Instructions:

- Attempt all questions.
- Show intermediate steps to receive partial credit.
- All the best!

Problem 1 (5 points)

Write your name and PID on top of each page. [5 points]

Problem 2 (15 points)

You are given two sets of 3D points. They have been obtained from two views of a scene using a depth sensor (such as Kinect). Thus, they differ from each other by a rigid body transformation, consisting of a rotation $R$ and translation $t$. The correspondence is known between the 3D points. The goal is to robustly estimate the rigid body transformation given by $R$ and $t$.

(a) Let $x \in \mathbb{R}^3$ be the points in the first view and and $x' \in \mathbb{R}^3$ be the points in the second view. The transformation between the points is given by $x' = Rx + t$. Express this rigid-body transformation as a single matrix, $A$. *Hint: use homogeneous coordinates.* [4 points]

$$A = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix}$$

(b) How many degrees of freedom does $A$ have? [4 points]

$$6 \ (3 \text{ for } R, \ 3 \text{ for } t)$$
(c) Now, suppose there are some outliers among the given correspondences and you wish to use RANSAC for outlier removal. What is the size of samples you need to pick? [2 points]

6

(d) Describe how you would implement RANSAC for this problem. Assume you have a method to compute \( A \) using the chosen samples. [3 points]

For large enough \( N \)
Randomly Sample 6 for correspondences
Estimate \( A \) (or \( R, t \))
Find number of points for which \( \| \mathbf{x} - (R\mathbf{x} + t) \| ^2 < \text{threshold} \)
Return \((R, t)\) with largest inlier set size

(e) On what factors does the required number of trials for RANSAC depend? [2 points]

Fraction of outliers
Size of samples
Probability of correctness

Problem 3  (15 points)

(a) Given corresponding points \( x_i \in \mathbb{R}^2 \) in image 1 and \( x'_i \in \mathbb{R}^2 \) in image 2, for \( i = 1, \cdots, n \). What is the minimum \( n \) for any method to be able to estimate the fundamental matrix? [2 points]

7

(b) What is the minimum \( n \) for any method to be able to estimate the essential matrix? [2 points]

5
(c) Suppose \( n = 100 \). Describe a method to estimate the fundamental matrix (you may assume no outliers among the correspondences).

\[
\begin{align*}
\text{Construct } & \mathbb{F} = 0. \\
\text{Use SVD to get } & \mathbb{F} . \\
\text{Reshape to } & \mathbb{F}_{3 \times 3} . \\
\text{Find closest } & \mathbb{F} \text{ rank 2 } \mathbb{F} \text{ by setting } s_{3,3} = 0 \text{ for } \mathbb{F}_{3 \times 3} . \\
\text{[If people say use RANSAC and 8-point, that’s also fine.]}
\end{align*}
\]

(d) Consider a camera that undergoes pure translation, by one unit along the direction of its principal axis. What is the essential matrix induced between the two views? \hspace{1cm} \text{[3 points]}

\[
\begin{align*}
\mathbb{E} &= (0,0,-1)^T, \quad R = I . \\
\mathbb{E} &= [\mathbb{E}]_x R = [\mathbb{T}]_x = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} . \\
\text{It’s okay if they do } & \mathbb{E} = (0,0,1)^T, \text{ then get } -\mathbb{E} . \text{ Give 2.5 points for it.}
\end{align*}
\]

(e) Suppose the above camera has a focal length of 100 pixels, aspect ratio 1, zero pixel skew and principal point at \((0,0)^T\). What is the fundamental matrix between the above two views? \hspace{1cm} \text{[3 points]}

\[
\begin{align*}
\mathbb{K} &= \begin{bmatrix} -100 & 0 & 0 \\ 0 & -100 & 0 \\ 0 & 0 & 1 \end{bmatrix} . \\
\mathbb{F} &= \mathbb{K}^{-1} \mathbb{E} \mathbb{K}^{-1} = \begin{bmatrix} 0 & 0.0001 & -1 \\ 0.0001 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix} . \\
\text{It’s okay if they start with } -\mathbb{K} , \text{ give full 3 points (same } \mathbb{E} \text{ anyway).}
\end{align*}
\]

(f) For the above fundamental matrix, what are the epipoles in the two views? Provide an intuitive explanation for the location of the obtained epipoles. \hspace{1cm} \text{[3 points]}

\[
\text{In 1st view, } \text{null}(\mathbb{F}) = \text{null}(\mathbb{F}^T) = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} . \quad \rightarrow \text{ 2 pts.}
\]

Pure forward motion, centered cameras, so camera center observed at middle of image plane. \quad \rightarrow \text{ 1pt.}
Problem 4  (15 points)

(a) Compute the second moment matrix for the $4 \times 4$ image regions highlighted in blue in the above two Images 1 and 2. Shaded squares have intensity value of 1 and unshaded squares have intensity value of 0. Use backward differencing for computing derivatives.  

\[ C_1 = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}, \quad C_2 = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \]

\[ I_x = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad I_y = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad I_x = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \]

(b) Use the above two second moment matrices to explain which of the above two image regions has greater “cornerness”?

\[ C_1 \text{ is rank 1, } C_2 \text{ is rank 2, so } C_2. \]

(Fine if people compute eigenvalues.)

(c) Draw the histograms of gradient directions for the highlighted $4 \times 4$ image regions in the above two Images 1 and 2. The histograms should have 8 bins, each having a width of $45^\circ$, centered at $0, 45^\circ, 90^\circ$ and so on. Use backward differencing for computing derivatives.  

\[ 3 \quad 1 \quad 1 \quad 1 \quad 4 \quad 0 \quad 45 \quad 90 \quad 135 \quad 180 \quad 225 \quad 270 \quad 315 \quad 360 \]
(d) How is invariance to image-plane rotations obtained in SIFT? [2 points]

Compute histogram of orientations.
Align image patch with dominant orientation.

(e) What is the utility of the ratio test in descriptor matching? [2 points]

Let $f_2$ and $f_2'$ be closest descriptors in image 2, compared to $f_1$ in image 1.

Ratio test: discard match if $\frac{\|f_2 - f_2'\|}{\|f_1 - f_2\|}$ is close to 1.

Utility: discard ambiguous matches (or retain discriminative ones).

[Fine if answer only contains utility.]

Problem 5 (15 points)

We will consider how to triangulate the position of a 3D point, given possibly noisy observations of in two known cameras. Consider two cameras, centered at known 3D positions $c_1$ and $c_2$. An unknown 3D point $Q$ is observed on the two cameras at points $\tilde{q}_1$ and $\tilde{q}_2$. If $f$ is the known focal length of the camera, then note that we can write the image points as 3D points, given by

$$q_1 = \begin{bmatrix} \tilde{q}_1 \\ f \end{bmatrix}, \quad q_2 = \begin{bmatrix} \tilde{q}_2 \\ f \end{bmatrix}.$$ 

A point on a line that passes through two 3D points $a$ and $b$ may be expressed as $x(\lambda) = a + \lambda b$. Let $x_1(\lambda_1)$ denote a 3D point that lies on the ray $l_1$ backprojected from the first camera center and
passing through the first image point. Similarly, let \( x_2(\lambda_2) \) denote a 3D point that lies on the ray \( l_2 \) backprojected from the second camera center and passing through the second image point.

(a) Write the expressions for \( x_1(\lambda_1) \) and \( x_2(\lambda_2) \), in terms of \( c_1, c_2, q_1, q_2 \). [2 points]

\[
\begin{align*}
x_1 &= (1-\lambda_1)c_1 + \lambda_1 q_1 \\
x_2 &= (1-\lambda_2)c_2 + \lambda_2 q_2
\end{align*}
\]

(b) In the presence of noise, will the rays \( l_1 \) and \( l_2 \) intersect, in general? [2 points]

\( \text{No} \).

(c) If the answer above is “no”, we need to find the point \( x \) that comes the closest to lying on both \( l_1 \) and \( l_2 \). Another way to express this is to say we want the \( \lambda_1^* \) and \( \lambda_2^* \), such that the distance between \( x_1 \) and \( x \) is minimized. Thereafter, the closest point to the two rays is simply given by

\[
x^* = \frac{x_1(\lambda_1^*) + x(\lambda_2^*)}{2}.
\]

Thus, we first wish to solve

\[
[\lambda_1^*, \lambda_2^*] = \arg \min_{\lambda_1, \lambda_2} g(\lambda_1, \lambda_2),
\]

where \( g(\lambda_1, \lambda_2) = \|x_1 - x\|_2^2 \). Write the expression for \( g \) in terms of \( c_1, c_2, q_1 \) and \( q_2 \). [2 points]

\[
g(\lambda_1, \lambda_2) = \left\| (1-\lambda_1) c_1 + \lambda_1 q_1 - (1-\lambda_2) c_2 - \lambda_2 q_2 \right\|_2^2
\]

\[
= \left\| c_1 + \lambda_1 (q_1 - c_1) - c_2 - \lambda_2 (q_2 - c_2) \right\|_2^2
\]

\[
= \left\| (c_1 - c_2) + \lambda_1 (q_1 - c_1) + \lambda_2 (q_2 - c_2) \right\|_2^2
\]
(d) What are the conditions on the partial derivatives of $g(\lambda_1, \lambda_2)$ for there to be a local minimum with respect to $\lambda_1$ and $\lambda_2$?

$$\frac{\partial g}{\partial \lambda_1} = 0 \quad \frac{\partial g}{\partial \lambda_2} = 0$$

(Five if stated in words.)

(e) Write the above conditions explicitly in terms of $c_1, c_2, q_1, q_2, \lambda_1$ and $\lambda_2$. What type of equation is each condition, in terms of the unknowns $\lambda_1$ and $\lambda_2$?

$$2(q_1-c_1)^T((c_1-c_2) + \lambda_1(q_1-c_1) - \lambda_2(q_2-c_2)) = 0$$
$$-2(q_2-c_2)^T((c_1-c_2) + \lambda_1(q_1-c_1) - \lambda_2(q_2-c_2)) = 0$$

(f) Is there a simple way to solve the above equations? It is sufficient to explain how, you do not have to actually solve them (but no problem if you do solve them).

The above are two linear equations in two variables $\lambda_1, \lambda_2$.

$$\alpha_1 \lambda_1 + \beta_1 \lambda_2 = \gamma_1$$
$$\alpha_2 \lambda_1 + \beta_2 \lambda_2 = \gamma_2$$

$$\alpha_1 = (q_1-c_1)^T(q_1-c_1), \quad \beta_1 = -(q_1-c_1)^T(q_2-c_2), \quad \gamma_1 = -(q_1-c_1)^T(c_1-c_2)$$
$$\alpha_2 = (q_2-c_2)^T(q_1-c_1), \quad \beta_2 = -(q_2-c_2)^T(q_2-c_2), \quad \gamma_2 = -(q_2-c_2)^T(c_1-c_2)$$

Fine to say just "two linear eqns in $\lambda_1, \lambda_2$", or from 2x2 linear system and invert, or similar variations.
Problem 6  (15 points)

(a) What are the two ways in which stereo rectification makes the disparity estimation problem more efficient? [3 points]

- Epipolar lines become scan lines.
  - 2D search reduces to 1D.
  - Efficiently implemented in hardware.

(b) What is the advantage of NCC over SSD? When is SSD preferable over NCC? [3 points]

- NCC → better invariance to illumination changes.
- SSD → faster to compute.

(c) What are the relative advantages and disadvantages of using a small or large window size in stereo matching? [3 points]

- Small → more detail, more noise in estimation.
- Large → less detail, less noise in estimation.

(d) What are the relative advantages and disadvantages of using a small or large baseline in two-view stereo? [3 points]

- Small → Easier matching, more uncertainty in depth.
- Large → Harder matching, less uncertainty in depth.

(d) Explain the possible needs for a smoothness prior in stereo estimation. [3 points]

- Textureless regions have uncertainty (more ambiguity) in disparity estimation.
- Noise or outliers may lead to inaccurate estimation.

Smoothness prior encourages similar values in a neighborhood, so provides a mechanism to reduce ambiguity and prevent inaccuracies due to noisy data.