Lecture: Hands-on AI Based 3D Vision
Assignment 01

Due: May 13th, 12:00am

1 3D Transformations (14 Points)

Recall that using homogeneous coordinates, a projective 3D transformation can be described by an invertible 4×4 matrix $\mathbf{H}$. There is a natural hierarchy of projective transformations, each of which less restrictive:

Type	Matrix	Preserves
Rigid	$\begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix}$	Lengths
Similarity	$\begin{bmatrix} s\mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix}$	Angles
Affine	$\begin{bmatrix} \mathbf{A} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix}$	Parallelism
Projective	$\mathbf{H}$	Straight lines

Here $\mathbf{t} \in \mathbb{R}^3$, $\mathbf{R} \in \mathrm{SO}(3)$, $\mathbf{A} \in \mathrm{GL}(3)$, $s \in \mathbb{R}$.

(a) For each type of transformation show the preserving property. I.e. show that

   i) rigid transformations preserve lengths (2P)

ii) similarity transformations preserve angles (2P)

iii) affine transformations preserve parallelism (2P)

iv) projective transformations preserve straight lines (2P $\star$ **bonus** $\star$)

(b) Let $\mathbf{x_1x_2}$, and $\mathbf{y_1y_2}$ be two line segments in 3D space, and let $\mathbf{H}$ be an affine transformation matrix.

i) Show that if the line segments are parallel, then their length ratio is preserved under $\mathbf{H}$. (2P)

ii) Give a counterexample that proves that this is not true for non-parallel line segments. (2P)

(c) For each type of transformation, determine the number of degrees of freedom. Explain your answers briefly. (1P for each)
Hint: Consider the fact that projective transformations act in homogeneous coordinates, where $\tilde{\mathbf{x}} \simeq k\tilde{\mathbf{x}}$.

2 Estimating Camera Pose from 3D-2D Correspondences (15 Points)

In this exercise we want to in some sense invert the projection process. Usually, we are given a camera matrix $\mathbf{P} = \mathbf{K}\begin{bmatrix}\mathbf{R} & \mathbf{t}\end{bmatrix} \in \mathbb{R}^{3\times4}$, and a set of 3D points $\{\mathbf{x}_i\}_{i\in I}$, and want to compute the 2D image points $\mathbf{u}_i$ satisfying

$$\begin{bmatrix}\mathbf{u}_i\\1\end{bmatrix} \propto \mathbf{P}\begin{bmatrix}\mathbf{x}_i\\1\end{bmatrix}. \tag{1}$$

In this exercise, we are going to work on the other direction: If we are given the 3D points paired with their image points $\{\mathbf{x}_i, \mathbf{u}_i\}_{i\in I}$, how can we recover the camera parameters $\mathbf{K}, \mathbf{R}$, and $\mathbf{t}$? We will for the sake of simplicity assume that $\mathbf{K} = \mathbf{I}$, and therefore

$$\mathbf{P} = \begin{bmatrix}\mathbf{R} & \mathbf{t}\end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix}.$$

In this exercise you will design a linear approach to solving for the entries of $\mathbf{P}$. Therefore, we flatten $\mathbf{P}$ into a single vector of unknowns

$$\boldsymbol{\theta}^T = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 & r_{21} & r_{22} & r_{23} & t_2 & r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix}$$

(a) (1P) For this part, we want you to show how to go from the general case (where $\mathbf{K} \neq \mathbf{I}$, but $\mathbf{K}$ is known), to the calibrated case (where $\mathbf{K} = \mathbf{I}$). Given a set of image points $\{\mathbf{u}_i\}_{i\in I} \subseteq \mathbb{R}^2$, that were captured by a general camera with known $\mathbf{K}$, derive the image points $\{\mathbf{u}_i^c\}_{i\in I}$ as a calibrated camera with $\mathbf{K}^c = \mathbf{I}$ would observe them.

(b) (6P) Given a single correspondence pair $\mathbf{x} \in \mathbb{R}^3$, $\mathbf{u} \in \mathbb{R}^2$, rewrite eq. (1) to obtain one or multiple linear equations of the form

$$\mathbf{v}^T \boldsymbol{\theta} = 0$$

(c) (1P) How many equations can you get from one correspondence pair? How many correspondences do you need to setup a system of linear equations

$$\mathbf{M}\boldsymbol{\theta} = \mathbf{0}$$

such that we can solve for $\boldsymbol{\theta}$?

(d) (1P) In order to make the method more robust to noise, we want to use more than the minimum number of correspondences from (b). Suppose we have an overparameterized system of the form $\mathbf{M}\boldsymbol{\theta} = \mathbf{0}$ where $\mathbf{M} \in \mathbb{R}^{D\times 12}$. Explain the steps necessary to get the least squares solution, i.e. the vector $\boldsymbol{\theta}^* \in \mathbb{R}^{12}$ which solves the optimization problem

$$\min_{\|\boldsymbol{\theta}\|=1} \|\mathbf{M}\boldsymbol{\theta}\|_2^2$$

Hint: SVD

(e) (4P) Explain how to recover $\mathbf{R}$ and $\mathbf{t}$ from $\boldsymbol{\theta}^*$ in (c).
Hint: What are the properties of $\mathbf{R}$ *that make it a rotation matrix? How can we enforce them? Think of SVD again!*

(f) (2P) Given $\mathbf{R}$ and $\mathbf{t}$, what is the camera-to-world rotation matrix, and the camera position in world frame.

(g) Implement your algorithm in the notebook.