}_x \mathbf{R}_y & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \\
&= \begin{bmatrix} 0 & 0 & 1 & -1 \\ -1 & 0 & 0 & 3 \\ 0 & -1 & 0 & 2 \end{bmatrix}
\end{aligned}
$$

The origin gets mapped to $\mathbf{T}[0, 0, 0, 1]^\top = [-1, 3, 2, 1]^\top$.

4. (*20 points*) In an image, two vanishing points of a plane are given by $(3, -2, 2)^\top$ and $(5, 1, 0)^\top$ in homogeneous coordinates. Find the line at infinity corresponding to this plane. We want to remove the projective distortion from this image. How would we do it? Explain and provide the transformation you will use. What is the new vanishing line after the transformation is applied? Where do the vanishing points get mapped to?

Solution:

The line at infinity is given by

$$\begin{aligned} l_\infty &= \begin{bmatrix} 3 \\ -2 \\ 2 \end{bmatrix} \times \begin{bmatrix} 5 \\ 1 \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} -2 \\ 10 \\ 13 \end{bmatrix} \end{aligned}$$

The transformation to be applied to image points for removing the projective distortion is $\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 10 & 13 \end{bmatrix}$.

It can be easily verified that the new vanishing line is $(0, 0, 1)^\top$ and the VPs get mapped to $(3, -2, 0)^\top$ and $(5, 1, 0)^\top$ respectively.

5. (*20 points*) What would the essential matrix look like when the camera undergoes (a) a pure translation along the y-axis, (b) a pure translation along the z-axis and (c) a pure rotation about the y-axis. Can you find the epipoles in each of the three cases? If yes, list them, else explain why not. For case (c), the images are also related through a homography. Derive the expression for this homography assuming the intrinsic matrix as $\mathbf{K}$ and rotation matrix $\mathbf{R}$.

Solution:

(a) $\mathbf{E}_{\mathbf{t}_y} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}$

The epipoles are the null space and left null space of the matrix. In this case both are at $(0, 1, 0)^\top$ and therefore are at infinity.

(b) $\mathbf{E}_{\mathbf{t}_z} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

The epipoles in this case are at $(0, 0, 1)^\top$ and are coinciding with the principal point.

(c) $\mathbf{E}_{\mathbf{R}_y} = [\mathbf{t}]_\times \mathbf{R}_y = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{R} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

The epipoles cannot be found out as the Essential matrix is zero.

Let $\mathbf{X}$ be the 3D point in the camera coordinate frame and $\mathbf{x}_1$ and $\mathbf{x}_2$ be the corresponding projections in the first and second images. For a pure rotation, we have the following relationship

$$\begin{aligned} \mathbf{x}_1 &= \mathbf{K}[\mathbf{I} \ \ |\mathbf{0}]\mathbf{X} \\ \mathbf{x}_2 &= \mathbf{K}[\mathbf{R} \ |\mathbf{0}]\mathbf{X} = \mathbf{K}\mathbf{R}\mathbf{K}^{-1}\mathbf{K}[\mathbf{I} \ \ |\mathbf{0}]\mathbf{X} \\ \Rightarrow \mathbf{x}_2 &= \mathbf{K}\mathbf{R}\mathbf{K}^{-1}\mathbf{x}_1 = \mathbf{H}\mathbf{x}_1 \\ \Rightarrow \mathbf{H} &= \mathbf{K}\mathbf{R}\mathbf{K}^{-1} \end{aligned}$$

6. (*20 points*) *Extra Credit*: Suppose you know the vanishing line on the ground plane and the vertical vanishing point, and a reference height along the vertical direction. Using cross ratios, derive the expression for computing the height of another object on the ground plane.

Solution: We define the notation first. Assume homogeneous coordinate representation.

- l_∞ be the vanishing line on the ground plane.
- o_b and r_b be the image point corresponding to the bottom of the object and reference respectively.
- o_t and r_t be the image point corresponding to the top of the object and reference respectively.
- v_z be the vanishing point in the vertical direction.

The vanishing point in the direction of line joining o_b and r_b is

$$v_b = l_\infty \times (r_b \times o_b).$$

The projection of o_t on the line joining the bottom and top of the reference is given by

$$\widehat{o}_t = (v_b \times o_t) \times (r_b \times r_t)$$

The height of the object, then would be

$$H = R\frac{||\widehat{o}_t - r_b||\ ||v_z - r_t||}{||r_t - r_b||\ ||v_z - \widehat{o}_t||}$$

Here the $||\cdot||$ signifies the l_2 distance *after* transforming the points to the Cartesian coordinates.