IIST, Thiruvananthapuram	AVD624 - Computer Vision

Lecture Notes: Week - 11

01/04/2022 - 08/04/2022

Prof. Dr. Deepak Mishra — *Summarised by: Vatsalya Gupta*

This is a weekly summary of the lectures given by *Dr. Deepak Mishra*, Professor and Head of Department of Avionics, for the course on *Computer Vision* conducted during the even semester 2021-22.

1 Stereo / Multiview Geometry

Multiview stereo refers to the task of reconstructing a 3D shape from calibrated overlapping images captured from different viewpoints. Various representations of 3D shape can be used. For example, dense 3D point cloud or surface mesh representations are common in applications that synthesize a new photorealistic image of the scene using computer graphic rendering techniques. The topics of multiview stereo and multi-baseline stereo matching share key concepts related to the recovery of dense 2D pixel correspondences in multiple images.

Disparity occurs when eyes fixate on one object; others appear at different visual angles.

2 Triangulation

Create two points on the ray: find the camera center; and apply the pseudo-inverse of P on x. Then connect the two points. This procedure is called **backprojection**.

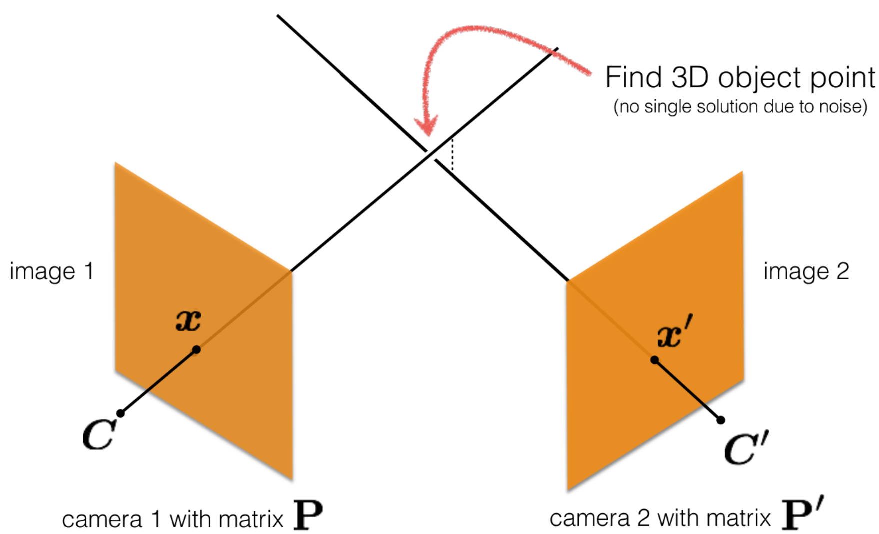

Figure 1: Stereo camera setup.

The triangulation problem is to estimate the 3D point $\mathbf{X}$, given a set of (noisy) matched points $\{\mathbf{x_i}, \mathbf{x_i'}\}$ and camera matrices $\mathbf{P}$, $\mathbf{P}'$. We can use the similarity relation $\mathbf{x} = \alpha\mathbf{PX}$, same direction but differs by a scale factor. So using cross-product rule $\mathbf{x} \times \mathbf{PX} = \mathbf{0}$. On expanding out the camera matrix and points, we get

$$\begin{bmatrix} yp_3^T X - p_2^T X \\ p_1^T X - xp_3^T X \\ xp_2^T X - p_1^T X \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

It can be seen that the third line is a linear combination of the first and second lines. Hence, one 2D to 3D point correspondence give you 2 equations. We can concatenate the 2D points from both images to get,

$$\begin{bmatrix} yp_3^T - p_2^T \\ p_1^T - xp_3^T \\ y'p_3'^T - p_2'^T \\ p_1'^T - x'p_3'^T \end{bmatrix} X = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

We need to solve the homogeneous linear system $\mathbf{AX} = \mathbf{0}$.

3 Stereo Camera System

Using the idea of perspective projection for two cameras - left (subscript *l*) and right (subscript *r*). We get,

$$(u_l, v_l) = (f_x\frac{x}{z} + o_x, f_y\frac{y}{z} + o_y) \qquad (u_r, v_r) = (f_x\frac{x-b}{z} + o_x, f_y\frac{y}{z} + o_y)$$

$$\implies z = \frac{bf_x}{u_l - u_r} \implies z \propto \frac{1}{u_l - u_r}$$

$(u_l - u_r)$ is called **Disparity**, and is proportional to **Baseline**.

4 Orientation Estimation

It involves assessing various problems such as epipolar geometry, estimating fundamental matrix, finding dense correspondence and computing depth.

4.1 Computing Extrinsic Parameters

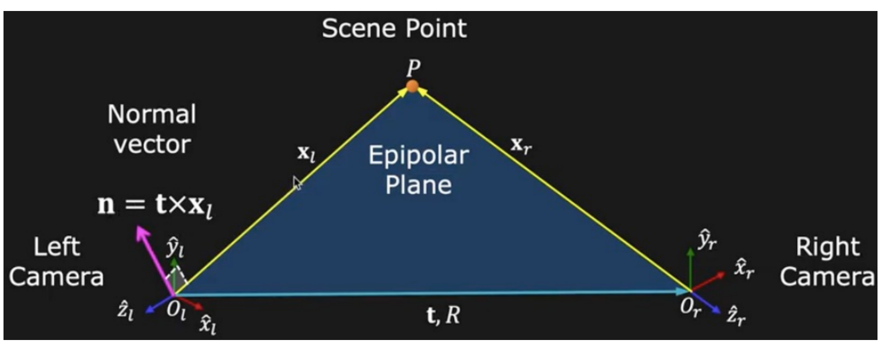

Figure 2: Epipolar geometry.

We write the epipolar constrain in matrix form as $\mathbf{x_l}.(\mathbf{t} \times \mathbf{x_l}) = \mathbf{0} \implies \begin{bmatrix} x_l \\ y_l \\ z_l \end{bmatrix} \begin{bmatrix} t_y z_l - t_z y_l \\ t_z x_l - t_x z_l \\ t_x y_l - t_y x_l \end{bmatrix} = 0$. Similarly, extending the idea for the right camera, we get $\mathbf{x_l} = \mathbf{R}\mathbf{x_r} + \mathbf{t}$, where $\mathbf{t_{3\times 1}}$ is the position of right camera in left's frame and $\mathbf{R_{3\times 3}}$ is the orientation of left camera in right's frame. Substituting into the epipolar constraint will give us the essential matrix $\mathbf{E} = \mathbf{T_x}\mathbf{R}$. Since, $\mathbf{T_x}$ is skew-symmetric and orthornormal, we can apply Singular Value Decomposition (SVD) on $\mathbf{T_x}$ and $\mathbf{R}$. We get,

$$\mathbf{x_l^T}\mathbf{E}\mathbf{x_r} = 0$$

Expand and write in terms of image coordinates and known camera matrix $\mathbf{K}$.

$$\begin{bmatrix} u_l & v_l & 1 \end{bmatrix} z_l \mathbf{K_l^{-1^T}} \mathbf{E} \mathbf{K_r^{-1}} z_r \begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} = 0 \implies \mathbf{E} = \mathbf{K_l^T}\mathbf{F}\mathbf{K_r}$$

For calibrated cameras, 12 parameters are needed for angle preserving mapping and for uncalibrated cameras, 22 parameters are needed for straight-line preserving mapping. First case can be resolved by using 3(4) known control points and computing via two separate spatial resection/P3P steps. The second case can be computed via two step DLT process using 6 known control points.

4.2 Coplanarity Constraint

Coplanarity can be expressed by $\begin{bmatrix} O'X' & O'O'' & O"X" \end{bmatrix} = 0$. The normalized directions of the vectors O"X" & O'X' are

$${}^{n}x' = (R'^{-1})(K')^{-1}(x')$$

As the normalised projection,

$${}^{n}x' = \begin{bmatrix} I_3 | -X_{o'} \end{bmatrix} \mathbf{X}$$

Upon solving this using the idea of cross product, we get the following expressions.

$$x'^T F x" = 0 \implies F = (K')^{-T} R' S_b R"^T (K")^{-1} \implies F = A'^{-T} S_{b'_{12}} A"^{-1}$$

Here, F is the fundamental matrix and S_b is a skew-symmetric matrix satisfying ${}^{n}x'^T S_b {}^{n}x" = 0$. Also, projection matrices are defined as $P' = [A'|a'] = [K'R'| - K'R'X'_O]$, for $b'_{12} = A'^{-1}a' - A"^{-1}a"$.

■■■