Exam on Computer Vision part of *Human Vision and Computer Vision* course
Master CIMET - June 2010

(All documents authorized - duration: 2 hours)

Problem 1: Image Projection

The equation below represents the transformation 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: Why do the two vectors on each side contain *four* elements not just three?

See Lecture 1

Question 2: What is the four-elements vector known as?

Question 3: Concerning the matrix: What does the symbol R represent and what are its dimensions? What does the symbol T represent and what are its dimensions?

Question 4: Is the entire matrix: (a) 3x3, (b) 4x4, or (c) 4x3?

Question 5: Now write out the matrix in full for the case of a coordinate transformation of: a horizontal translation (in x) by 1 m, and *roll* about the depth (z) axis of 90 degrees.

Question 6: Did the use of these 4-D vectors/matrices simplify your task?

Problem 2: Image Projection

The relationship between a 3D point at world coordinates (X, Y, Z) 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:

$$\mathbf{P} = \mathbf{K}[\mathbf{R} \mid \mathbf{T}] = \begin{bmatrix} fk_u & 0 & u_0 \\ 0 & fk_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix}$$

Question 1: What are the intrinsic camera parameters and the extrinsic camera parameters in this equation?

See Lecture 3 Part 2

Given the 3x4 camera matrix

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

and a 3D point in homogeneous coordinates

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

Question 2: What are the Cartesian coordinates of the point **X** in 3D?

(0/1, 2/1, 2/1) = (0, 2, 2)

What are the Cartesian image coordinates, (u, v), of the projection of **X**?

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

An ideal pinhole camera has focal length 5mm. Each pixel is 0.02 mm x 0.02 mm and the image principal point is at pixel (500, 500). Pixel coordinates start at (0, 0) in the upper-left corner of the image.

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

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

Question 4: 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 is the 3 x 4 matrix that represents the extrinsic, rigid body transformation between the camera coordinate system and the world coordinate system?

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

Question 5: Combining your results from questions 3 and 4, compute the projection of scene point (100, 150, 800) into image coordinates.

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

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

Problem 3 : Computer Vision models

See Lecture2 part2

Question 1: David Marr’s idea of visual *process* proposed three levels of representation, name them, and describe the types of features Marr used to express the *first two* levels.

Question 2: Describe the Gestalt Theory and give some examples that support this Perceptual Organization Theory

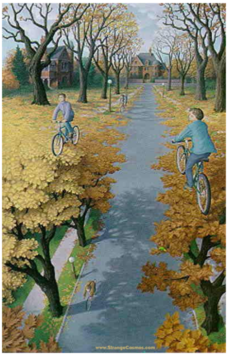
Question 3: Observers have a tendency to interpret shape and size in 3D - often unaware of 2D size. Give an example to illustrate this tendency.

Question 4: “Marr's theory for 3D vision must be wrong”. Describe Marr’s theory and discuss this statement in some detail.

Question 5: Draw on each of the following images which are the main geometrical cues which create the illusion of a 3D world?

Perspective, one vanishing point, size of trees

Perspective, one vanishing point, shadows, size of objects, etc.


www.StrangeCosmos.com


Seurat