Master CIMET and 3DMT - Computer Vision course

Exam May 2015 - 2h without documents

(3 parts with a total of 18 questions accounting for 24 points, the exam will be scored for 20 points)

Part 1 (*12 points*) : 3D reconstruction

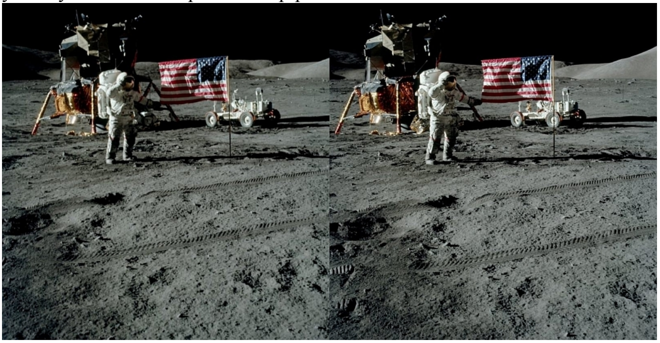
Question 1 *(2 points)***:** Show how to map the coordinates of a 3D point in the scene to the coordinates of a 2D point in the image.

Question 2 *(1 point)* : Is it possible to recover depth from a monocular point of view? Justify

Question 3 *(2 points)*: Show how to map the coordinates of interest points in multiple 2D images of the same 3D scene viewed from different points of view.

Question 4 *(1 point)*: Is it possible to recover depth from a binocular point of view?

Question 5 *(2 points)*: The two images below are a stereo pair taken using parallel cameras, aligned horizontally. Describe a framework (i.e. a solution) for matching corresponding features in the two images and estimating their disparities. State clearly how your method exploits the epipolar constraint.



Question 6 *(2 points)*: Explain how to obtain estimates of the 3-D positions of the objects in the scene above from stereo disparities. What information about the cameras is required for this?

Question 7 *(2 points)*: If only one of the two images in the pair above was available, what computational processes might be used to make inferences about the 3-D structure of the scene? What assumptions are involved in applying these processes?

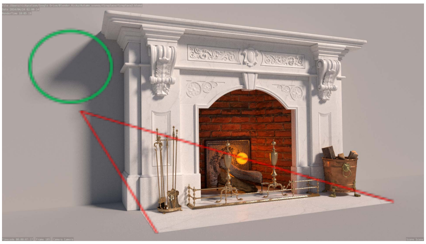
See correction of exam 2014

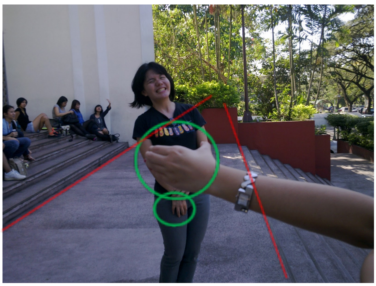
Part 2 *(5 points)* **:** vanishing point perspective

Question 1 *(3 points)***:** For each image shown below, estimate (draw) manually the position(s) (in the image plane) of the major vanishing point(s), corresponding to the building(s) orientations.

Question 2 *(2 points)* : What are the main cues of 3D perception in each of the images below? Draw on each image these main pictorial cues.

See suggests indicated below





Part 3 *(7 points)***: 3D** Projection on a 2D Plane

The equation below represents the transformation M1 of a point in one three-dimensional cartesian reference frame (denoted w) into another (denoted c).

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$

Question 1 *(0.5 point)* : Why do the two vectors on each side contain *four* elements not just three?

See course / use of homogeneous coordinates

Question 2 *(0.5 point)* : Concerning the matrix: What does the symbol R represent ? What does the symbol T represent?

See course

Question 3 *(0.5 point)* : What is the dimensionality of the matrix: (a) 3x3, (b) 4x4, or (c) 4x3?

(b) R is of dim 3x3 and T of dim 3x1

Question 4 *(0.5 point)* : Now write out the matrix in full for the case of a coordinate transformation of: a <u>horizontal</u> translation (in x) by 1 m, and a <u>roll</u> rotation about the depth (z) axis of 90 degrees.

See course

Question 5 *(1 point)* : The relationship between a 3D point at world coordinates (X_w, Y_w, Z_w) and its corresponding 2D pixel at image coordinates (u, v) can be defined as a projective transformation using a 3x4 camera projection matrix P. P can be decomposed into a product of matrices that contain elements expressed in terms of the intrinsic and extrinsic camera parameters, as follows: $P = K\,\Pi_0\,M1$ with K is a 3x3 matrix and Π_0 a 3x4 matrix. What represent the coefficents of the matrix K ?

See course

Question 6 *(1 point)* : Given the 3x4 camera matrix:

$$P = \begin{bmatrix} 5 & -14 & 2 & 17 \\ -10 & -5 & -10 & 50 \\ 10 & 2 & -11 & 19 \end{bmatrix}$$

and a 3D point of the reference frame (denoted w) defined in homogeneous coordinates by:

$\mathbf{X} = [0\ 2\ 2\ 1]^T$

What are the cartesian coordinates (denoted (u, v)) of the projection of X on the image plane ?

$su = -7$, $sv = 20$, and $s = 1$, so $(u, v) = (-7/1, 20/1) = (-7, 20)$

Solution

Question 7 *(1 point)* : Let us consider that: - we use an ideal pinhole camera with a focal length 5mm; - each pixel is of size 0.02 mm x 0.02 mm and; - the image principal point is at pixel (500, 500). Note that pixel coordinates start at (0, 0) in the upper-left corner of the image.

What is the 3x3 camera calibration matrix, **K**, for this camera configuration?

Solution

$$K = \begin{bmatrix} fk_u & 0 & u_0 \\ 0 & fk_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 5\,\frac{1}{0.02} & 0 & 500 \\ 0 & 5\,\frac{1}{0.02} & 500 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 250 & 0 & 500 \\ 0 & 250 & 500 \\ 0 & 0 & 1 \end{bmatrix}$$

Question 8 *(1 point)* : Assuming the world coordinate frame is aligned with the camera coordinate frame (i.e. their origins are the same and their axes are aligned) and the origins are at the camera's pinhole, what are the coefficients of the 3x4 matrix M1 that represents the extrinsic, rigid body transformation between the camera coordinate system and the world coordinate system ?

Solution

$$P_r = \begin{bmatrix} r_{11} & r_{12} & r_{13} & T_x \\ r_{21} & r_{22} & r_{23} & T_y \\ r_{31} & r_{32} & r_{33} & T_z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Question 9 *(1 point)* : Combining your results from questions 7 and 8, compute the projection of a scene point of coordinates (100, 150, 800) into image plane coordinates.

Solution

$$\begin{bmatrix} su \\ sv \\ s \end{bmatrix} = KP_r \begin{bmatrix} 100 \\ 150 \\ 800 \\ 1 \end{bmatrix} = \begin{bmatrix} 250 & 0 & 500 & 0 \\ 0 & 250 & 500 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 100 \\ 150 \\ 800 \\ 1 \end{bmatrix} = \begin{bmatrix} 425000 \\ 437500 \\ 800 \end{bmatrix}$$

so, the scene point projects to pixel (425000/800, 437500/800) = (531, 547)