Master COSI, CIMET, MLDM and 3DMT - Computer Vision course

Exam March 2016 - 3h without documents

(3 parts with a total of 12 questions accounting for 23 points, the exam will be scored for 20 points)

Part 1: 3D reconstruction from stereo vision

Question 1 *(2 points)***:** A schematic picture of a stereo pair of cameras looking at a scene point is shown in the figure below. Name all the labeled quantities 1-5.

Solution: remind explanations given during lectures

1= epipolar plane
2= 2D (image) projection of the scene point
3= optical centers of cameras
4= epipoles on epipolar lines
5= baseline

Question 2 *(1 point)***:** What is the process of converting the images taken by a pair of cameras with the geometry below, to those taken by a pair of parallel cameras called?

Solution: rectification

Question 3 *(1 point)***:** What constraint reduces stereo correspondence to a line search? With reference to the figure, explain how this constraint arises.

Solution:
- Epipolar Constraint: – Matching points lie along corresponding epipolar lines; – Reduces correspondence problem to 1D search along conjugate epipolar lines; – Greatly reduces cost and ambiguity of matching.
- Remind explanations (and figures) given during lectures about the matching of corresponding points using the epipolar geometry. Using a block matching technique (correlation measure) we can easily match corresponding

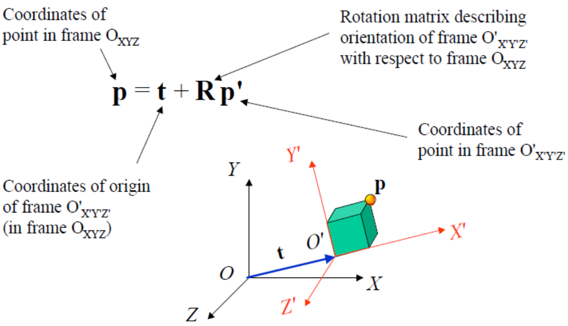
features (computed from interest point technique) along epipolar line.
- Epipolar lines must be parallel to the baseline (i.e. stereo images must be rectified).

Question 4 *(1 point)***:** Why is correspondence hard? Describe how the SSD (Sum of Squared Differences) algorithm for correspondence works.

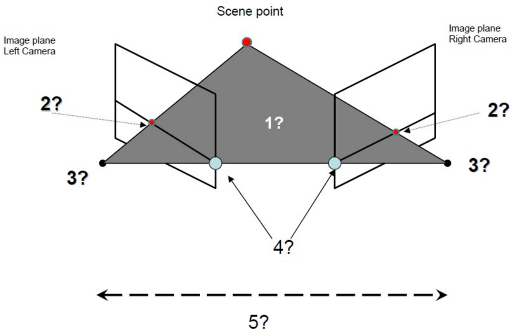
Solution:
- In case of occlusion, texture less areas, or similar patterns the matching process of corresponding points can fail.
- One technique to match corresponding points is to minimize a cost function (e.g. the SSD between intensity values of pairs of points) and to compute differences from the neighbor of interest points (block matching process). See figure / question 5 of exam 2014.

Question 5 *(1 point)***:** A scene point P has coordinates (X1,Y1,Z1) in the coordinate system that is centered at one location and (X2,Y2,Z2) in the coordinate system centered at another location. Write down a general rigid transformation that relates the coordinates of the point P in these two coordinate systems.

Solution: remind example below given during lecture 4



Question 6 *(1 point)***:** Express this transformation as a matrix-vector product using homogeneous coordinates. Describe how you would get real ("Euclidean") coordinates from the homogeneous coordinates.



Solution: remind example below given during lecture 4

When $\boldsymbol{X}_c = R\boldsymbol{X}_w + T$

then

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

Remind that this transform is defined up to a scaling factor (s)

Part 2: Calibration

We seek to calibrate an undistorted camera with a planar calibration object. The object possesses M distinct features. The location of those features in the object coordinate frame are unknown (and they are not arranged in a checkerboard).

In answering the questions below, you might want to use the following notation/equations

---

$$\begin{pmatrix} \tilde{X}^C \\ \tilde{Y}^C \\ \tilde{Z}^C \end{pmatrix} = R \begin{pmatrix} X^W \\ Y^W \\ Z^W \end{pmatrix} + \begin{pmatrix} T_X \\ T_Y \\ T_Z \end{pmatrix}$$

$$R = \begin{pmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \cos\varphi & 0 & \sin\varphi \\ 0 & 1 & 0 \\ -\sin\varphi & 0 & \cos\varphi \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\psi & \sin\psi \\ 0 & -\sin\psi & \cos\psi \end{pmatrix}$$

$$\begin{pmatrix} x_{\text{im}} \\ y_{\text{im}} \end{pmatrix} = \left\lfloor \begin{pmatrix} \frac{f}{s_x} \frac{\tilde{X}^C}{\tilde{Z}^C} + o_x + 0.5 \\ \frac{f}{s_y} \frac{\tilde{Y}^C}{\tilde{Z}^C} + o_y + 0.5 \end{pmatrix} \right\rfloor$$

Question 1 *(2 points)***:** What are the free (i.e. intrinsic and extrinsic) parameters that can be recovered through the calibration process?

Let us assume that there are 2M parameters for the unknown feature locations $\{X^W_j; Y^W_j\}$ on the planar calibration object and that $Z^W_j = 0$ for all features j. So, we have to add the object parameters $\{X^W_j; Y^W_j\}$ to the set of free parameters. However, as the definition of the object reference frame is somewhat arbitrary we can define $X^W_1 = Y^W_1 = 0$, and $X^W_2 = 0$. This reduces the number of recoverable parameters by 3. This definition constrains the location of all other features on the board.

Solution:

There are five intrinsic parameters but only four can be recovered. The recoverable intrinsic parameters are $\{\frac{f}{s_x}, \frac{s_x}{s_y}, o_x, o_y\}$.

There are $2M$ parameters for the unknown feature locations on the planar calibration object. This is because we know $\tilde{Z}^W_j = 0$ for all features j. So we add the object parameters $\{X^W_j, Y^W_j\}$ to the set of free parameters. However, the definition of the object reference frame is somewhat arbitrary. We can therefore define $X^W_1 = Y^W_1 = 0$, and $X^W_2 = 0$. This reduces the number of recoverable parameters by 3. This definition constrains the location of all other features on the board.

The extrinsic parameters are $\{\phi_i, \varphi_i, \psi_i, T_{X,i}, T_{Y,i}, T_{Y,i}\}$, for each image i.

In summary, we have $4 + 6K + 2M - 3 = 6K + 2M + 1$ unknown parameters, with K being the nymber of images and M the number of features on the calibration pad.

Question 2 *(2 points)***:** Let be K the number of images acquired and M the number of features on the planar calibration object. Compute the number of unknown parameters defined above.

Question 3 *(4 points)***:** How many parameters have to be estimated to calibrate the camera from this calibration object? How many constraints are being provided by each image of the calibration pattern? What are the lower bounds for M and K? Provide exact formulae, one of the form K 3… and one of the form M 3… For K = 3 images, what is the minimum M? For M = 3 features, what is the minimum K?

Solution:

We have to estimate $4 + 6K + 2M - 3$ parameters from $2KM$ constraints. Hence we have

$$\begin{aligned} 2KM \geq 1 + 6K + 2M \quad &\Longleftrightarrow \quad 2KM - 6K \geq 1 + 2M \\ &\Longleftrightarrow \quad K(2M - 6) \geq 1 + 2M \\ &\Longleftrightarrow \quad K \geq \frac{1 + 2M}{2M - 6} = \frac{7 + 2M - 6}{2M - 6} = 1 + \frac{7}{2M - 6} (1) \end{aligned}$$

and (assuming $M > 3$):

$$\begin{aligned} K \geq 1 + \frac{7}{2M - 6} \quad &\Longleftrightarrow \quad K - 1 \geq \frac{7}{2M - 6} \\ &\Longleftrightarrow \quad 2M - 6 \geq \frac{7}{K - 1} \\ &\Longleftrightarrow \quad 2M \geq 6 + \frac{7}{K - 1} \\ &\Longleftrightarrow \quad M \geq 3 + \frac{7}{2K - 2} \end{aligned}$$

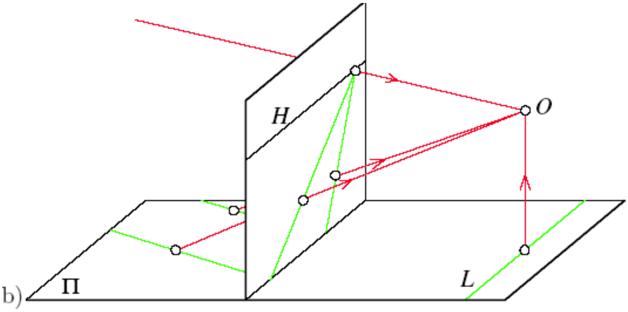
For $K = 3$ images, we need $M \geq 5$ features. With only $M = 3$ features, we cannot calibrate, no matter how many images. We need at least $M = 4$ features.

Part 3: Perspective projection

Question 1 *(2 points)***:** In general, under what conditions will a line viewed with a pinhole camera have its vanishing point at infinity? May an image have more than three orthogonal vanishing points?

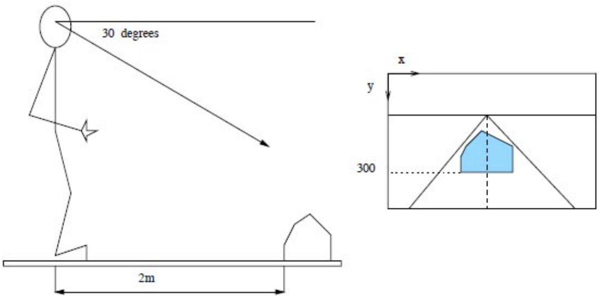
Solution: remind explanations given during lectures
- Parallel lines in 3D world intersect at infinity in 2D Projective Space
- In projective space, all lines intersect in a point, even parallel lines.
- Geometrically we can only have three orthogonal vanishing points with a 3D perspective projection

Question 2 *(2 points)***:** Explain why in the following image the green lines drawn in the image plane are not parallel, oppositely to the green lines in the world coordinate frame?



Solution: remind explanations given during lectures
- Parallel lines in 3D world intersect in the 2D Projective Space, the location of the vanishing point depends of the focal (f) and of the orientation (X,Y) of lines (see equation of the pinhole model)

Consider a person standing on the road viewing the road at the viewing angle α.



Let us assume that the y image coordinate is on the horizon line and that the observer coordinate of the horizon is related to the coordinate of the vanishing point, which is an intersection of the two parallel lines in the ground plane. Let us also assume that the y-axis is pointing downwards, x-axis is oriented to the right and the z-axis out of plane. Then, the coordinates of the vanishing point are:

$$x = \frac{X_c + \lambda v_1}{Z_c + \lambda v_3} \text{ and } y = \frac{Y_c + \lambda v_2}{Z_c + \lambda v_3}$$

where $v_c = [v_1, v_2, v_3]^T$ is the direction vector of a line in the camera coordinate frame and $X = [X_c, Y_c, Zc]^T$ is the base point of the line.

Consider two lines in the world coordinate frame which lie in the ground plane with direction vector $v_w = [0, 0, 1]^T$. Then, the same direction vector in the camera frame will be $v_c = [0, -\sin\alpha, \cos\alpha]$. Suppose that y' is the actual retinal coordinate of the horizon, where:

$$y' = (y - 200)/f$$

Then the y' coordinate of the vanishing point (and hence the horizon) can be obtained by letting $y' \to \infty$:

$$y' =_{\lambda\to\infty} \frac{Y_c - \lambda \sin\alpha}{Z_c + \lambda\cos\alpha} = \frac{-\sin\alpha}{\cos\alpha}. \quad (1)$$

So $\alpha = -\text{atan}(y')$ can be directly computed from the y' coordinate of the horizon.

Question 3 *(4 points)***:** Suppose that computed viewing angle is 30° and that there is an obstacle in front the person at the distance of 2 meters from the feet. Consider that the parameters of the imaging system can be well approximated by a pinhole camera where the resulting image is of resolution 400×400, the focal length is f = 30 and the image of the projection center is the center of the image. The y-coordinate of the obstacle in the image is 300 pixels. How tall is the person?

There are couple ways how to answer to this question.

The first step consists to compute the coordinate transformation between the feet coordinate frame {w} and the eye coordinate frame {c} using the following form:

Xw = R Xc + T = ? Xc + ? (2)

a) What are the values missing in the equation (2)?

The second step consists to compute the inverse transformation of (2) in order de estimate Xc from Xw.

b) Inverse the transformation (2) in order to compute Xc from Xw

If we suppose that the coordinate of the obstacle in the feet coordinate system is $[X_w, 0, Z_w]^T$, then the third step consists to compute the eye y-coordinate y' (computed as in (1)) in function of its 3D counterpart.

c) Compute *y' in function of Y_c and Z_c, next deduce t_y*

Solution:

There are couple ways how this can be done. The coordinate transformation between the feet coordinate frame {w} and the eye coordinate frame {c} has the following form

$$\boldsymbol{X}_w = R\boldsymbol{X}_c + T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha \\ 0 & -\sin\alpha & \cos\alpha \end{bmatrix} \boldsymbol{X}_c + \begin{bmatrix} 0 \\ -t_y \\ 0 \end{bmatrix}$$

where t_y is the height of the person and α is the viewing angle. The inverse transformation is then expressed as $\boldsymbol{X}_c = R^T\boldsymbol{X}_w - R^T T$

$$\boldsymbol{X}_c = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{bmatrix} \boldsymbol{X}_w + \begin{bmatrix} 0 \\ t_y\cos\alpha \\ t_y\sin\alpha \end{bmatrix}$$

Suppose that the coordinate of the obstacle in the feet coordinate system is $[X_w, 0, Z_w]^T$, then the retinal y-coordinate y' (computed as in a)) is related to its 3D counterpart as

$$y' = \frac{Y_c}{Z_c} = \frac{-\sin\alpha Z_w + \cos\alpha t_y}{\cos\alpha Z_w + \sin\alpha t_y}.$$

Since all the above quantities except t_y are known we can compute it as

$$t_y = \frac{y'\cos\alpha Z_w + \sin\alpha Z_w}{\cos\alpha - y'\sin\alpha}.$$