Stereopsis

What is stereopsis?
Stereo vision refers to the ability to infer information on the 3D structure and distance of a scene from two or more images taken from different viewpoints.
We shall be focusing more specifically on binocular vision.
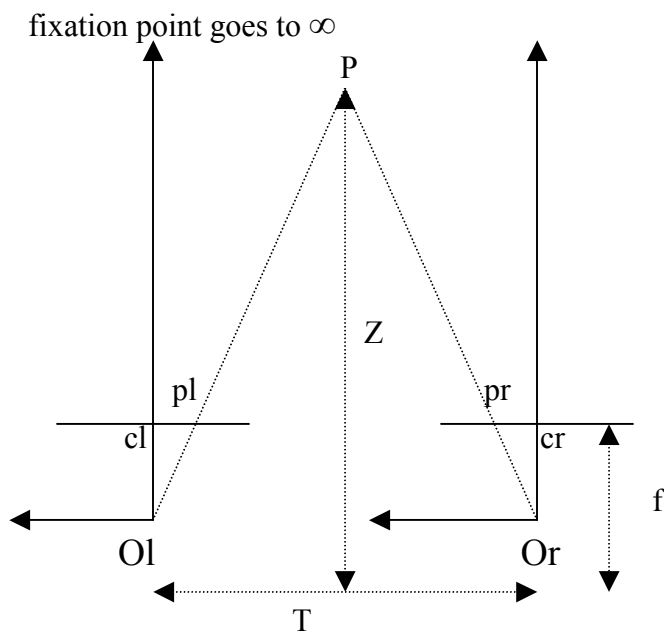
The Two Problems Of Stereo:
- Correspondence:
    Determining which item in the left eye corresponds to which item in the right eye.
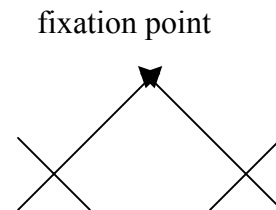- Reconstruction:
    Conversion into a 3D map of the scene based on our knowledge of the *geometry of the stereo system* and on the *disparity map*. *Disparity* is the computed difference between corresponding objects.

Simple Stereo System:

T is the **baseline.**
Ol and Or are the **optical centers**.
Z is the distance between P and the baseline.
f is the **focal length**.

Properties:

From the similar triangles (pl, P, Pr) and (Ol, P, Or), we have

$$\frac{T + xl - xr}{Z - f} = \frac{T}{Z}$$

xl and xr are the coordinates of pl and pr respectively.

we then obtain

$$Z = f \frac{T}{d}$$

$d = xr - xl$ is the *disparity*

⇨ Depth Z is inversely proportional to disparity.
⇨ Note that **disparity** is the sum of the displacements of pl and pr from their origin, i.e. $|xl| + |xr|$, since xl < 0 we have xr − xl.
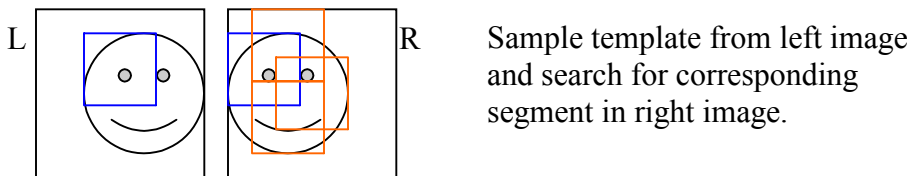
Parameters Of A Stereo System:
• Intrinsic parameters characterize the transformation mapping an image point from camera to pixel coordinates, in each camera
• Extrinsic parameters describe the relative position of the two cameras.

So far, we have **f**, **T**, **cl** and **cr** as the parameters of our system. Finding these values is part of **stereo calibration**. We shall see that it is possible to compute much of the 3D information without any prior knowledge of these parameters, this is know as **uncalibrated stereo**. For this, we need to understand **epipolar geometry**.

The Correspondence Problem:

Idea: Find which parts of the left and right images correspond.
Technique: The basic idea would be to take an element from say the left image and finding the corresponding element in the right image.

Sample template from left image and search for corresponding segment in right image.

Problem: Which element should we choose?
        How should the search be carried out?
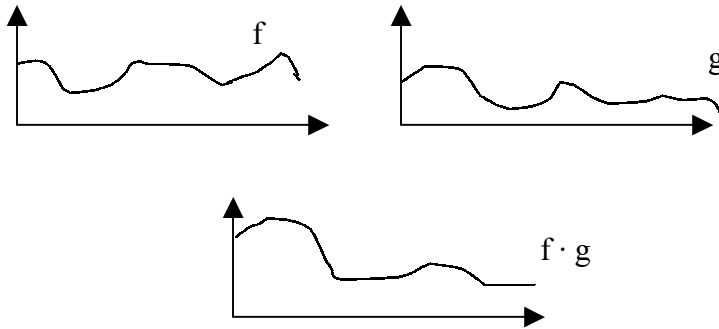Methods: Two general methods will be described below:

Correlation-Based Methods
1. Notes about these methods:
   - Deal with raw data, simple to work with.
   - It is area based and is dense.
   - There are no features involved.
   - There is a need for textures; homogeneous surfaces may yield incorrect results.

2. **Cross-Correlation**
   From 1D point of view with two functions f and g:

Formula: $$\int f \cdot g \, dx$$

If we generalize this to the case of two 2D arrays of pixels we have:

$$\Psi(I_L, I_R) = \Sigma \, \Sigma \, I_L \cdot I_R$$

Our objective here is to **maximize** the above formula.
If the two functions are similar, then the integral will be maximal whereas if the functions are dissimilar, $f \cdot g$ will be minimal.

3. **Sum of Squared Differences (SSD)**
   This method is similar to the previous except for the formula, which is described as follows:

$$\int (f - g)^2 \, dx$$

Generalizing to images yields:

$$\Psi(I_L, I_R) = \Sigma \, \Sigma \, (I_L - I_R)^2$$

Our objective here is to **minimize** this sum.
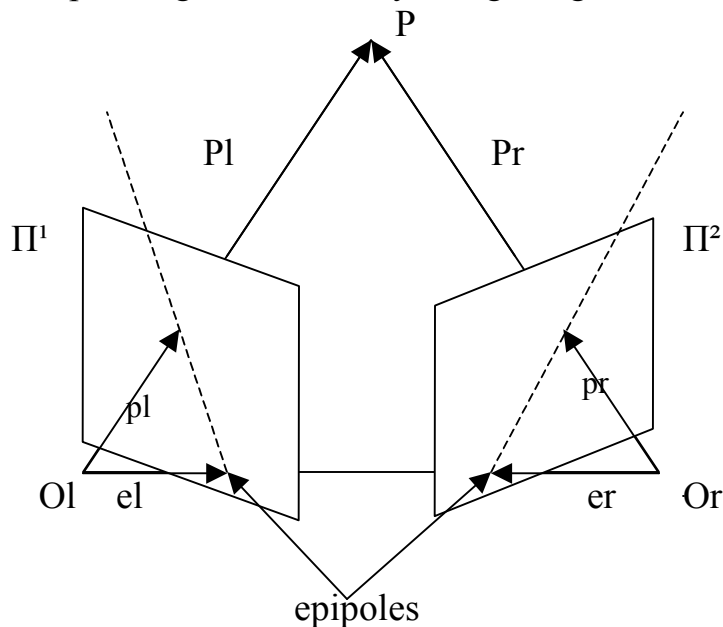
Feature-Based Methods

Notes on this method:
- Abstraction of data.
- Focus on prominent features such as corners.
- Based on local properties.
- Sparse.

Most methods narrow down the number of possible matches for each feature by enforcing certain constraints on feasible matches. They can be *geometric* constraints like the *epipolar constraint,* or ***analytical*** like the *uniqueness* or *continuity constraints*.

Generally, we simply choose a prominent feature like *edges, lines* or *corners*. Then determine the feature descriptor for a line for example, which would contain information such as length, orientation, coordinates or average contrast along the line. Then find the corresponding feature in the other image using some similarity verification scheme.

Epipolar Geometry:

Epipolar geometry allows us to clarify what information is needed in order to perform the search for corresponding elements only along image lines.

The triangle lies in the **epipolar plane**.

The lower corners of the triangle (Ol & Or) are the **optical centers**.

The intersections of the triangle's base with the planes are the **epipoles**.

> An epipole represents the image of the center of the opposite camera on the image plane.

The dotted lines are the **epipolar lines**.

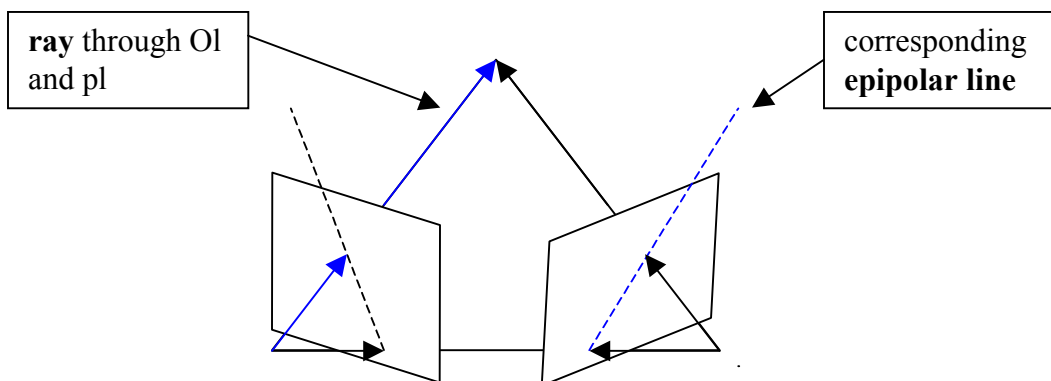> Represents the image of the opposite line on the image plane.

Basics:
- **Extrinsic parameters** are the translation vector **T = (Or – Ol)** and the rotation matrix **R**. Remember the extrinsic parameters relate the reference frames of the right and left camera.
- Relation between Pr and Pl is defined by **Pr = R (Pl – T)   (Ξ)**
- Relation between a point and its projection (using perspective projection):
  - **pl = Pl · (fl / Zl)   (ΩL)**
  - **pr = Pr · (fr / Zr) (ΩR)**

The essence of Epipolar Geometry: The Epipolar Constraint
The practical importance of epipolar geometry is based on the fact that the epipolar plane intersects each image in a line called the epipolar line. So, given pl, we know P can lie anywhere on the ray through Ol and pl. But, since the image of this ray in the right image is the epipolar line going through the corresponding point pr, *the correct match must lie on the epipolar line.* This is known as the *epipolar constraint*.
It can be summarized as follows:
The corresponding point Pr for a point Pl must lie on the epipolar line.

⇨This allows us to restrict the search for the match of pl along the corresponding epipolar line. The correspondence problem has therefore been reduced to a 1D problem.

The Essential Matrix E

<u>Idea:</u> The essential matrix establishes a natural link between the epipolar constraint and the extrinsic parameters of the stereo system.

<u>Getting E:</u>

- **Pl**, **T** and **(Pl – T)** are coplanar vectors found in the epipolar plane.

- Therefore **T x Pl** will yield a vector perpendicular to the plane.

- Hence, $(\mathbf{Pl} - \mathbf{T})^T \cdot \mathbf{Pl} \times \mathbf{T} = \mathbf{0}$ (Dot product of perpendicular vectors = 0)

- From **(Ξ)** the above becomes $(\mathbf{R}^T \mathbf{Pr})^T \mathbf{T} \times \mathbf{Pl} = \mathbf{0}$ **(Λ)**

- Define **T x Pl** = S**Pl** where S is the rank deficient matrix (rank = 2):

$$S = \begin{pmatrix} 0 & -Tz & Ty \\ Tz & 0 & -Tx \\ -Ty & Tx & 0 \end{pmatrix}$$

- Plug it into **(Λ)**: $\mathbf{Pr}^T \mathbf{E} \mathbf{Pl} = \mathbf{0}$ where $\mathbf{E} = \mathbf{R}S$

- Replace Pr and Pl using **(ΩR)** & **(ΩL)** and divide result by **Zl·Zr**:

$$\boxed{\mathbf{pr}^T \ \mathbf{E} \ \mathbf{pl} = \mathbf{0}} \qquad (\Phi)$$

- In the equation above, **E pl** can be seen as the projective line in the right plane that goes through **pr** and the epipole **er**. Let **ur** be this projective line:

$$\boxed{\mathbf{u_r} = \mathbf{E} \ \mathbf{p_L}}$$

⇨ The essential matrix **E** is the mapping between points and epipolar lines we were looking for.

The Fundamental Matrix F:

<u>Idea:</u> We will now show that the mapping between points and epipolar lines can be obtained from corresponding points only. This with no prior information on the stereo system.

<u>Reminder:</u> Intrinsic Parameters are the parameters necessary to link the pixel coordinates of an image point with the corresponding coordinates in the camera reference frame.

<u>Getting F:</u>

- Let **Ml** and **Mr** be the matrices of the **intrinsic parameters** of the left and right camera respectively.
- Let **pl** and **pr** be the points in pixel coordinates corresponding to **pl** and **pr** in camera coordinates:

$\Rightarrow$ **pl** = **Ml**$^{-1}$ **pl**   and **pr** = **Mr**$^{-1}$ **pr**

Plug the two equations above into **(Φ)**: $\Rightarrow$ $\boxed{\underline{\mathbf{pr}}^{\,T}\mathbf{F}\,\underline{\mathbf{pl}} = 0}$

Where $\boxed{\mathbf{F} = \mathbf{Mr}^{-T}\mathbf{E}\ \mathbf{Ml}^{-1}}$ **F** is the *fundamental matrix.*

- **F pl** can be thought of as the equation of the projective epipolar line that corresponds to the point **pl**. Let **ur** be this epipolar line.

$$\boxed{\underline{\mathbf{ur}} = \mathbf{F}\,\underline{\mathbf{pl}}}$$

<u>E & F:</u> The difference
The important difference between E and F is that:
- The fundamental matrix is defined in terms of pixel coordinates,
- The essential matrix in terms of camera coordinates.

$\Rightarrow$ Consequently, if you can estimate F from a number of point matches in pixel coordinates; *you can reconstruct the epipolar geometry with no information on the intrinsic and extrinsic parameters*

<u>Computing E and F:</u>
1. Establish n point correspondences between the two images.

2. Each correspondence gives a homogeneous linear equation like:

$\underline{\mathbf{pr}}^{\,T}\mathbf{F}\,\underline{\mathbf{pl}} = 0$

3. From these equations, setup an nxn matrix A of coefficients.

4. Compute SVD(A) = U DV^T (Notes on SVD are available on the last page)

5. At this point, the entries of F are the components of the column of V corresponding to the last singular value of A.

6. To enforce singularity constraint, compute SVD(F) = UDV^T

7. Set smallest singular value of D to 0 to obtain D'.

8. Corrected estimate of F is

$\boxed{ \atop {}^{T}}$

F' = U D' V

<u>3D Reconstruction:</u>

Cases:

- Know Intrinsic and Extrinsic parameters: Unambiguous and Absolute
- Know Intrinsic parameters only: Euclidean reconstruction, i.e. not absolute depth but relative depth.
- Know neither: Projective reconstruction.

<u>Notes on SVD – Singular Value Decomposition</u>

The idea behind SVD is the fact that any mxn matrix A can be rewritten in a special form:

$$\mathbf{A_{mxn} = U_{mxm} \ D_{mxn} \ V_{nxn}}^{T}$$

Where:

- **D** is a diagonal matrix, with entries $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$
- **U** is an mxn matrix of mutually orthogonal unit vectors as its column.
- **V** is an nxn matrix of mutually orthogonal unit vectors as its column.