COMS30030: Image Processing and Computer Vision

Stereo Problem Sheet

Solutions

1. Practical benefits of epipolar lines. We can search along epipolar lines for corresponding points, avoiding searching everywhere in the image. This can speed up the process, reducing the amount of computation. It can also reduce mismatches, as we are restricting where we look, hence potentially removing points that would have been mismatches.

2. (a) Two possible ways of solving this question. First, using the relationship between depth and disparity in the X-Z plane: $Z = fT_x/d$. Disparity $d = 0.2-(-0.2) = 0.4$, hence $Z = 2/0.4 = 5$. Now we can compute the 3-D wrt left camera using perspective projection equation: $X/Z = x/f \rightarrow X = Zx/f = 5 \times 0.2 = 1$ and similarly $Y = Zy/f = 5 \times 0.1 = 0.5$. Giving the 3-D point $\mathbf{P} = (1, 0.5, 5)$.

   Alternatively using summation of vectors in stereo set up, i.e. that scaled image plane vector in left camera must equal T plus scaled version of image plane vector in right camera (Note cameras have the same orientation and so we don't need to correct for different camera coordinate systems). This can be expressed as:

$$a \begin{bmatrix} 0.2 \\ 0.1 \\ 1 \end{bmatrix} = b \begin{bmatrix} -0.2 \\ 0.1 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}$$

   which gives $a = b = 5$ and hence wrt left camera $\mathbf{P} = 5(0.2, 0.1, 1) = (1, 0.5, 5)$.

   (b) Use the epipolar constraint equation: $\mathbf{p}_R^T E \mathbf{p}_L = 0$, noting that $E = SR$, $R = I$ in this simple stereo set up and S is given by

$$S = \begin{bmatrix} 0 & -T_Z & T_Y \\ T_Z & 0 & -T_X \\ -T_Y & T_X & 0 \end{bmatrix}$$

   Hence for $\mathbf{T} = (2, 0, 0)$ we get

$$\begin{bmatrix} x_R & y_R & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -2 \\ 0 & 2 & 0 \end{bmatrix} \begin{bmatrix} x_L \\ y_L \\ 1 \end{bmatrix} = \begin{bmatrix} x_R & y_R & 1 \end{bmatrix} \begin{bmatrix} 0 \\ -2 \\ 2y_L \end{bmatrix} = 0$$

   Giving $-2y_R + 2y_L = 0 \rightarrow y_R = y_L$. Hence all epipolar lines are horizontal since all possible corresponding points in the right image will have the same y coordinate as the point of interest in the left image.

3. For parallel image planes, $R = I$ and hence epipolar constraint becomes $\mathbf{p}_R^T E \mathbf{p}_L = \mathbf{p}_R^T S \mathbf{p}_L = 0$, which becomes

$$\left[\begin{array}{ccc} x_R & y_R & 1 \end{array}\right] \left[\begin{array}{ccc} 0 & -T_Z & T_Y \\ T_Z & 0 & -T_X \\ -T_Y & T_X & 0 \end{array}\right] \left[\begin{array}{c} x_L \\ y_L \\ 1 \end{array}\right] = \left[\begin{array}{ccc} x_R & y_R & 1 \end{array}\right] \left[\begin{array}{c} T_Y - T_Z y_L \\ T_Z x_L - T_X \\ T_X y_L - T_Y x_L \end{array}\right] = 0$$

which gives

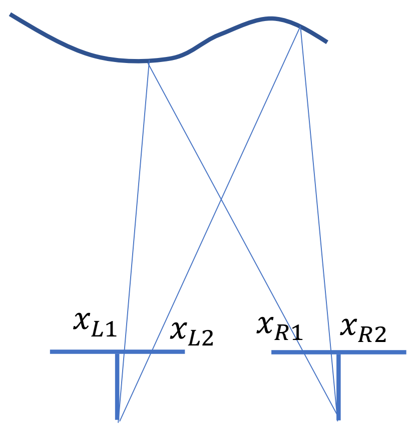
$$x_R(T_Y - T_Z y_L) + y_R(T_Z x_L - T_X) + (T_X y_L - T_Y x_L) = 0$$

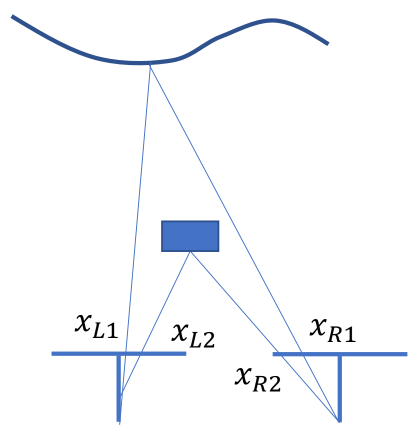
Hence in $y = mx + c$ form

$$y_R = \frac{(T_Y - T_Z y_L)}{(T_X - T_Z x_L)} x_R + \frac{(T_X y_L - T_Y x_L)}{(T_X - T_Z x_L)}$$

These epipolar lines will be parallel when the slope does not depend on x_L and y_L, which we can force by making $T_Z = 0$, i.e. when $\mathbf{T} = (T_X, T_Y, 0)$ and so image planes are coplanar, i.e. lie in the same plane.

4. Order of corresponding points in both views is mainted when depth is similar as shown on l;eft below, whilst oreder is switched when second point corresponds to projection from object in the foregropund as shown on the right.





5. From slide 5 in lecture 2:

$$H_{WL} = \left[\begin{array}{cc} R_L & -R_L T_L \\ \mathbf{0} & 1 \end{array}\right] \quad H_{WR} = \left[\begin{array}{cc} R_R & -R_R T_R \\ \mathbf{0} & 1 \end{array}\right]$$

where R_L is rotation taking left coordinate system back to world coordinate system and T_L defines the position of the left COP in the world coordinate system (and similar for R_R and T_R). Note that

$$H_{WR}^{-1} = \left[\begin{array}{cc} R_R^T & \mathbf{T}_R \\ \mathbf{0} & 1 \end{array}\right]$$

and so

$$H_{RL} = \left[\begin{array}{cc} R_L & -R_L \mathbf{T}_L \\ \mathbf{0} & 1 \end{array}\right] \left[\begin{array}{cc} R_R^T & T_R \\ \mathbf{0} & 1 \end{array}\right] = \left[\begin{array}{cc} R_L R_R^T & R_L \mathbf{T}_R - R_L \mathbf{T}_L \\ \mathbf{0} & 1 \end{array}\right]$$

$$= \begin{bmatrix} (R_R R_L^T)^T & R_L(\mathbf{T}_R - \mathbf{T}_L) \\ \mathbf{0} & 1 \end{bmatrix} = \begin{bmatrix} (R_{RL}^T & \mathbf{T}_{RL} \\ \mathbf{0} & 1 \end{bmatrix}$$

where $R_{RL} = R_R R_L^T$ is the rotation difference between R_R and R_L (recall the R^T is the inverse rotation of R) and hence the rotation that takes the right coordinate system into the left one, and $\mathbf{T}_{RL} = R_L(\mathbf{T}_R - \mathbf{T}_L)$ defines the position of the right COP wrt the left coordinate system, ie $(\mathbf{T}_R - \mathbf{T}_L)$ is in the world coordiante system but then applying R_L converts it into the left coordinate system.

6. Image plane coordinates are related to pixel coordinates by $x = s_x(\hat{x} - \hat{o}_x)$ and $y = s_y(\hat{y} - \hat{o}_y)$. This can be written as a matrix times a vector:

$$\mathbf{p} = M\hat{\mathbf{p}} = \begin{bmatrix} s_x & 0 & -s_x\hat{o}_x/f \\ 0 & s_y & -s_y\hat{o}_y/f \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{x} \\ \hat{y} \\ f \end{bmatrix} = \begin{bmatrix} x \\ y \\ f \end{bmatrix}$$

7. All epipolar lines pass through epipoles and vector $\mathbf{T}$ defining right-hand COP w.r.t left hand COP is in all epipolar planes, hence moving COP along the direction of $\mathbf{T}$ has no impact on the epipolar lines or epipolar planes. This can be shown by considering the epipolar constraint equation:

$$\mathbf{p}_R^T E \mathbf{p}_L = \mathbf{p}_R^T R S \mathbf{p}_L = 0 \Rightarrow \mathbf{p}_R^T R c S \mathbf{p}_L = 0$$

where c is a scalar and in the last step we have multiplied both sides of the equation by c. Hence taking c inside S yields an essential matrix E with a scaled version of T and so the constraint equation remains unchanged and so the epipolar lines and hence planes are the same.

8. Forming the epipolar constraint equation $\mathbf{p}_R^T E \mathbf{p}_L = 0$

$$E\mathbf{p}_L = \begin{bmatrix} 0 & -0.4 & 0 \\ 0 & 0 & -10 \\ 0 & 0.9 & 0 \end{bmatrix} \begin{bmatrix} 0.1 \\ 0.2 \\ 1 \end{bmatrix} = \begin{bmatrix} -0.08 \\ -10 \\ 0.18 \end{bmatrix}$$

Now test each right camera point to see which is closest to satisfying epipolar constraint

(a)

$$\begin{bmatrix} 0.1 & 0.02 & 1 \end{bmatrix} \begin{bmatrix} -0.08 \\ -10 \\ 0.18 \end{bmatrix} = -0.208 + 0.18 = -0.028$$

(b)

$$\begin{bmatrix} 0.2 & 0.01 & 1 \end{bmatrix} \begin{bmatrix} -0.08 \\ -10 \\ 0.18 \end{bmatrix} = -0.116 + 0.18 = 0.064$$

(a)

$$\begin{bmatrix} 0.4 & 0.012 & 1 \end{bmatrix} \begin{bmatrix} -0.08 \\ -10 \\ 0.18 \end{bmatrix} = -0.152 + 0.18 = 0.028$$

Hence (a) and (c) are equally likely correspondences. Note: positive or negative deviation from the epipolar constraint are equivalent in terms of deviation from the constraint.

9. (a) The key part about determining the configuration of the cameras is to note that the rotation matrix R defines a 3-D rotation about the y-axis. For an angle θ such a rotation in the clockwise direction is defined by the following matrix:

$$R = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix}$$

To see that this is a rotation about the y-axis, note that if you multiply a 3-D vector by it then the y component of remains unchanged. See `http://en.wikipedia.org/wiki/Rotation_matrix` for more details on rotation matrices.

In the question $\theta = 60\,\mathrm{deg}$ and hence the configuration of the cameras is as follows:

ie to transform the vector $\mathbf{P}_L$, defined wrt the left camera, into $\mathbf{P}_R$, defined wrt the right camera, we shift by T and then rotate by $60\,\mathrm{deg}$. Note that the right camera is rotated counter-clockwise wrt the left camera, ie the reverse of R in the coordinate transformation equation.

(b) We need to use the essential matrix relationship, ie

$$\mathbf{p}_r^T E \mathbf{p}_l = 0 \qquad\qquad E = RS$$

where $\mathbf{p}_r$ and $\mathbf{p}_l$ are the projections of the 3-D point into each camera in image plane coordinates and

$$S = \begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & -\sqrt{3} \\ 0 & \sqrt{3} & 0 \end{bmatrix}$$

We can compute $E = RS$:

$$E = \frac{1}{2}\begin{bmatrix} 1 & 0 & \sqrt{3} \\ 0 & 2 & 0 \\ -\sqrt{3} & 0 & 1 \end{bmatrix}\begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & -\sqrt{3} \\ 0 & \sqrt{3} & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & -\sqrt{3} \\ 0 & \sqrt{3} & 0 \end{bmatrix}$$

We are given $\mathbf{p}_l = (0.02, 0.01, 0.1)$ (NB $\mathbf{p}_l$ lies in the left camera image plane and hence it's z component equals the focal length) and so we can use the above epipolar relationship to establish a constraint on the position of the projection in the right camera, ie

$$E\mathbf{p}_l = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & -\sqrt{3} \\ 0 & \sqrt{3} & 0 \end{bmatrix}\begin{bmatrix} 0.02 \\ 0.01 \\ 0.1 \end{bmatrix} = \begin{bmatrix} 0.01 \\ 0.02 - \sqrt{3}/10 \\ \sqrt{3}/100 \end{bmatrix}$$

$$\mathbf{p}_r^T E \mathbf{p}_l = \begin{bmatrix} x_r & y_r & 0.1 \end{bmatrix}\begin{bmatrix} 0.01 \\ 0.02 - \sqrt{3}/10 \\ \sqrt{3}/100 \end{bmatrix} = \frac{x_r}{100} + \frac{1 - 5\sqrt{3}}{50} y_r + \frac{\sqrt{3}}{1000} = 0$$

Hence we get the equation of the epipolar line in the right image upon which $\mathbf{p}_r$ must lie, ie

$$y_r = \frac{1}{10\sqrt{3}-2}[x_r + \sqrt{3}/10]$$

If we insert $x_r = -0.02$ then we get from above $y_r = 0.01$, hence confirming that the point $(-0.02, 0.01)$ does indeed lie on the epipolar line.

(c) For 3-D reconstruction we need the following equation

$$a\mathbf{p}_l - bR^T\mathbf{p}_r - c(\mathbf{p}_l \times R^T\mathbf{p}_r) = \mathbf{T}$$

where $\times$ denotes the dot product and a, b and c are to be determined. Note that

$$\mathbf{p}_l \times R^T\mathbf{p}_r = MR^T\mathbf{p}_r$$

where M is given by

$$M = \begin{bmatrix} 0 & -z_l & y_l \\ z_l & 0 & -x_l \\ -y_l & x_l & 0 \end{bmatrix} == \begin{bmatrix} 0 & -0.1 & 0.01 \\ 0.1 & 0 & -0.02 \\ -0.01 & 0.02 & 0 \end{bmatrix}$$

We can solve for a, b and c by rearranging:

$$[\mathbf{p}_l \qquad - R^T\mathbf{p}_r \qquad - MR^T\mathbf{p}_r] \begin{bmatrix} a \\ b \\ c \end{bmatrix} = Q \begin{bmatrix} a \\ b \\ c \end{bmatrix} = T$$

which gives

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = Q^{-1}\mathbf{T} = \begin{bmatrix} 3.2232 & 0.8378 & 9.2716 \\ 9.641 & -0.8378 & -1.8444 \\ 6.2225 & 95.3324 & -10.7778 \end{bmatrix} \begin{bmatrix} \sqrt{3} \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 14.8543 \\ 14.8543 \\ 0 \end{bmatrix}$$

The parameter $c = 0$ indicates that the rays from projecting out from each camera through the corresponding points cross and the 3-D coordinates of the point wrt the left camera are given by $a\mathbf{p}_l = (0.2971, 0.1485, 1.4854)$.

(d) For the noisy corresponding points using the above equations gives $a = 14.8995$, $b = 14.2152$ and $c = 0.6526$. Hence the rays no longer cross since $c \neq 0$ as the points are no longer in perfect correspondence.

The estimate of the 3-D point can then be found finding the midpoint between $a\mathbf{p}_l = (0.3725, 0.1639, 1.49)$ and $bR^T\mathbf{p}_r + \mathbf{T} = (0.3730, 0.1706, 1.4892)$, ie find the average of the two vectors to give $(0.3728, 0.1672, 1.4896)$.

Andrew Calway
November 14, 2025