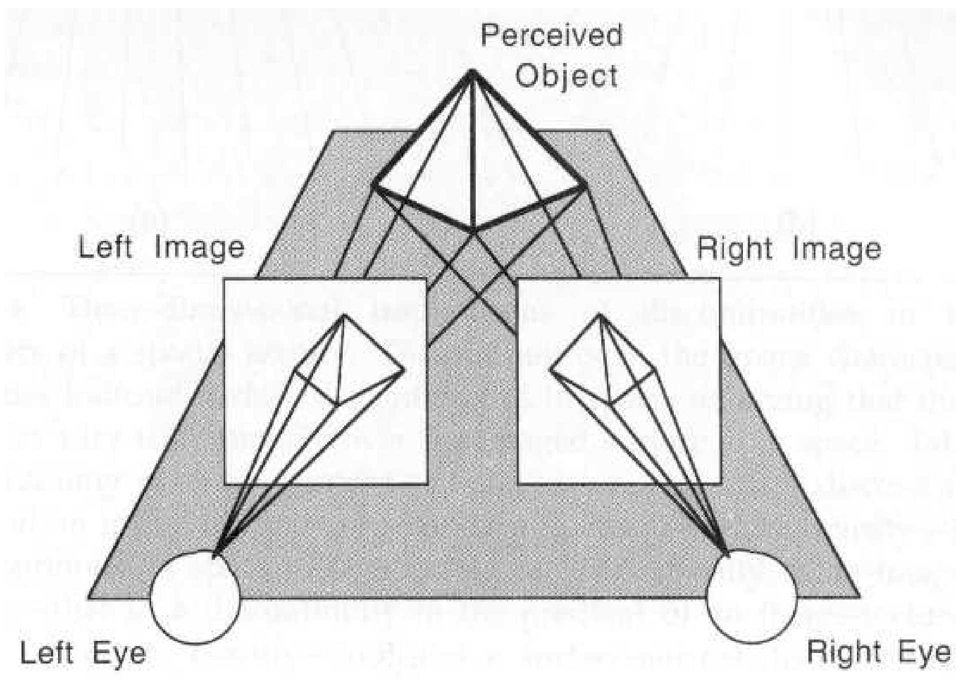
Stereo Vision

• What is the goal stereo vision?

- The recovery of the 3D structure of a scene using two or more images of the 3D scene, each acquired from a different viewpoint in space.

- The images can be obtained using muliple cameras or one moving camera.

- The term binocular vision is used when two cameras are employed.



• Stereo setup and terminology

Fixation point: the point of intersection of the optical axis.

Baseline: the distance between the centers of projection.

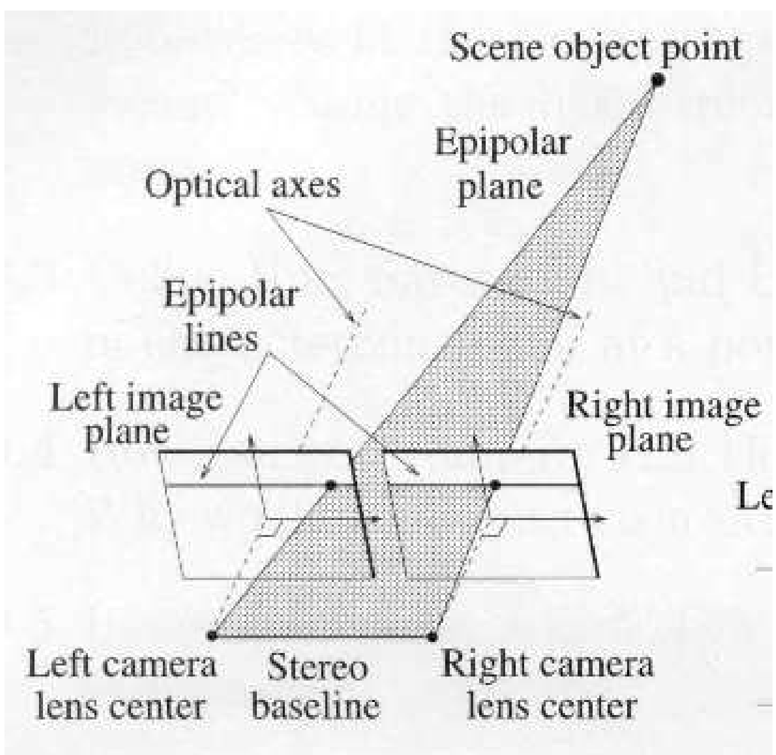
Epipolar plane: the plane passing through the centers of projection and the point in the scene.

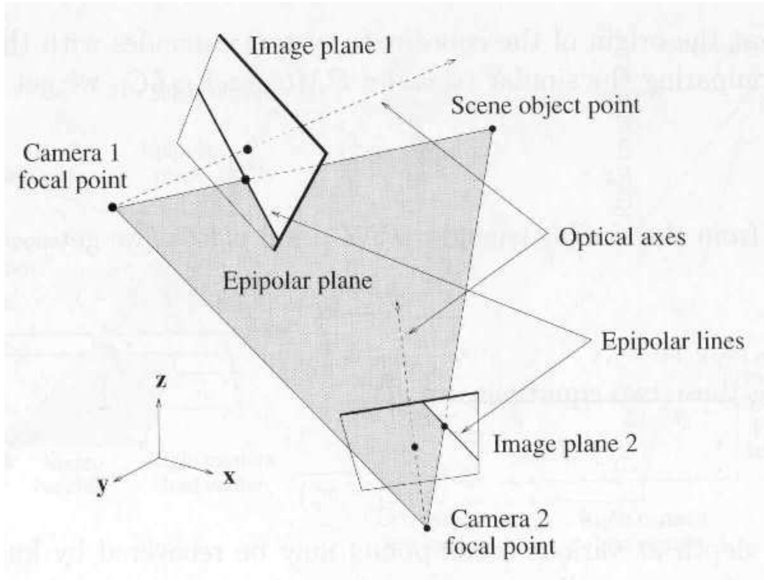
Epipolar line: the intersection of the epipolar plane with the image plane.

Conjugate pair: any point in the scene that is visible in both cameras will be projected to a pair of image points in the two images.

Disparity: the distance between corresponding points when the two images are superimposed.

Disparity map: the disparities of all points form the disparity map (can be displayed as an image).



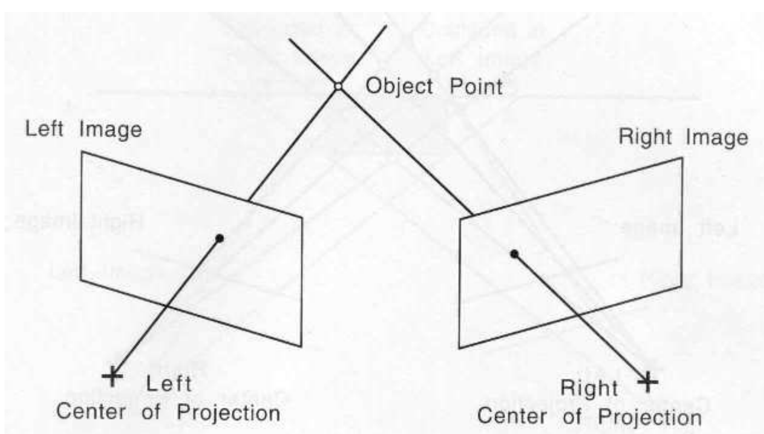


(two cameras in arbitrary position and orientation)

• Triangulation - the principle underlying stereo vision

- The 3D location of any visible object point in space is restricted to the straight line that passes through the center of projection and the projection of the object point.

- Binocular stereo vision determines the position of a point in space by finding the intersection of the two lines passing through the center of projection and the projection of the point in each image.



- **The two problems of stereo**

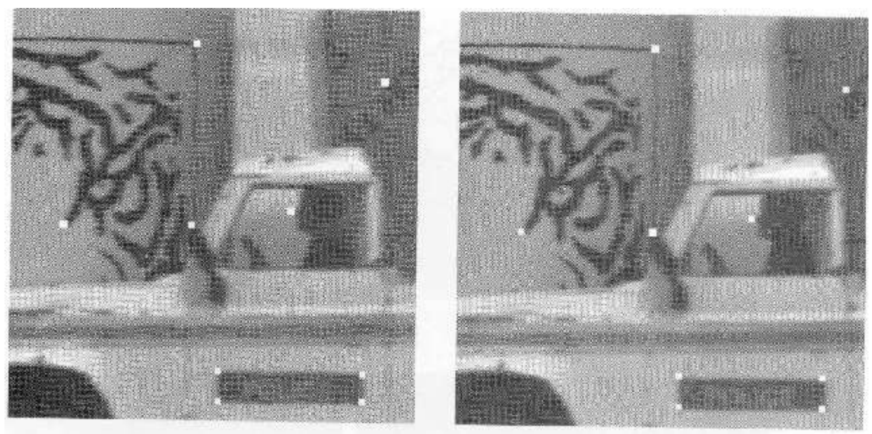
  (1) The *correspondence* problem.

  (2) The *reconstruction* problem.

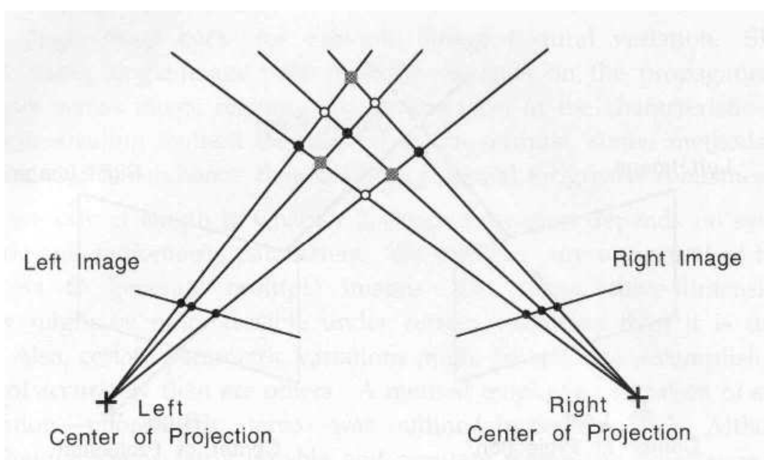
- **The correspondence problem**

- Finding pairs of matched points such that each point in the pair is the projection of the same 3D point.

- Triangulation depends crucially on the solution of the correspondence problem.



- Ambiguous correspondence between points in the two images may lead to several different consistent interpretations of the scene.



• The reconstruction problem

- Given the corresponding points, we can compute the disparity map.

- The disparity map can be converted to a 3D map of the scene (i.e., recover the 3D structure) if the stereo geometry is known.

• Recovering depth (reconstruction)

- Consider recovering the position of P from its projections p_l and p_r.

$$x_l = f\frac{X_l}{Z_l} \text{ or } X_l = \frac{x_l Z_l}{f} \text{ and } x_r = f\frac{X_r}{Z_r} \text{ or } X_r = \frac{x_r Z_r}{f}$$

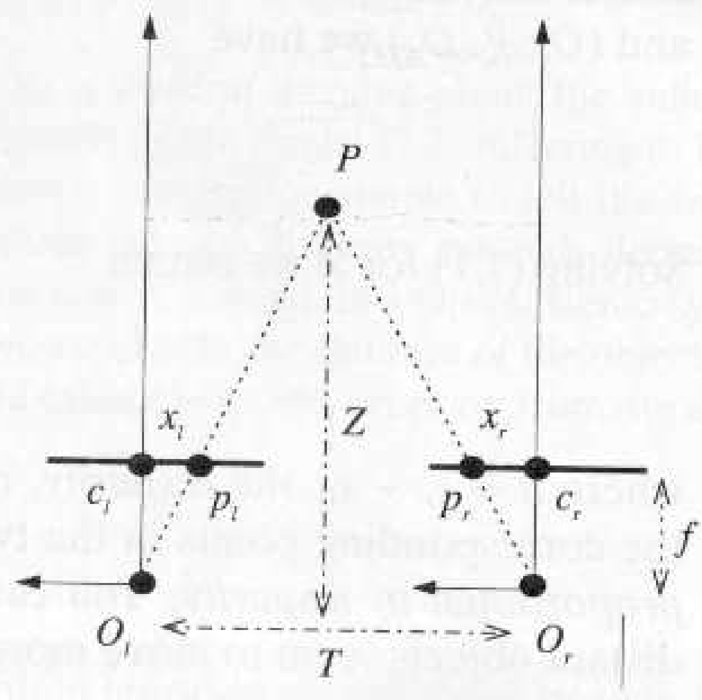
- In general, the two cameras are related by the following transformation:

$$P_r = R(P_l - T)$$

- Using $Z_r = Z_l = Z$ and $X_r = X_l - T$ we have:

$$\frac{x_l Z}{f} - T = \frac{x_r Z}{f} \quad \text{or} \quad Z = \frac{Tf}{d}$$

where $d = x_l - x_r$ is the *disparity* (i.e., the difference in the position between the corresponding points in the two images)



(b)

• Stereo camera parameters

Intrinsic parameters: characterize the transformation from image plane coordinates to pixel coordinates, in each camera.

Extrinsic parameters (R, T): describe the relative position and orientation of the two cameras

$$P_r = R(P_l - T) \text{ (aligns right camera with left camera)}$$

- Can be determined from the extrinsic parameters of each camera:

$$R = R_r R_l^T$$

$$T = T_l - R^T T_r$$