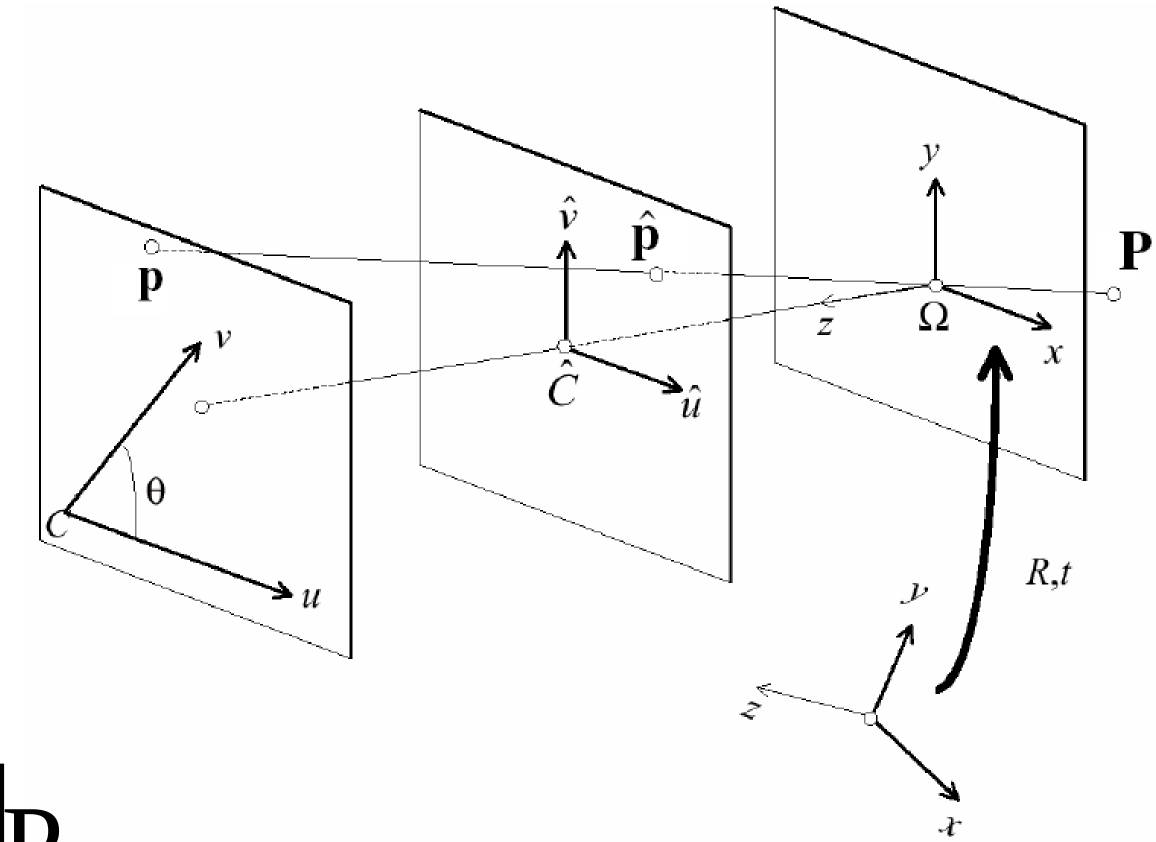
MAPI – Computer Vision

Multiple View Geometry

Geometry of Multiple Views
2- and 3- view geometry

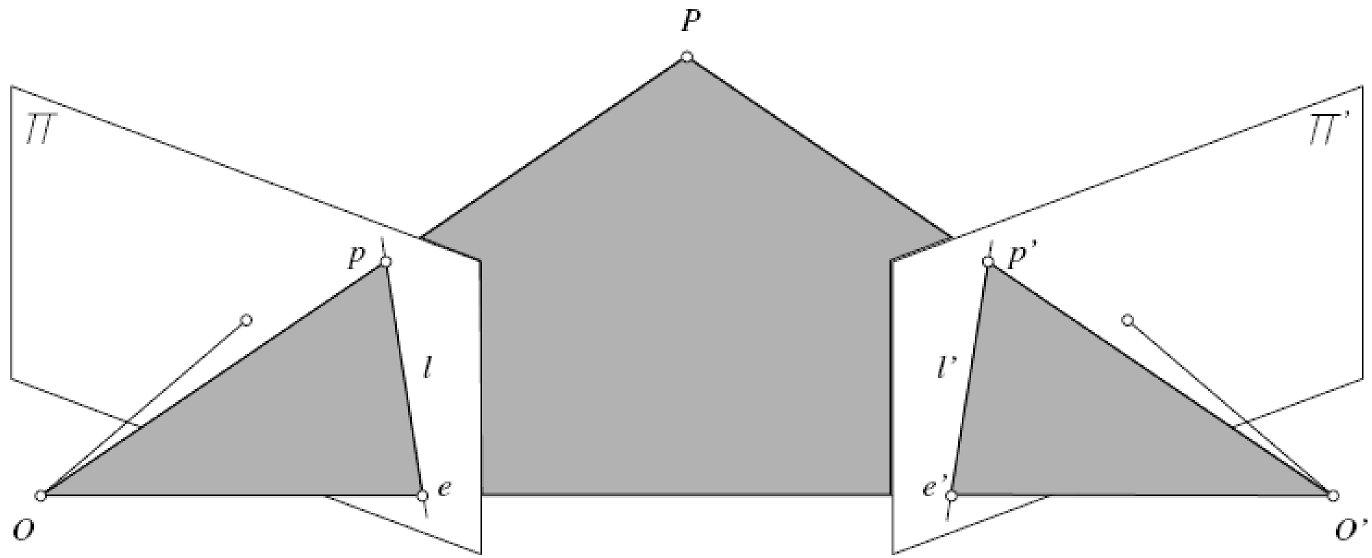


$$p \equiv K\hat{p}$$

$$p \equiv K\left[R \middle| t\right]P$$

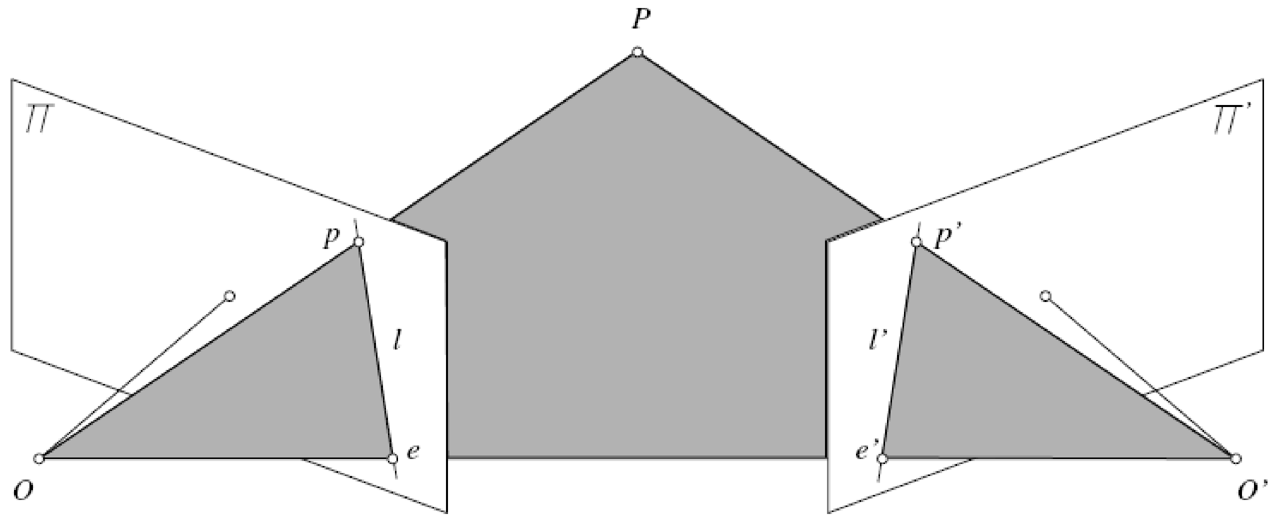
Geometry of Multiple Views
2- and 3- view geometry

- Epipolar Geometry
  - The epipolar geometry is the intrinsic projective geometry between two views.
  - It is independent of scene structure, and only depends on the cameras' internal parameters and relative pose.
  - Consider the images *p* and *p'* of a point *P* observed by two cameras with optical centers *O* and *O'*.



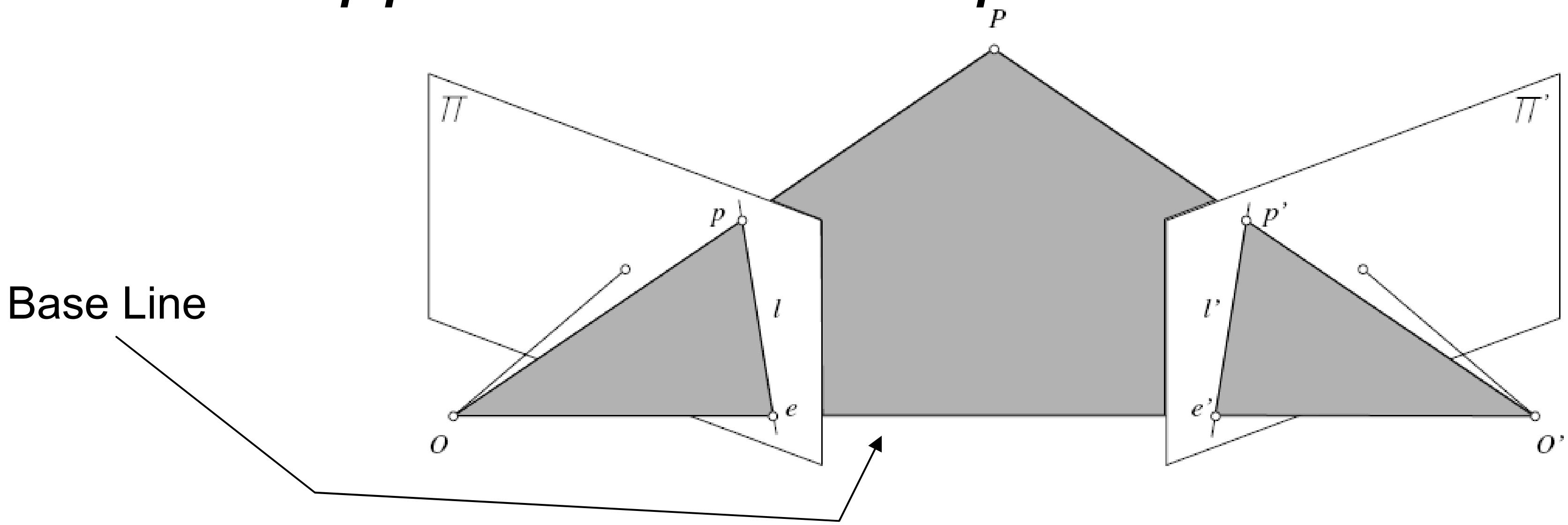
Geometry of Multiple Views
2- and 3- view geometry

- Epipolar Geometry
  - These **five points** all belong to the ***epipolar plane*** defined by the two intersecting rays ***OP*** and ***O'P***.
  - In particular, the point ***p'*** lies on the line ***l'*** where this plane and the retina ***Π'*** of the second camera intersect.
  - The **line *l'*** is the ***epipolar line*** associated with the point ***p',*** and it passes through the point ***e'*** where the baseline joining the optical centers ***O*** and ***O'*** intersects ***Π***.
  - Likewise, the point ***p*** lies on the ***epipolar line l*** associated with the point ***p***, and this line passes through the intersection ***e*** of the baseline with the plane ***Π***.



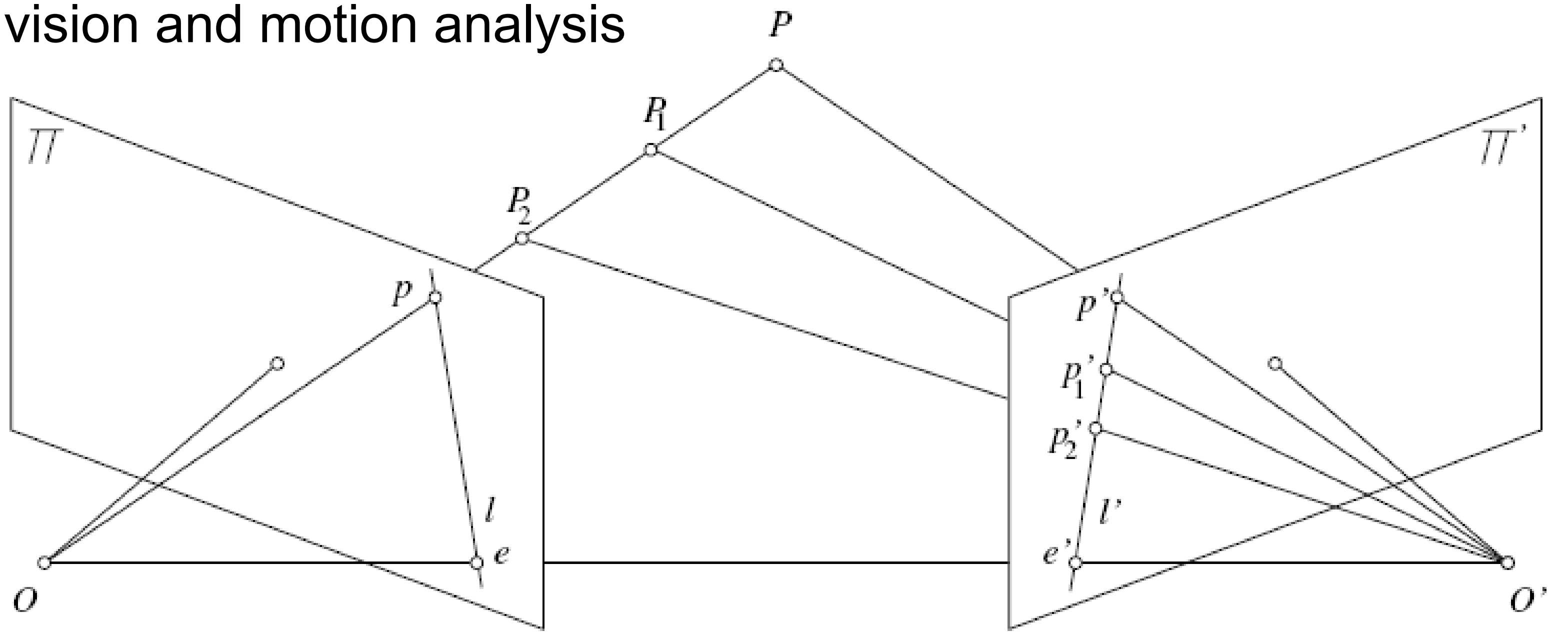
Geometry of Multiple Views
2- and 3- view geometry

- Epipolar Geometry
  - The points ***e*** and ***e'*** are called the ***epipoles*** of the two cameras.
  - The epipole ***e'*** is the ***(virtual) image*** of the ***optical center O*** of the first camera in the image observed by the second camera, and vice versa.
  - if ***p*** and ***p'*** are images of the same point, then ***p'*** must lie on the ***epipolar line*** associated with ***p***.



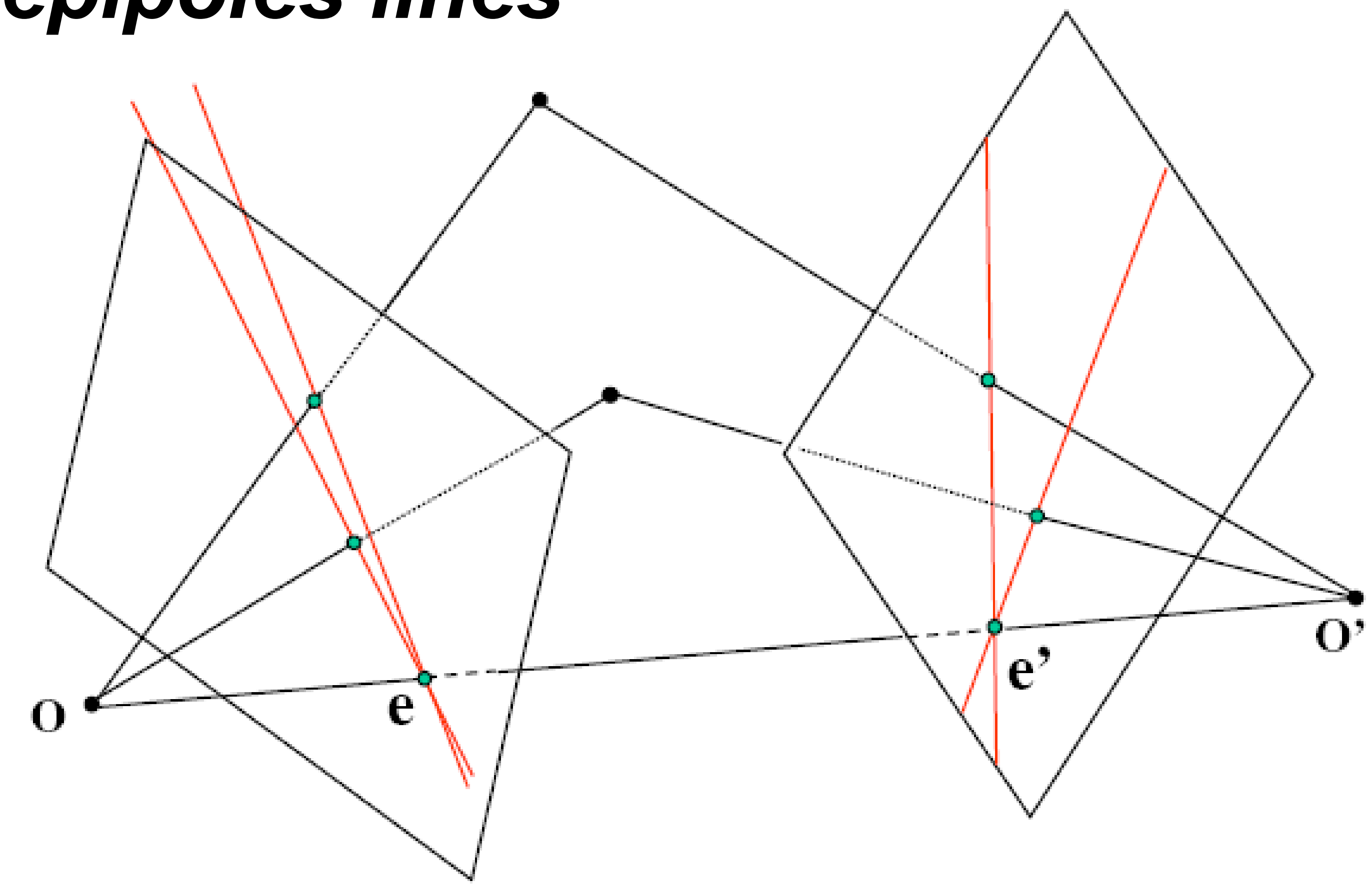
Geometry of Multiple Views
2- and 3- view geometry

- 2- Camera Geometry
  - Given the projection of ***P*** in one image, its projection in the other image is ***constrained to be on a line***, ***epipolar line*** associated with ***p***
  - This ***epipolar constraint*** plays a fundamental role in stereo vision and motion analysis



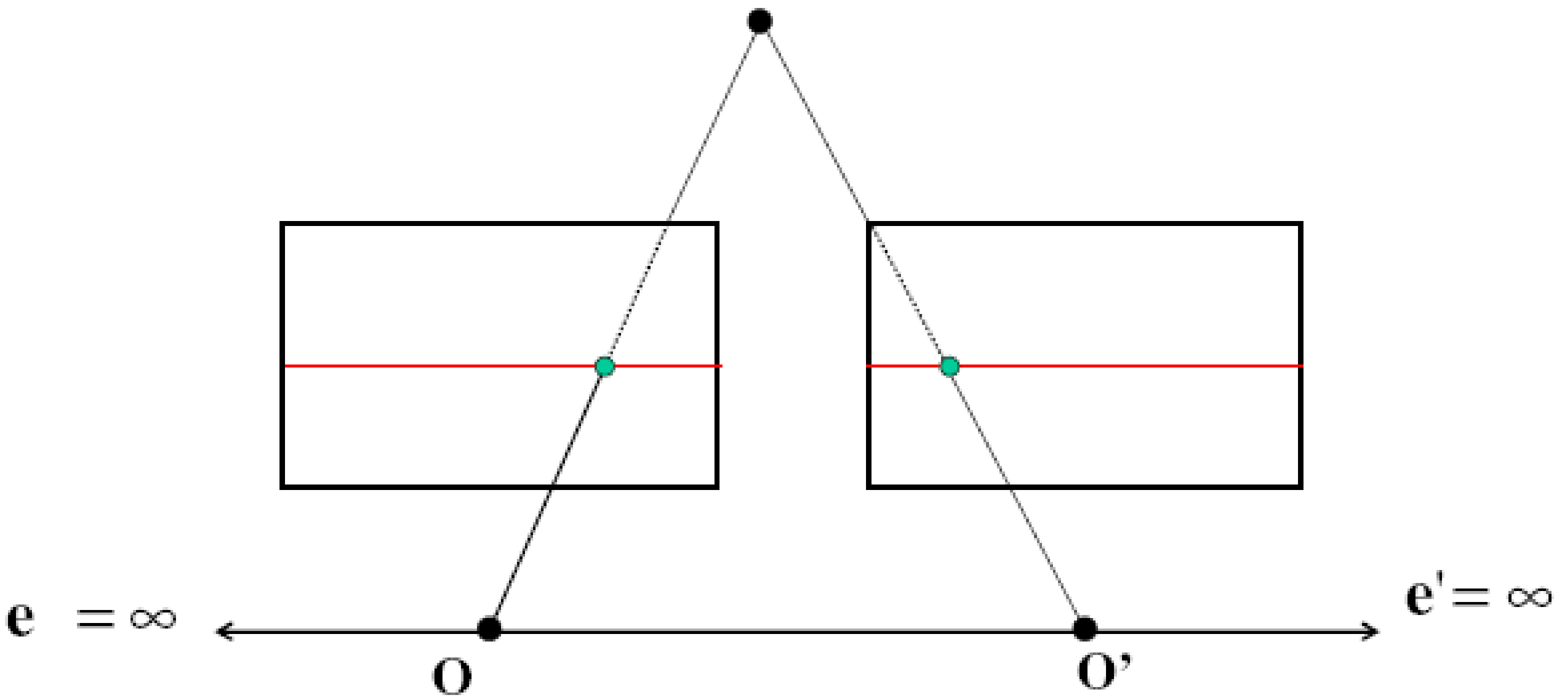
Geometry of Multiple Views
2- and 3- view geometry

- ***Epipoles e*** and ***e'*** are at the intersection of the ***epipoles lines***



Geometry of Multiple Views
2- and 3- view geometry

- Special case: frontoparallel cameras.
  - Epipoles are at infinity



Geometry of Multiple Views
2- and 3- view geometry

- Fundamental matrix F
  - The fundamental matrix is the algebraic representation of epipolar geometry.
  - **Geometric derivation**
    - The mapping from a point in one image to a corresponding epipolar line in the other image may be decomposed into two steps.
    - **first step**, the point **x** is mapped to some point **x'** in the other image lying on the epipolar line **l'**. This point **x'** is a potential match for the point **x**.
    - **second step**, the epipolar line **l'** is obtained as the line joining **x'** to the epipole **e'**.

Geometry of Multiple Views
2- and 3- view geometry

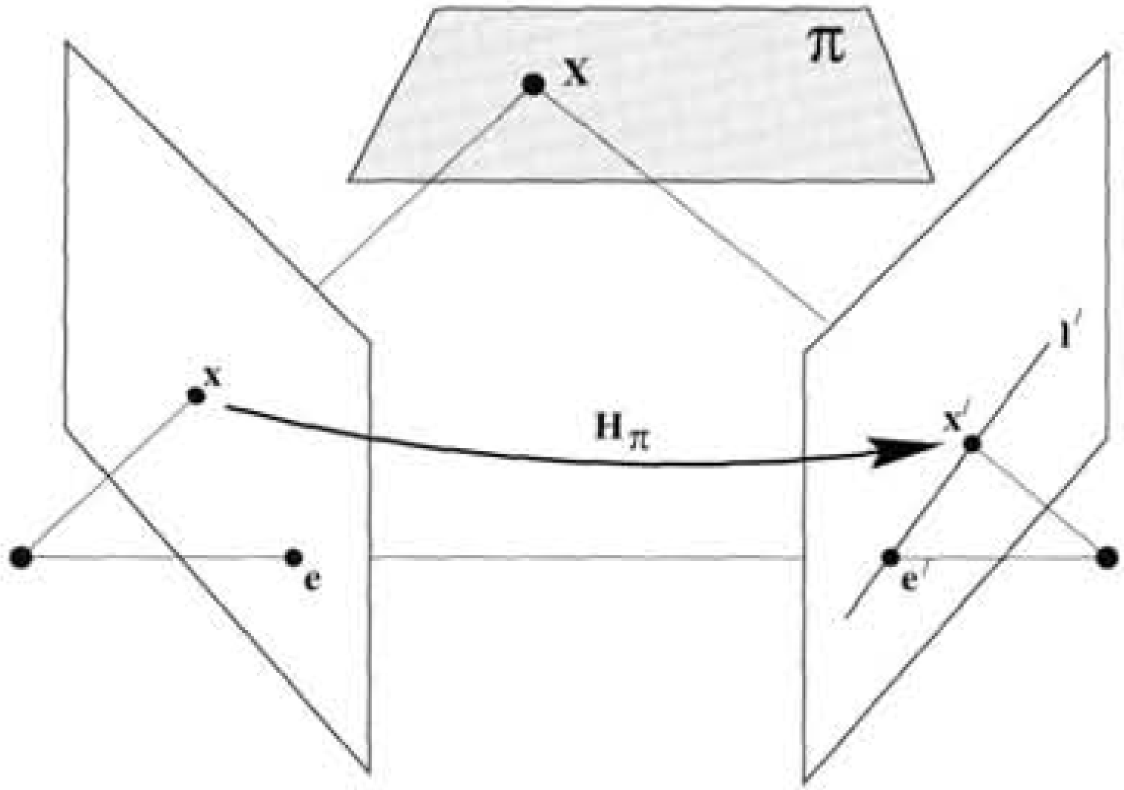
- Fundamental matrix F
  - **Geometric derivation**

$$l' = e' \times x' = [e']_\times x'$$

$$x' = Hx$$

$$l' = [e']_\times Hx = Fx$$

$$[v]_\times = \begin{bmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{bmatrix}$$



Geometry of Multiple Views
2- and 3- view geometry

- Fundamental matrix F
  - **Correspondence condition: map x→l'**
    - *The fundamental matrix satisfies the condition that for any pair of corresponding points* x ↔ x' *in the two images*

$$\mathrm{x}' = \begin{bmatrix} x' & y' & 1 \end{bmatrix}$$
$$\mathrm{x} = \begin{bmatrix} x & y & 1 \end{bmatrix}$$

$$\mathrm{x}'^{\mathrm{T}}\,\mathrm{F}\mathrm{x} = 0$$

$$\begin{bmatrix} x' & y' & 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = 0$$

$$x'xf_{11} + x'yf_{12} + x'f_{13} + y'xf_{21} + y'yf_{22} + y'f_{23} + xf_{31} + yf_{32} + f_{33} = 0$$

Geometry of Multiple Views
2- and 3- view geometry

- Fundamental matrix F
  - Properties
    - **Transpose:** If **F** is the **fundamental matrix** of the **pair of cameras (P, P')**, then $\mathbf{F}^T$ is the fundamental matrix of the pair in the **opposite order: (P', P)**.
    - **Epipolar lines:** For any **point x** in the first image, the corresponding epipolar line is $\mathbf{l}' = \mathbf{F}\mathbf{x}$. Similarly, $\mathbf{l} = \mathbf{F}^T\mathbf{x}'$ represents the **epipolar line** corresponding to **x'** in the second image.
    - The **epipole:** for any **point x** (other than e) the epipolar line $\mathbf{l}' = \mathbf{F}\mathbf{x}$ contains the epipole **e'**. Thus **e'** satisfies $\mathbf{e}'^T(\mathbf{F}\mathbf{x}) = (\mathbf{e}'^T\mathbf{F})\mathbf{x} = 0$ for all **x**. It follows that $\mathbf{e}'^T\mathbf{F} = 0$, i.e. **e'** is the left null-vector of **F**. Similarly $\mathbf{F}\mathbf{e}=0$, i.e. e is the right null-vector of **F**.
    - F has seven degrees of freedom: a 3 x 3 homogeneous matrix has eight independent ratios; however, F also satisfies the constraint detF=0 which removes one degree of freedom.
    - **F is a *correlation,* a projective map taking a point to a line**. In this case **a point in the first image x defines a line in the second l' = Fx**, which is the epipolar line of x. **If l and l' are corresponding epipolar lines then any point x on l is mapped to the same line l'.** This **means there is no inverse mapping**

Geometry of Multiple Views
2- and 3- view geometry

- Fundamental matrix F
  - Epipolar line equation:

$$\mathrm{l}' \equiv ax + by + c = 0$$

$$(a, b, c)^T$$

$$\mathrm{l}' = \mathrm{F}\mathrm{x}$$

$$\mathrm{l}' = \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

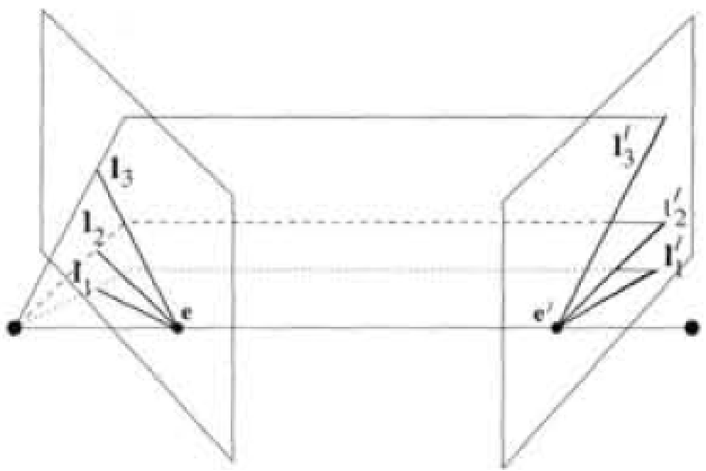
$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} f_{11}x + f_{12}y + f_{13} \\ f_{21}x + f_{22}y + f_{23} \\ f_{31}x + f_{32}y + f_{33} \end{bmatrix}$$

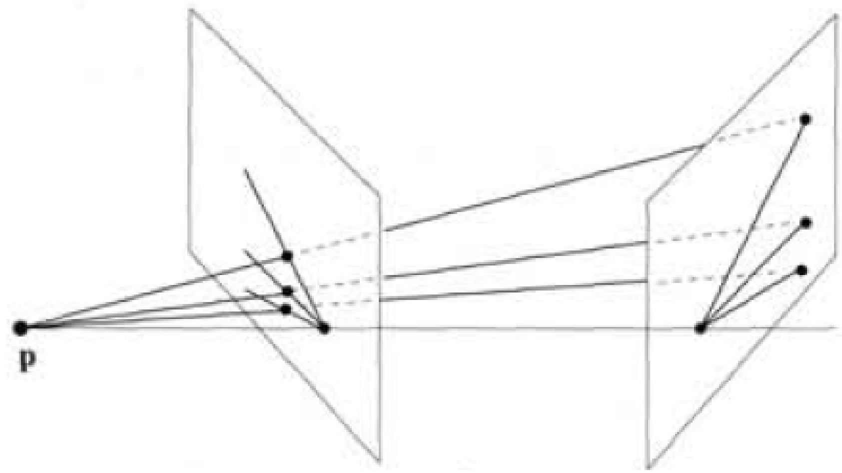
  - x belongs to l’

$$\mathrm{x}^{\mathrm{T}}\mathrm{l}' = 0$$

Geometry of Multiple Views
2- and 3- view geometry

- Fundamental matrix F
  - **The epipolar line homography**
    - The **set of epipolar lines in each of the images** forms a **pencil of lines passing through the epipole**.
    - Such a pencil of lines may be considered as a 1-dimensional projective space.
    - It is clear from figure that **corresponding epipolar lines** are **perspectively related**, so that there is a **homography between the pencil of epipolar lines centred at e** in the first view, and **the pencil centred at e' in the second**. A homography between two such 1-dimensional projective spaces has 3 degrees of freedom.





Geometry of Multiple Views
2- and 3- view geometry

- Fundamental matrix F
  - **The epipolar line homography**
    - The **7** degrees of freedom of the fundamental matrix can thus be counted as follows:
      - **2 for e, 2 for e',**
      - **3 for the epipolar line homography** which maps a line through **e** to a line through **e'**.

Geometry of Multiple Views
2- and 3- view geometry

- Fundamental matrix F

$$\begin{bmatrix} x' & y' & 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = 0$$

$$x'xf_{11} + x'yf_{12} + x'f_{13} + y'xf_{21} + y'yf_{22} + y'f_{23} + xf_{31} + yf_{32} + f_{33} = 0$$

$$(x'x, x'y, x', y'x, y'y, y', x, y, 1)\mathrm{f} = 0$$

$$\mathrm{Af} = 0$$

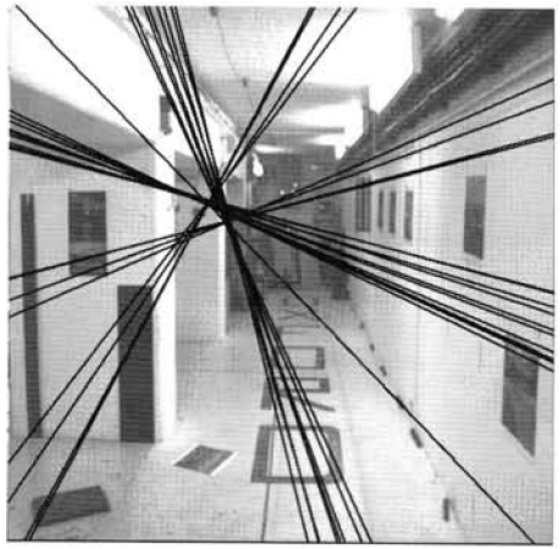
$$\begin{bmatrix} x'_1 x_1 & x'_1 y_1 & x'_1 & y'_1 x_1 & y'_1 y_1 & y'_1 & x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x'_n x_n & x'_n y_n & x'_n & y'_n x_n & y'_n y_n & y'_n & x_n & y_n & 1 \end{bmatrix}$$

Geometry of Multiple Views
2- and 3- view geometry

- Fundamental matrix F
  - For a solution to exist, matrix **A** must have rank at most 8, and if the rank is **exactly 8**, then the **solution is unique**, and can be found **by linear methods**
  - If the data is not exact, because of noise in the point coordinates, then the rank of **A may be greater than 8** (in fact equal to 9, since A has 9 columns). In this case, one finds a **least-squares solution**.
  - The **least-squares solution** for **f** is the **singular vector corresponding to the smallest singular value of A.**
  - The solution vector **f** found in this way minimizes $||Af||$ subject to the condition $||f|| = 1$.
  - The algorithm just described is the essence of a method called the **8-point algorithm for** computation of the fundamental matrix.

Geometry of Multiple Views
2- and 3- view geometry

- Fundamental matrix F
  - **The singularity constraint**
    - An important property of the **fundamental matrix** is that it is singular, in fact of rank 2.
    - Furthermore, the left and right null-spaces of F are generated by the vectors representing (in homogeneous coordinates) the **two epipoles in the two images**.
    - If the **fundamental matrix is not singular** then **computed epipolar lines** are not coincident, as is demonstrated by figure

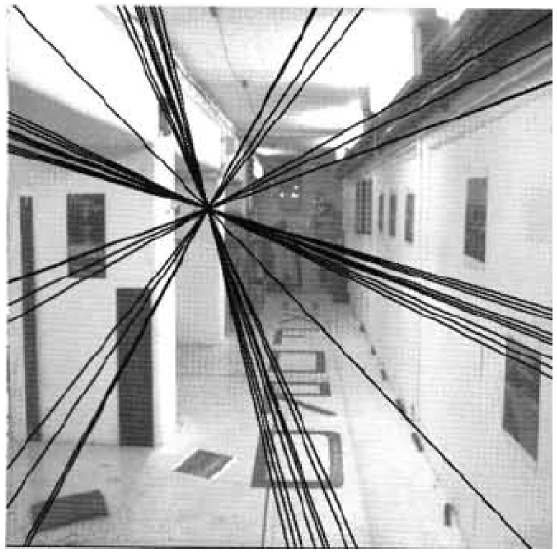


Geometry of Multiple Views 2- and 3- view geometry

- Fundamental matrix F
  - **The singularity constraint**
    - The **matrix F found** by solving the set of linear equations **will not in general have rank 2**, and we should take steps to enforce this constraint.
    - The most convenient way to do this **is to correct the matrix F found by the SVD solution from A**.
    - Matrix F is replaced by the matrix F' that minimizes the Frobenius norm ||F — F'|| subject to the condition detF'=0

SVD – single value decomposition solution of over-determined systems of equations

$$\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2}$$



Geometry of Multiple Views
2- and 3- view geometry

- Fundamental matrix F
  - The 8-point algorithm for computation of the fundamental matrix may be formulated as consisting of two steps, as follows.
    - **Linear solution.** A solution F is obtained from the vector f corresponding to the smallest singular value of A,
    - **Constraint enforcement.** Replace F by F', the closest singular matrix to F under a Frobenius norm. This correction is done using the SVD.

Geometry of Multiple Views
2- and 3- view geometry

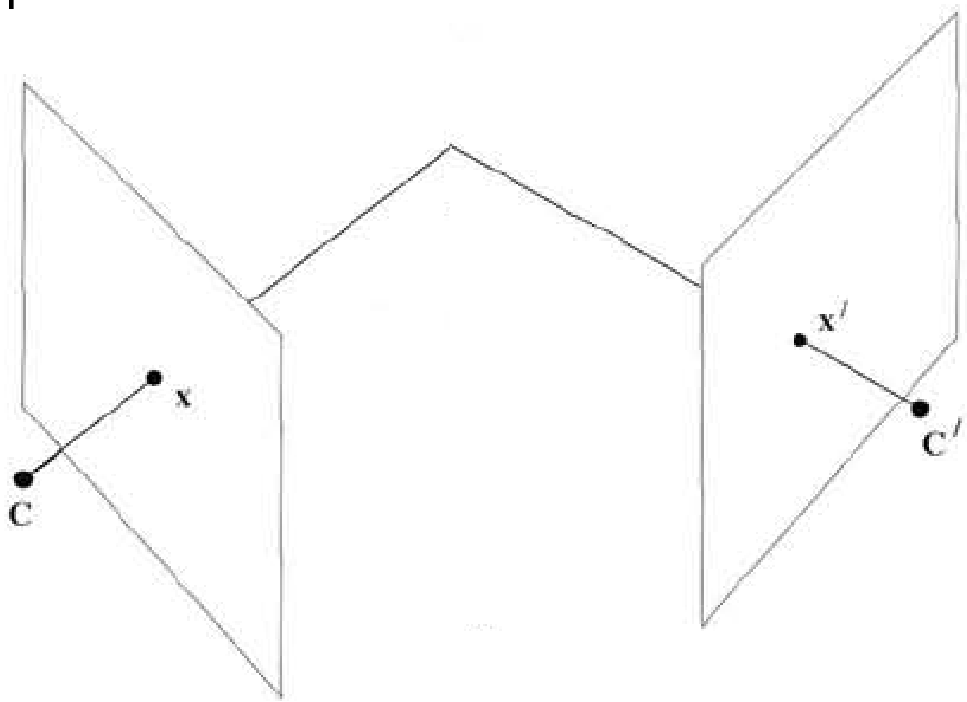
- Fundamental matrix F
  - The 8-point algorithm assume that we know the correspondences between the images
  - In general, we do not know the correspondences and we must find them automatically
    - Solution:
      - RANSAC = RANdom SAmple Consensus

Geometry of Multiple Views
2- and 3- view geometry

- Fundamental matrix F
  - Automatic computation of F
    - RANSAC
      - Do *k* times:
        - » Draw set with minimum number of correspondences
        - » Fit ***F*** to the set
        - » Count **the number *d* of correspondences that are closer than *t* to the fitted epipolar lines**
        - » **If *d* > *d*min, recompute fit error using all the correspondences**
      - Return best fit found

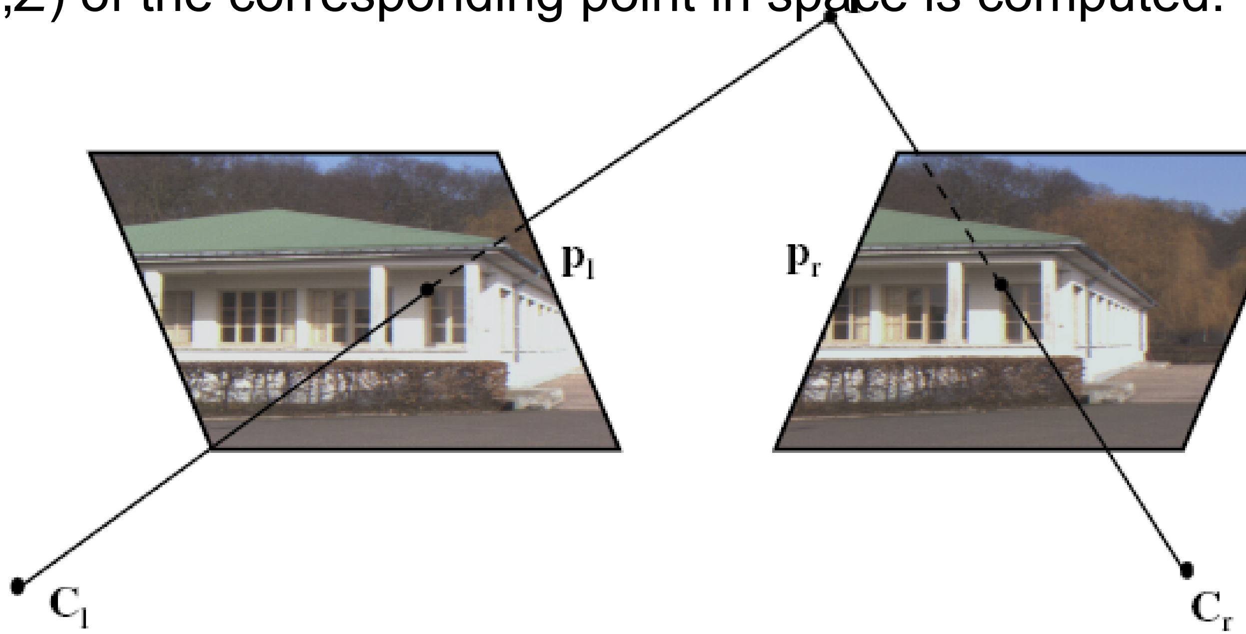
Geometry of Multiple Views
2- and 3- view geometry

- Structure Computation
  - Back-projecting rays
    - From the measured image points
    - Triangulation

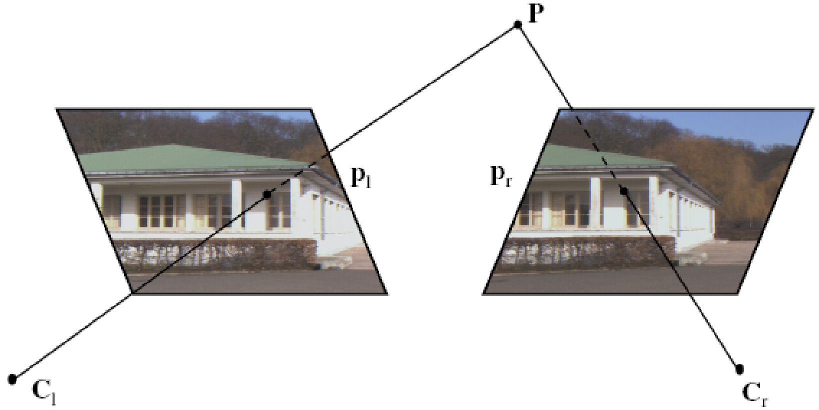


Stereo Vision

- Stereo Vision
  - Stereo vision is the process of recovering the three-dimensional location of points in the scene from their projections in images. More precisely, if we have two images I_l and I_r (left and right from the left and right eyes), given a pixel $\mathbf{p}_l$ in the left image and the corresponding pixel $\mathbf{p}_r$ in the right image, then the coordinates (X,Y,Z) of the corresponding point in space is computed.



Stereo Vision



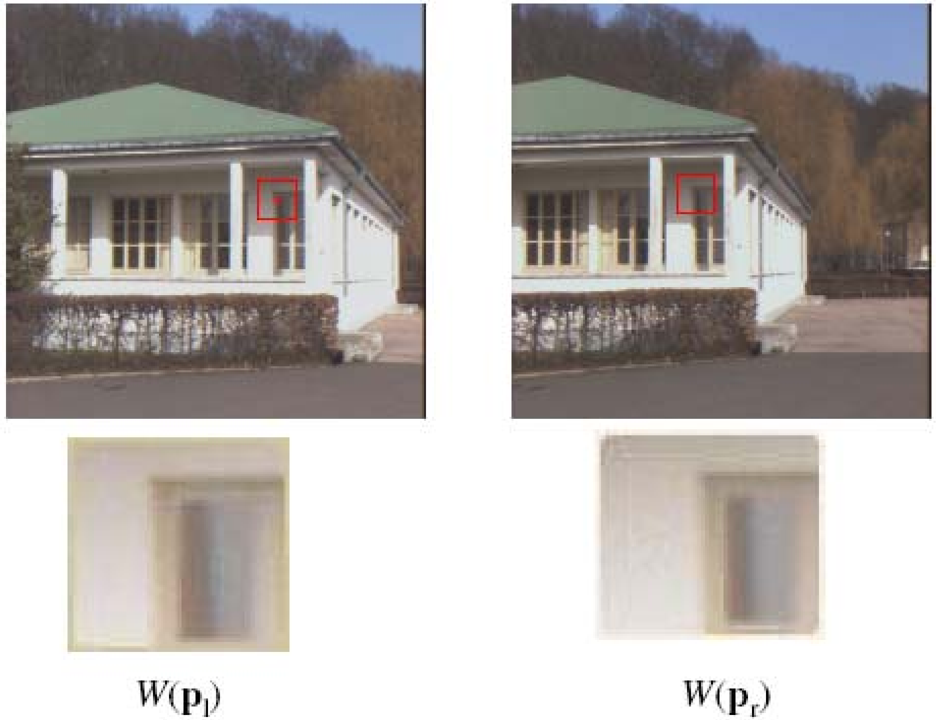
- Stereo Vision
  - Geometrically, given $\mathbf{p}_l$, we know that the point **P** lies on the line $\mathbf{L}_l$ joining $\mathbf{p}_l$ and the left optical center $\mathbf{C}_l$ (this line is the *viewing ray*), although we don't know the distance along this line.
  - Similarily, we know that **P** lies along a line $\mathbf{L}_r$ joining $\mathbf{p}_r$ and **P**.
  - knowing exactly the parameters of the cameras (intrinsic and extrinsic), we can explicitly compute the parameters of $\mathbf{L}_l$ and $\mathbf{L}_r$. Therefore, we can compute the intersection of the two lines, which is the point **P.**
  - This procedure is called *triangulation*.

Stereo Vision

- Stereo Vision
  - *Correspondence*:
    - Given a point $\mathbf{p}_l$ in one image, find the corresponding point in the other image.
  - *Reconstruction*:
    - Given a correspondence $(\mathbf{p}_l,\mathbf{p}_r)$, compute the 3-D coordinates of the corresponding point in space, $\mathbf{P}$.

Stereo Vision

- *Correspondence*
  - Given $\mathbf{p}_l$, finding the corresponding point $\mathbf{p}_r$ involves searching the right image for the location $\mathbf{p}_r$ such that the right image around $\mathbf{p}_r$ "looks like" the left image around $\mathbf{p}_l$.
  - Take a small window *W* around $\mathbf{p}_l$ and compare it with the right image at all possible locations.
  - The position $\mathbf{p}_r$ that gives the best match is reported.
  - The fundamental operation, therefore, is to compare the pixels in a window $W(\mathbf{p}_l)$ with the pixels in a window $W(\mathbf{p}_r)$.
    - Sum of Absolute Differences
    - Sum of Squared Differences
    - Normalized Correlation



Stereo Vision

- **Matching Functions**

SSD:

$$\psi(I_l(x,y), I_r(x+d,y)) = (I_l(x,y) - I_r(x-d,y))^2$$

SAD:

$$\psi(I_l(x,y), I_r(x+d,y)) = \left| I_l(x,y) - I_r(x-d,y) \right|$$

Correlation:

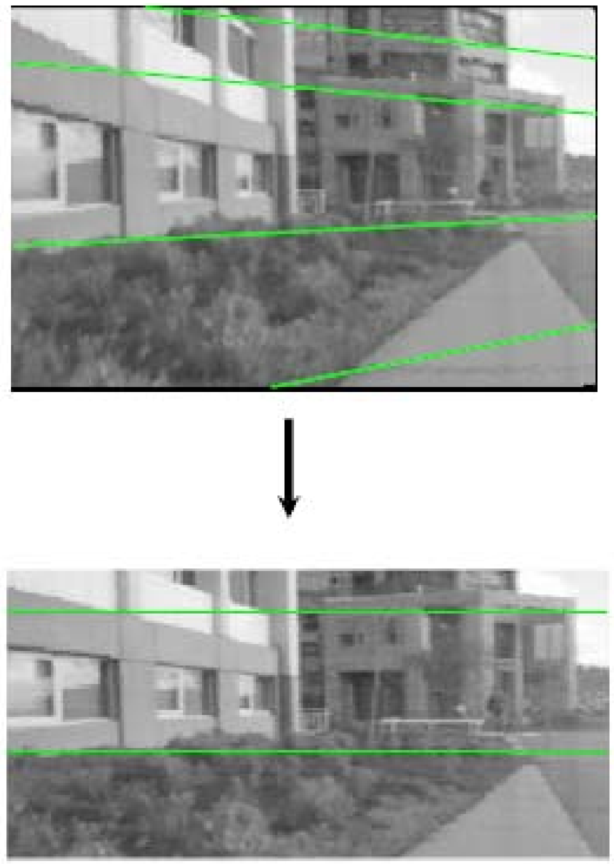
$$\psi(I_l(x,y), I_r(x+d,y)) = I_l(x,y).I_r(x-d,y)$$

Normalized Correlation:

$$\psi(I_l(x,y), I_r(x+d,y)) = \frac{I_l(x,y).I_r(x-d,y) - \bar{I}_l\bar{I}_r}{\sigma_l\sigma_r(d)}$$

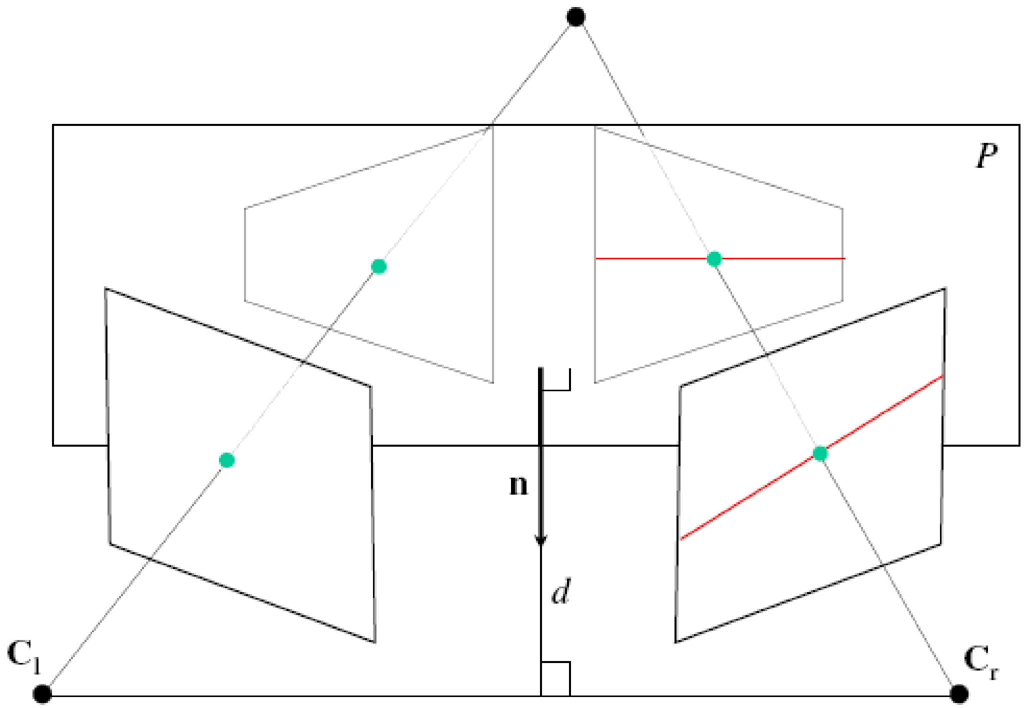
Stereo Vision

- *Correspondence*
  - **Rectification**
    - Searching along epipolar lines at arbitrary orientation is intuitively expensive.
    - Always search along the rows of the right image.
      - Given the epipolar geometry of the stereo pair, there exists in general a transformation that maps the images into a pair of images with the epipolar lines parallel to the rows of the image. This transformation is called *rectification*.
    - Images are almost always rectified before searching for correspondences in order to simplify the search.
    - The exception is when the epipole is inside one of the images. In that case, rectification is not possible.



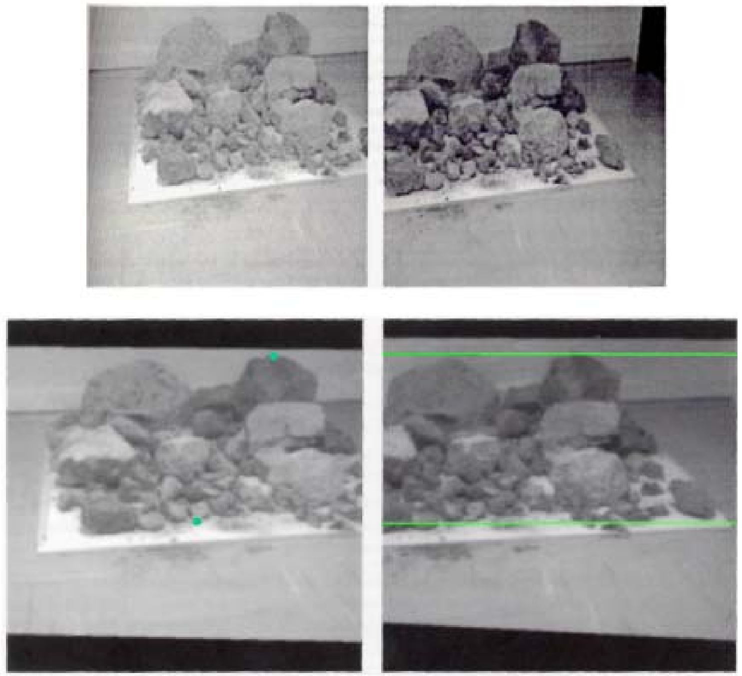
Stereo Vision

- *Correspondence*
  - **Rectification**
    - Given a plane **P** in space, there exists two homographies $\mathbf{H}_l$ and $\mathbf{H}_r$ that map each image plane onto **P**. That is, if $\mathbf{p}_l$ is a point in the left image, then the corresponding point in **P** is **Hp** (in homogeneous coordinates).
    - If we map both images to a common plane **P** such that P is parallel to the line $\mathbf{C}_l\mathbf{C}_r$, then the pair of virtual (rectified) images is such that the epipolar lines are parallel.



Stereo Vision

- *Correspondence*
  - **Rectification**
    - The algorithm for rectification is then:
      - Select a plane **P** parallel to $\mathbf{C}_r\mathbf{C}_l$
      - Define the left and right image coordinate systems on **P**
      - Construct the rectification matrices $\mathbf{H}_l$ and $\mathbf{H}_r$ from **P** and the virtual image's coordinate systems.

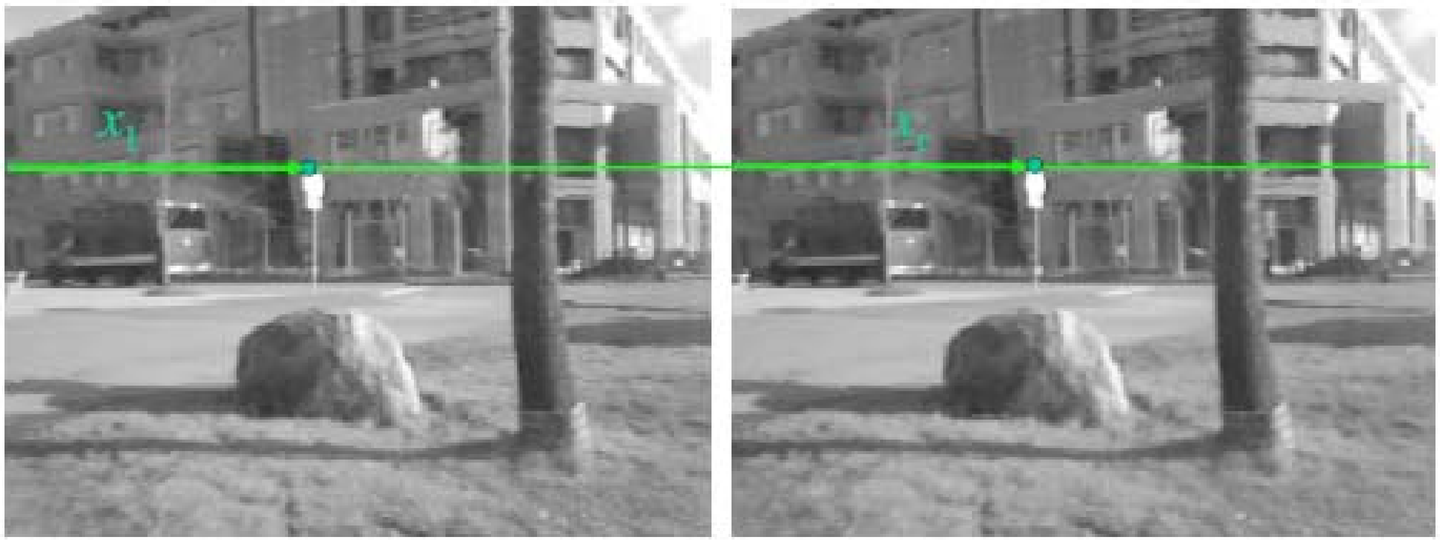


Stereo Vision

- *Disparity*
  - Assuming that images are rectified, given two corresponding points $\mathbf{p}_l$ and $\mathbf{p}_r$,
  - the difference of their coordinates along the epipolar line x_l-x_r is called the disparity d.
  - The disparity is the quantity that is directly measured from the correspondence.
  - The corresponding 3-D point $\mathbf{P}$ can be computed from $\mathbf{p}_l$ and d, assuming that the camera parameters are known.

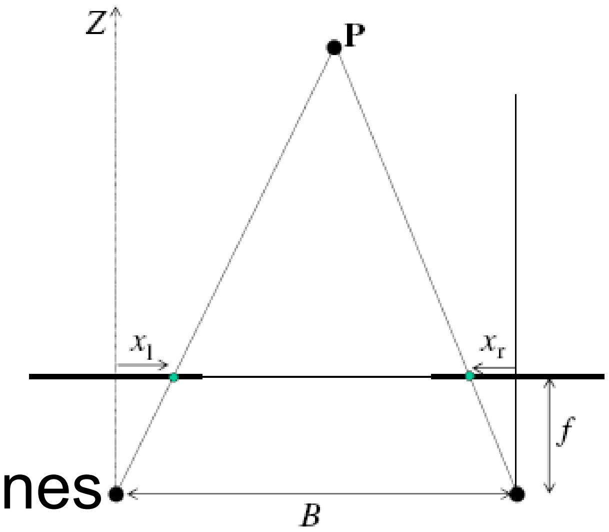
Stereo Vision

- *Disparity*



$$d = x_l - x_r$$

Stereo Vision



- Recover the 3-D coordinates
  - Two cameras with parallel image planes
    - Consider a point **P** of coordinates *X*,*Y*, *Z*.
    - **P** is projected to $\mathbf{p}_l$ and $\mathbf{p}_r$ in the two images.
    - The focal length of the cameras is denoted by *f*, and the distance between the optical centers (the baseline) is denoted by *B*.
    - Looking at the diagram, we see that the triangles ($\mathbf{C}_r$,$\mathbf{C}_l$,**M**) and ($\mathbf{p}_r$,$\mathbf{p}_l$,**P**) are similar triangles.
    - Therefore, the ratios of their height to base are equal

$$\frac{Z}{B} = \frac{Z - f}{B - x_l + x_r}$$

$$Z = Bf / d$$

Stereo Vision

- Recover the 3-D coordinates
  - Two cameras with parallel image planes
    - This relation is the fundamental relation of stereo.

$$Z = Bf \,/\, d$$

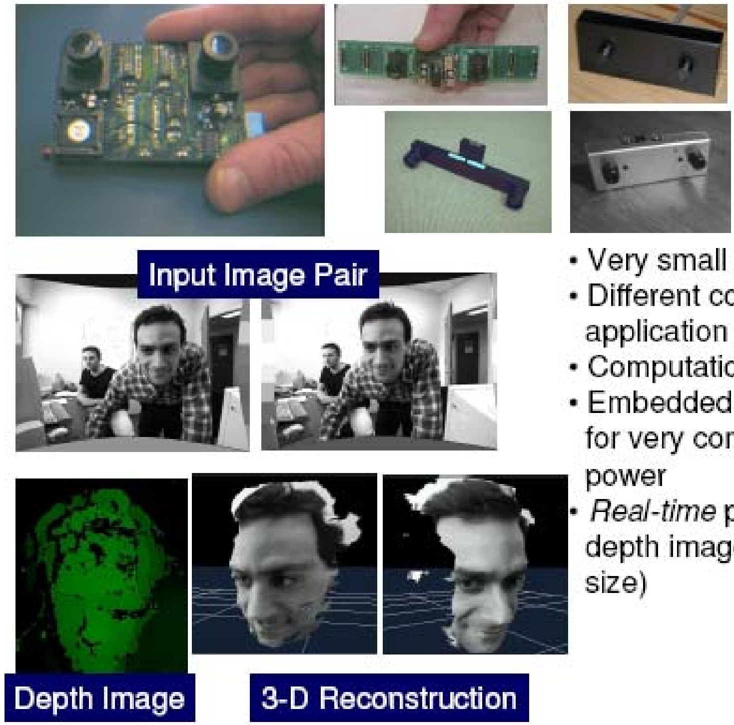
    - It basically states that the depth is inversely proportional to the disparity.
    - Once we know *Z*, the other two coordinates are derived using the standard perspective equations:

$$X = \frac{x_l Z}{f}$$

$$Y = \frac{y_l Z}{f}$$

Stereo Vision

- Recover the 3-D coordinates
  - Two cameras with parallel image planes
    - Commercial solutions

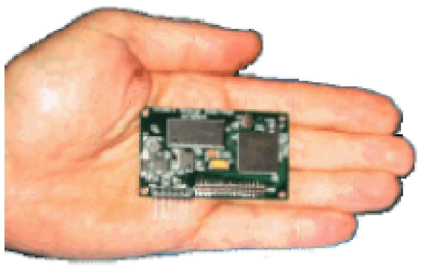


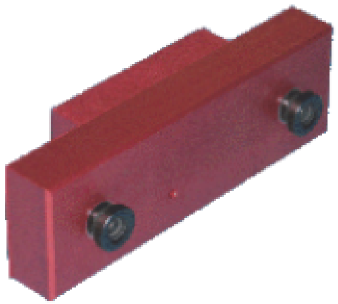
- Very small package
- Different configurations to meet application needs
- Computation on *conventional* PCs, or
- Embedded computing (DSP, FPGA) for very compact packaging and low power
- *Real-time* processing (e.g. 12-90 depth images/second, depending on size)

Stereo Vision

- Recover the 3-D coordinates
  - Two cameras with parallel image planes
    - Commercial solutions

Videre





Stereo Vision

- Recover the 3-D coordinates
  - Two cameras with parallel image planes
    - Commercial solutions

- Three cameras for increased depth precision
- Computation on standard PC
- Real-time (16 full-size depth images/second)
- Color cameras

Stereo Vision

- Recover the 3-D coordinates
  - Two cameras with parallel image planes
    - How accurately can those coordinates be computed?
      - Consider a matching pair of disparity *d* corresponding to a depth *Z*. If we make an error of one pixel (*d*+1 instead of *d*), the depth becomes *Z'*.
      - We want to evaluate D*Z*, the error in depth due to the error in disparity.
      - Taking the derivative of Z as a function of *d*, we get:

$$\left|\frac{\Delta Z}{\Delta d}\right| = \frac{Bf}{d^2} = \frac{Z^2}{Bf}$$

Stereo Vision

- Recover the 3-D coordinates
  - Two cameras with parallel image planes
    - The final relation illustrates the fundamental relation between baseline, focal length and accuracy of stereo reconstruction.
    - For an error of 1 pixel on disparity, we get an error in *Z* of:

$$\frac{Z^2}{Bf}$$

Stereo Vision

- Recover the 3-D coordinates
  - Two cameras with parallel image planes
    - *Depth:*
      - The resolution of the stereo reconstruction decreases quadratically with depth.
      - This implies severe limitation on the applicability of stereo.
      - If we assume sub-pixel disparity interpolation with a resolution Δd, the depth resolution becomes:

$$\frac{Z^2}{Bf}\Delta d$$

but it still remains quadratic in depth.

Stereo Vision

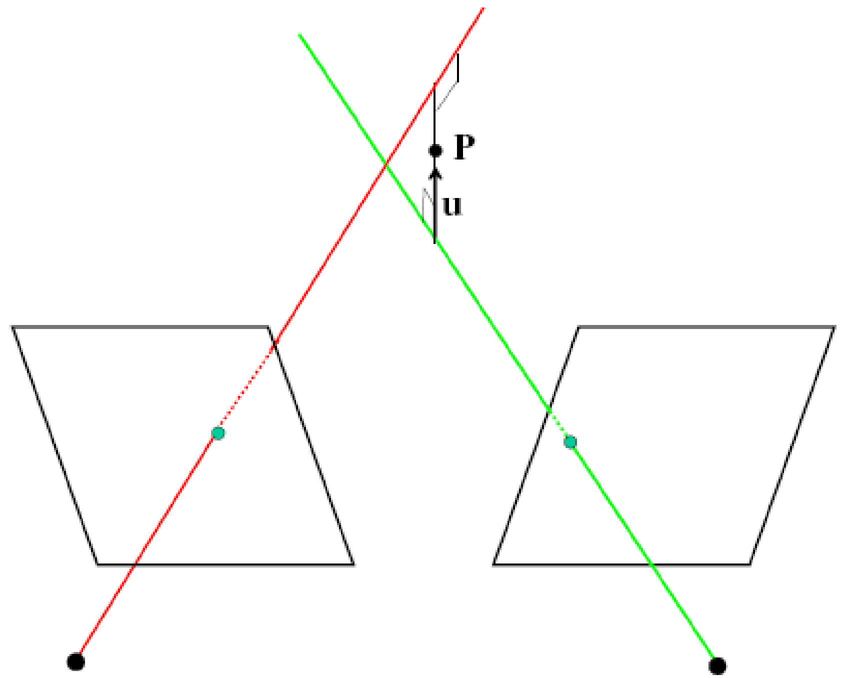
- Recover the 3-D coordinates
  - Two cameras with parallel image planes
    - *Baseline:*
      - The resolution improves as the baseline increases. We would be tempted to always use a baseline as large as possible.
      - However, the matching becomes increasingly difficult as the baseline increases (for a given depth) because of the increasing amount of distortion between the left and the right images.

$$\frac{Z^2}{Bf}$$

    - *Focal length:*
      - The resolution improves with focal length.
      - Intuitively, this is due to the fact that, for a given image size, the density of pixels in the image plane increases as *f* increases. Therefore, the disparity resolution is higher.

Stereo Vision

- Recover the 3-D coordinates
  - Two cameras with parallel image planes
    - The previous discussion assumes that the viewing rays from the left and right cameras intersect exactly.
    - That is not usually the case because of small errors in calibration.
    - The two viewing rays pass close to each other but do not exactly intersect.
    - The point **P** is reconstructed as the point that is the closest to both lines.



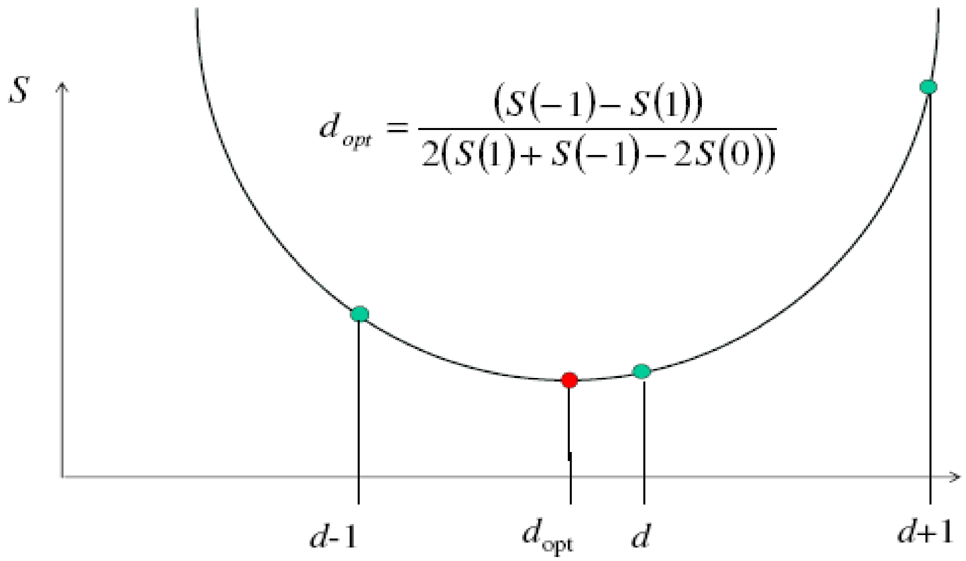
Stereo Vision

- Recover the 3-D coordinates
  - Two cameras with parallel image planes
    - Sub-pixel disparity
      - The disparity is computed by moving a window one pixel at a time.
      - As a result, the disparity is known only up to one pixel.
      - This limitation on the resolution of the disparity translates into a severe limitation on the accuracy of the recovered 3-D coordinates.
      - One effective way to get this problem is to recover the disparity at a finer resolution by interpolating between the pixel disparities using quadratic interpolation.

Stereo Vision

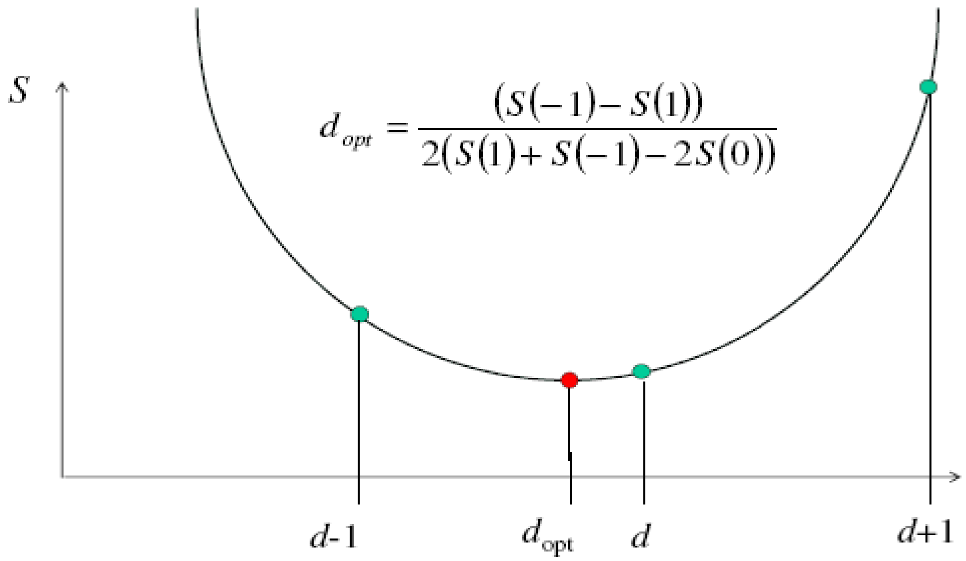
- Recover the 3-D coordinates
  - Two cameras with parallel image planes
    - Sub-pixel disparity

- Suppose that the best disparity at a pixel is obtained at $\boldsymbol{d_o}$ with a matching value (for example SSD) of $S(d_o)$.
- We can obtain a second order approximation of the (unknown) function $S(d)$ by approximating S by a parabola.
- At the position d_{opt} corresponding to the bottom of the parabola, we have $S(_{dopt})$<=$S(d_o)$. Therefore, d_{opt} is a better estimate of the disparity than d_o.
- The question remains as to how to find this approximating parabola.



Stereo Vision

- Recover the 3-D coordinates
  - Two cameras with parallel image planes
    - Sub-pixel disparity



- Let us first translate all the disparity values so that d_o=0.
- The equation of a general parabola is:
  - $S(d) = ad^2+bd+c.$
  - To recover the 3 parameters of the parabola we need 3 equations which we obtain by writing that the parabola passes through the point of disparity 0, -1, and +1:
  - $S(0) = c$ $S(1) = a+b+c$ $S(-1) = a-b+c$
  - Solving this, we obtain: $c = S(0)$ $a = (S(1) + S(-1) - 2S(0))/2$ $b = (S(1) - S(-1))/2$
  - The bottom of the parabola is obtained at *d*opt such that $S'(d) = 2ad+b = 0.$ Therefore, the optimal disparity is obtained

Stereo Vision

- Recover the 3-D coordinates
  - Two cameras with parallel image planes
    - **Matching Confidence**
      - Stereo matching assumes that there is enough information in the images to discriminate between different positions of the matching window.
      - That is not the case in regions of the image in which the intensity is nearly constant. In those regions, there is not a single minimum of the matching function and the disparity cannot be computed.
    - Detecting that the disparity estimate is unreliable
      - This can be done by compute the gradient of the image in the x direction, I_x and computing the local average of its squared magnitude: ΣI_x^2.
      - This quantity can then be used as a measure of reliability of matching (confidence measure.)

Stereo Vision

- Recover the 3-D coordinates
  - Two cameras with parallel image planes
    - **Lighting Issues**
      - A problem is that the lighting conditions maybe substantially different between the left and right image.
      - Because, for example, of different exposures or different settings of the camera.
      - Because ψ measures directly the difference in pixel values, its value will be corrupted.

Stereo Vision

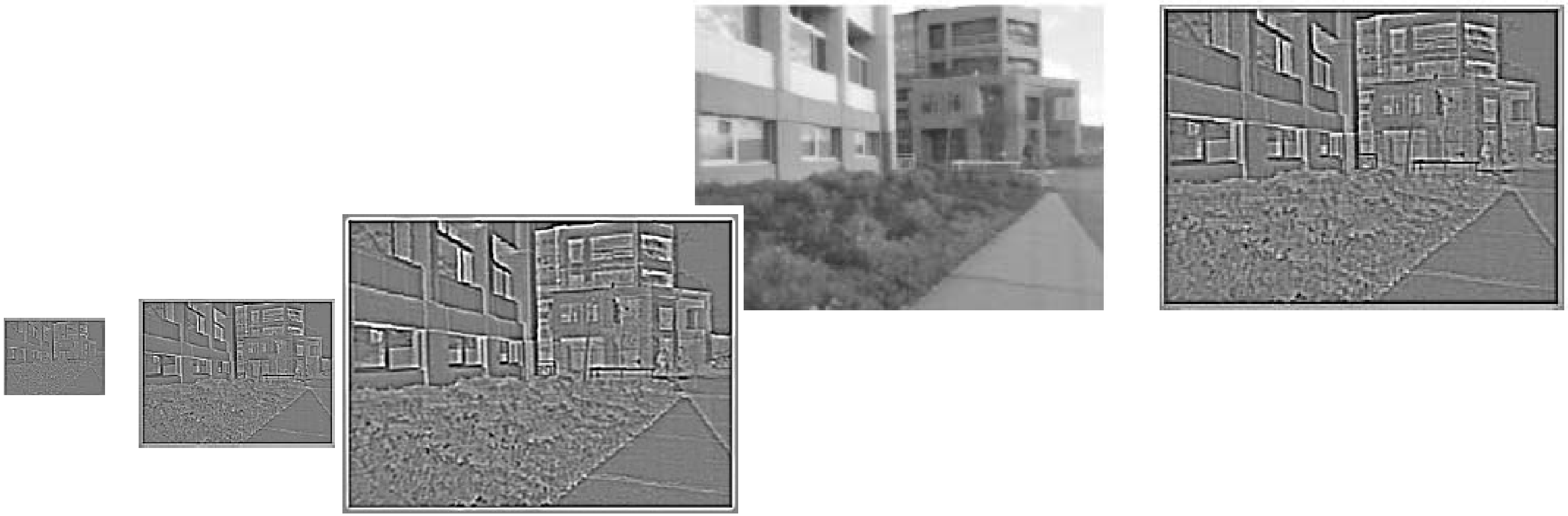
- Recover the 3-D coordinates
  - Two cameras with parallel image planes
    - **Lighting Issues**
      - The normalized correlation reduces this problem
      - **Laplacian of Gaussian (LOG)**
        - » smoothly varying parts of the image do not carry much information for matching. The useful information is contained in higher-frequency variations of intensity.

Stereo Vision

- Recover the 3-D coordinates
  - Two cameras with parallel image planes
    - **Laplacian of Gaussian (LOG)**
      - » smoothly varying parts of the image do not carry much information for matching. The useful information is contained in higher-frequency variations of intensity.
    - – So we want to eliminate the slowly-varying parts of the image (low-frequency), and this can be done by using a second derivatives.
    - – Also, we want to eliminate the high-frequency components which correspond to noise in the image, which suggests blurring with a Gaussian filter.
    - – The combination of the two suggests the use of the Laplacian of Gaussian (LOG) previously defined in the context of edge detection.
    - – This filter is technically a band-pass filter (removes both high and low frequencies.)
    - – In practice, the images are first convoluted with the LOG filter

Stereo Vision

- Recover the 3-D coordinates
  - Two cameras with parallel image planes
    - **Laplacian of Gaussian (LOG)**
      - » smoothly varying parts of the image do not carry much information for matching. The useful information is contained in higher-frequency variations of intensity.



Stereo Vision

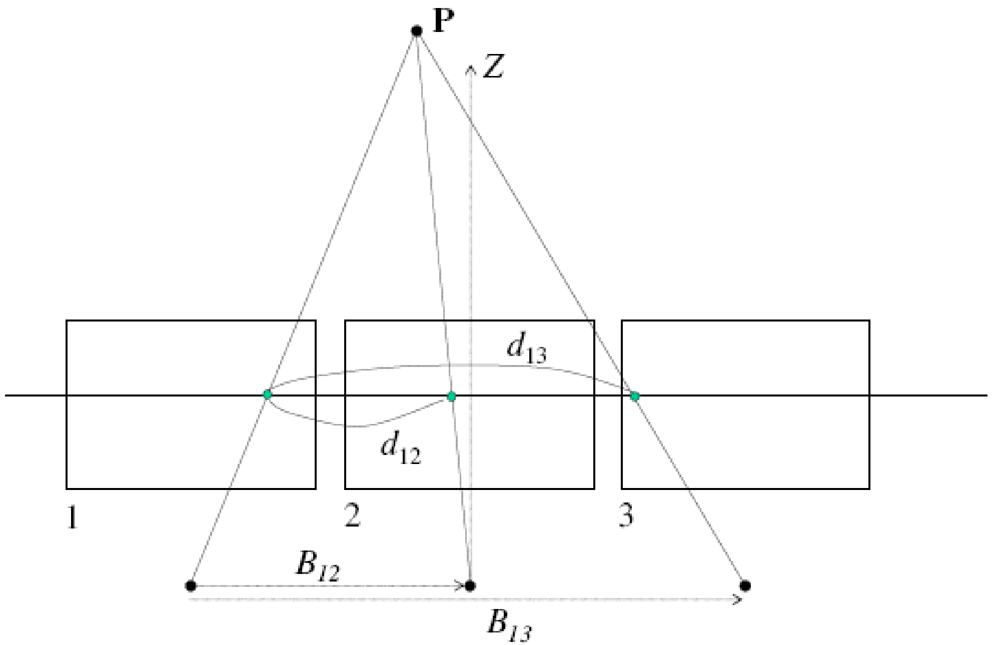
- Recover the 3-D coordinates
  - Two cameras with parallel image planes
    - **Effect of Window Size**
      - Window size has qualitatively the same effect as smoothing.
      - Localization is better with smaller windows. In particular, the disparity image is corrupted less near occluding edges because of better localized match.
      - Matching is better with larger windows because more pixels are taking into account, and there is therefore better discrimination between window positions.
      - Localization degrades as the window size increases, for the same reason.
      - For large values of *w*, the minimum is flatter (lower curvature), leading to lower confidence in the matching as defined before.

Stereo Vision

- Recover the 3-D coordinates
  - Two cameras with parallel image planes
    - **Ambiguity**
      - In many cases, several positions along the epipolar line can match the window around the left pixel.
      - In that case, there is an ambiguity as to which point leads to the correct 3-D reconstruction. If the ambiguous matches are far apart, they will correspond to very different points in space, thus leading to large errors.
      - In practice, it is nearly impossible to eliminate all such ambiguities, especially in environments with lots of regular structures. The solution is generally to use more than two cameras in stereo.

Stereo Vision

- Recover the 3-D coordinates
  - >2 cameras - “multibaseline” approach
    - Suppose that the three cameras are aligned.
    - A point P corresponds to three points in the images, aligned along the horizontal epipolar line.



Stereo Vision

- Recover the 3-D coordinates
  - >2 cameras - “multibaseline” approach
    - Suppose that the three cameras are aligned.
    - A point P corresponds to three points in the images, aligned along the horizontal epipolar line.
    - The disparities of those points d_{12} and d_{13} between the first image and images 2 and 3 are in general different because the baselines are different:

$$d_{12} = \frac{B_{12}}{Z}$$

$$d_{13} = \frac{B_{13}}{Z}$$

$$\frac{1}{Z} = \frac{d_{12}}{B_{12}} = \frac{d_{13}}{B_{13}}$$

Stereo Vision

- Recover the 3-D coordinates
  - >2 cameras - “multibaseline” approach
    - if we plot the matching curves ($S_{12}(d')$) not as a function of the disparity but at a function of $d' = d/B$, assuming that the matches are correct, all the curves should have the same minimum at $d' = 1/Z$.
    - Instead of using S_{12} and S_{13} separately, we can just combine them into a single matching function: $S(d') = S_{12} + S_{13}$
    - and find the minimum of $S(d')$.

Stereo Vision

- Recover the 3-D coordinates
  - >2 cameras - “multibaseline” approach
  - Using a multibaseline approach combines the advantages of both short and long baseline:
    - The short baseline will make the matching easier but leads to poorer localization in space. The longer baseline leads to higher precision localization but more difficult matching. Furthermore, ambiguous matches would not generally appear at multiple baselines, thus, by combining the matching function into a single function, we get a sharper, unique peak of the matching function.