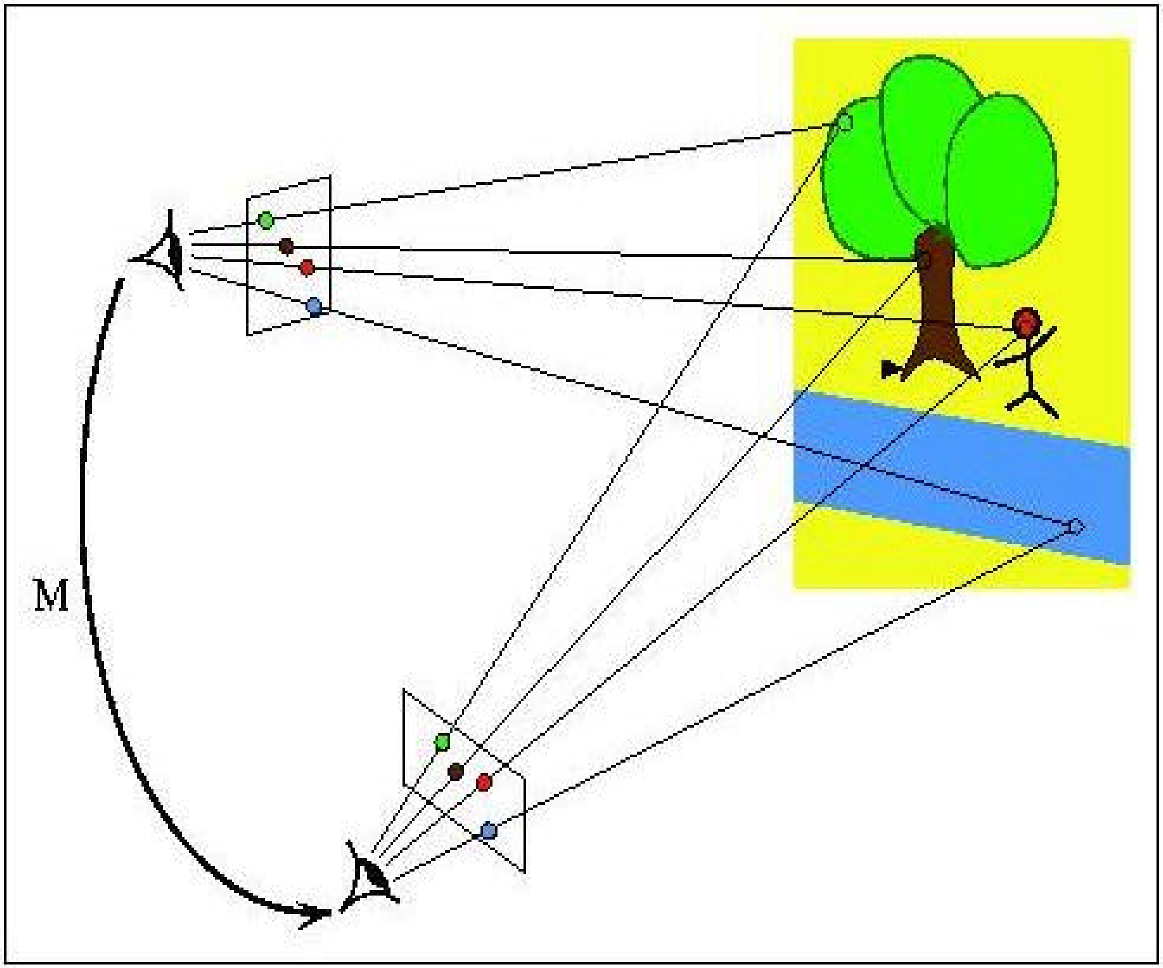
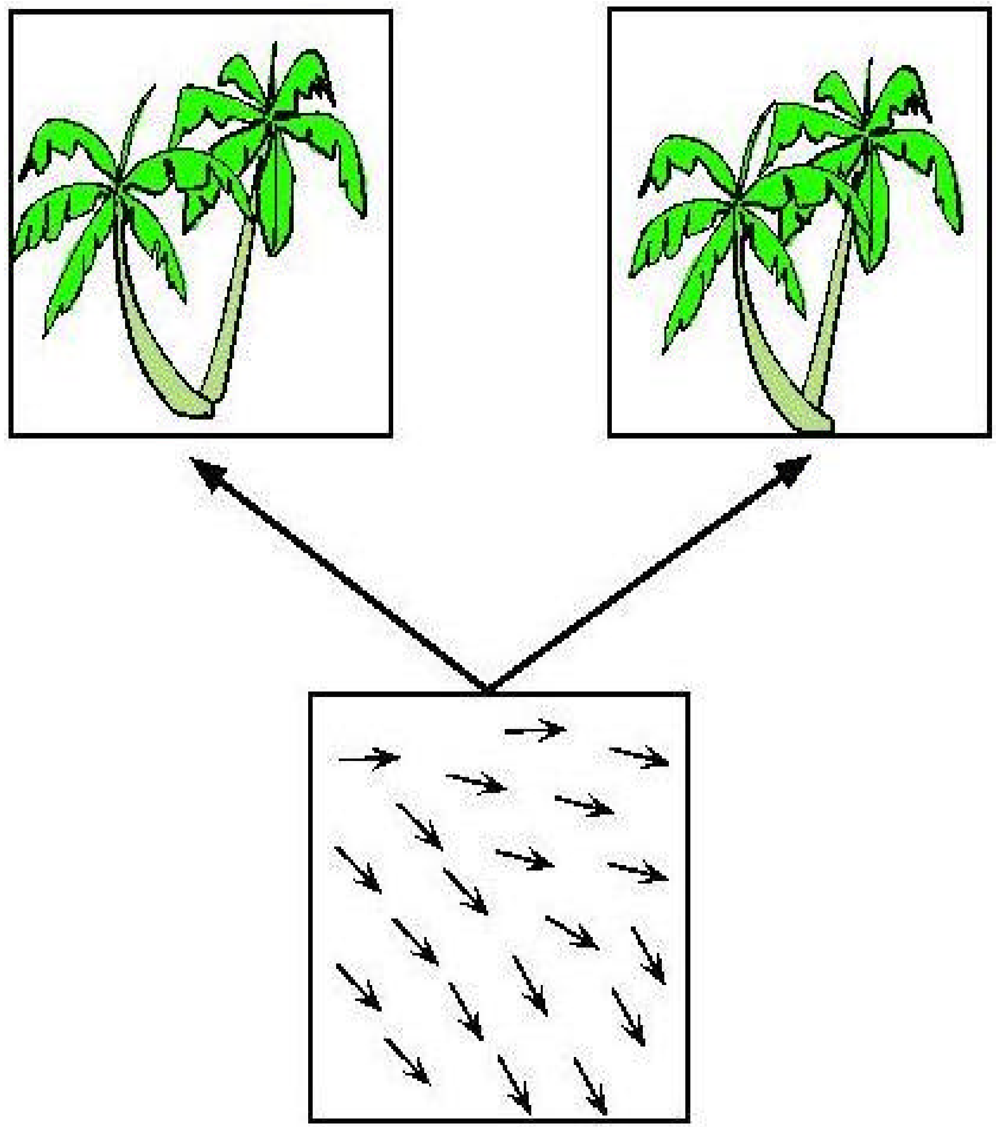
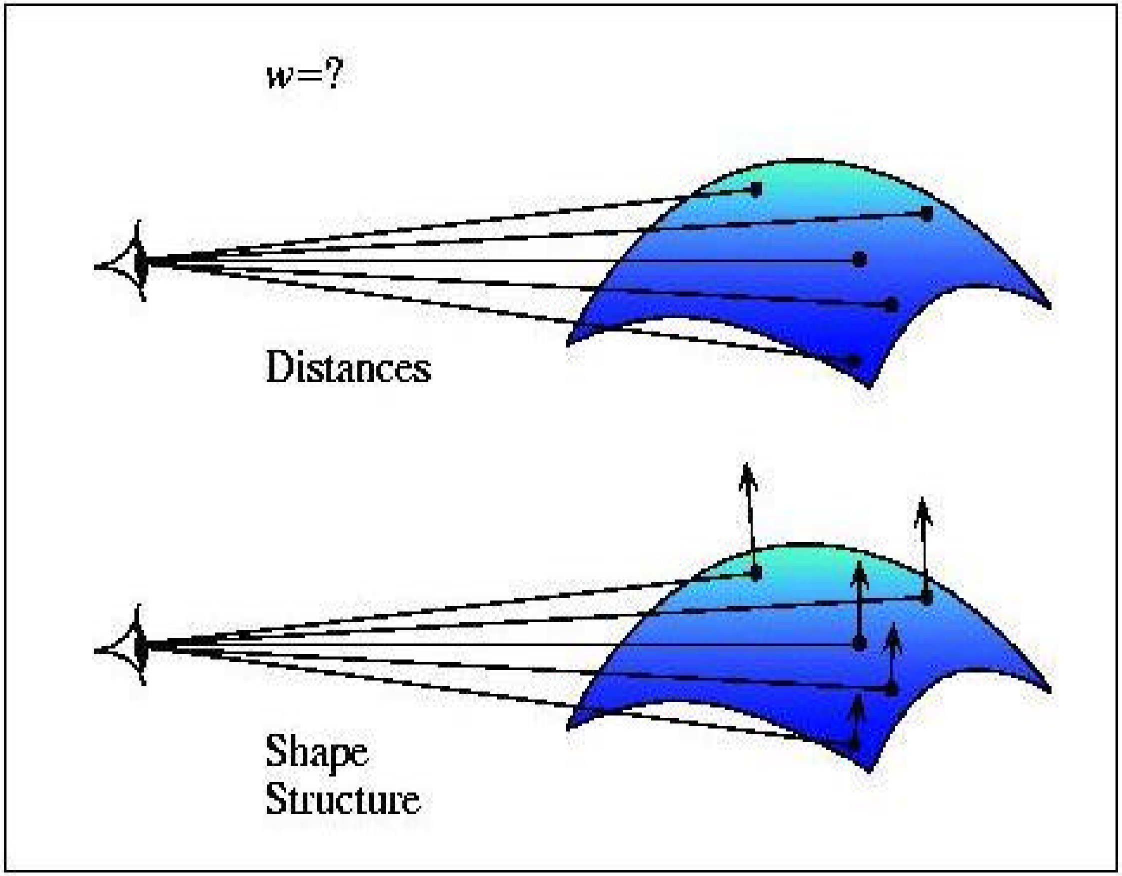
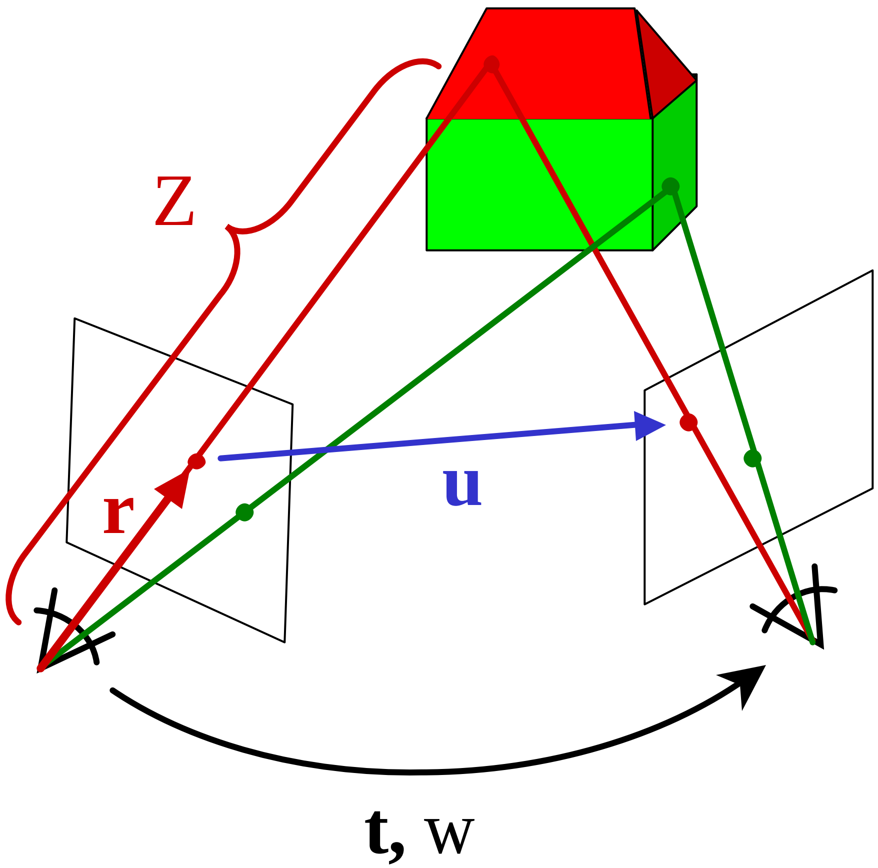
Motion and Flow II


M




w=?
Distances
Shape
Structure

Structure from Motion

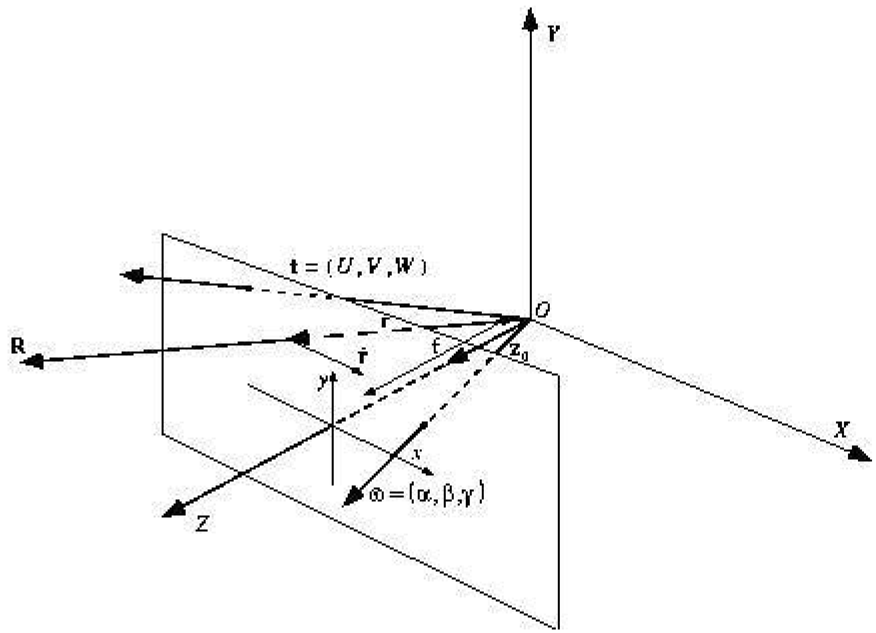


$$\mathbf{u} = \mathbf{u}_{\mathrm{tr}} + \mathbf{u}_{\mathrm{rot}}$$

$$\mathbf{u}_{\mathrm{tr}} = \frac{1}{Z}\left(\hat{\mathbf{z}} \times \left(\mathbf{t} \times \mathbf{r}\right)\right)$$

$$\mathbf{u}_{\mathrm{rot}} = \frac{1}{F}\left(\hat{\mathbf{z}} \times \left(\mathbf{r} \times \left(\left[\ ?\ \right]_{\times} \mathbf{r}\right)\right)\right)$$

Passive Navigation and Structure



The system moves with a rigid motion with translational velocity

$\mathbf{t} = (U, V, W)^T$ and rotational velocity $w = (a, b, g)^T$.

Scene points $\mathbf{R} = (X, Y, Z)^T$ project onto image points $\mathbf{r} = (x, y, f)$

and the 3D velocity $\dot{\mathbf{R}} = (U, V, W)$ of a scene point is observed in the image as velocity $\dot{\mathbf{r}} = (u, v.0)$.

Image Flow due to Rigid Motion

The velocity of a point with respect to the *XYZ* coordinate system is

$$\dot{\mathbf{R}} = -\mathbf{t} - \mathrm{w} \times \mathbf{R}$$

$$\dot{X} = -U - BZ + gY$$

$$\dot{Y} = -V - gX + aZ$$

$$\dot{Z} = -W - aX + bX$$

Let $f = 1$, then $x = \frac{X}{Z} \quad y = \frac{Y}{Z}; \quad u = \dot{x} \quad v = \dot{y}$

$$u = \left(\frac{\dot{X}}{Z}\right) = \frac{\dot{X}}{Z} - \frac{X\dot{Z}}{Z^2} = \left(-\frac{U}{Z} - b + gy\right) - x\left(-\frac{W}{Z} - ay + bx\right)$$

$$v = \left(\frac{\dot{Y}}{Z}\right) = \frac{\dot{Y}}{Z} - \frac{Y\dot{Z}}{Z^2} = \left(-\frac{V}{Z} - gx + a\right) - y\left(-\frac{W}{Z} - ay + bx\right)$$

$$u = \frac{-U + xW}{Z} + axy - b\left(x^2 + 1\right) + gy = \frac{u_{\mathrm{tr}}}{Z} + u_{\mathrm{rot}}$$

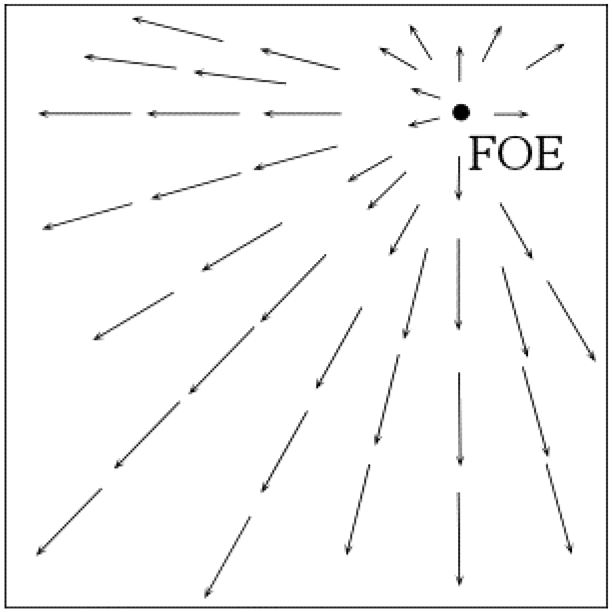
$$v = \frac{-V + yW}{Z} + a\left(y^2 + 1\right) - bxy - gx = \frac{v_{\mathrm{tr}}}{Z} + v_{\mathrm{rot}}$$

in vector notation : $\mathbf{r} = \frac{1}{\mathbf{R} \cdot \mathbf{z}_0}\mathbf{R}$, where $Z = \mathbf{R} \cdot \mathbf{z}_0$

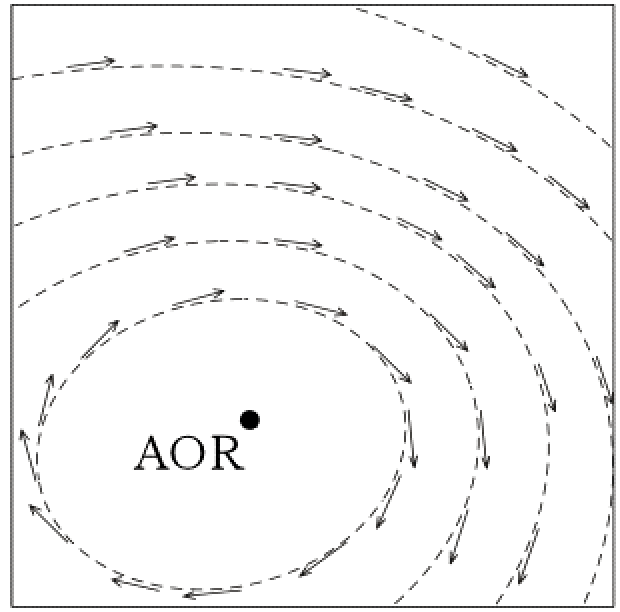
$$\dot{\mathbf{r}} = \frac{1}{\mathbf{R} \cdot \mathbf{z}_0}\left(\mathbf{z}_0 \times (\mathbf{t} \times \mathbf{r})\right) + \mathbf{z}_0 \times \left(\mathbf{r} \times (\mathrm{w} \times \mathbf{r})\right)$$

Scaling ambiguity: We can compute the translation only up to a scale factor ($K\mathbf{t}$, KZ) give the same flow as ($\mathbf{t}$, Z).

Translational flow field



Rotational flow field



$$\frac{\mathbf{u}_{tr}}{Z} = \left((x - x_0)\frac{W}{Z}, (y - y_0)\frac{W}{Z} \right)$$

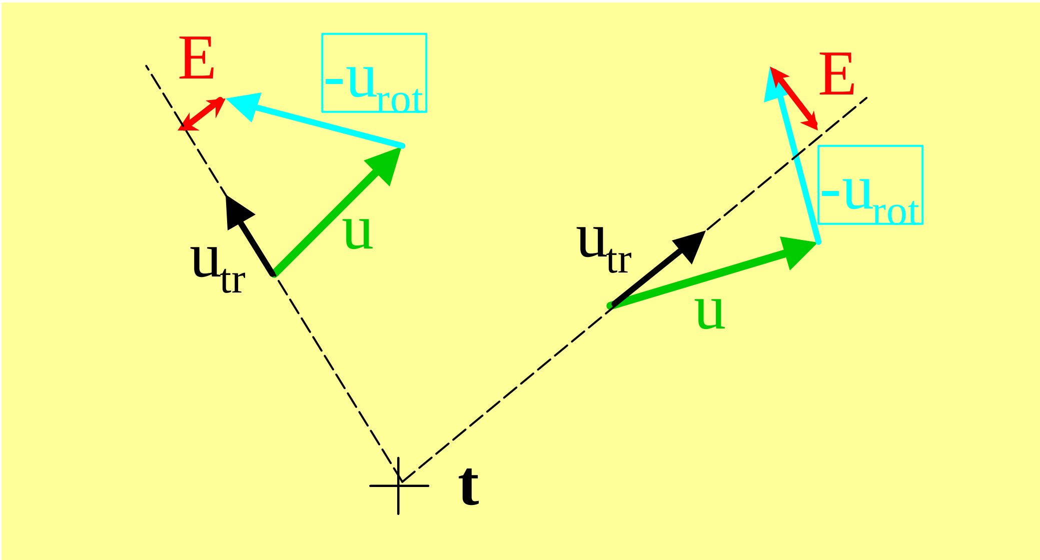
where $(x_0, y_0) = \left(\frac{U}{W} \cdot f, \frac{V}{W} \cdot f \right)$ is the focus of expansion (FOE)

or focus of contraction (FOC).

$\left(\frac{a}{g} f, \frac{b}{g} f \right)$ is the point where the rotation axis pierces the image plane (AOR).

Classical Structure from Motion

- Established approach is the epipolar minimization: The "derotated flow" should be parallel to the translational flow.



Uniqueness

Let there be two translations $\mathbf{t}_1, \mathbf{t}_2$ and two surfaces Z_1, Z_2

$$\mathbf{t}_1 = (U_1, V_1, W_1) \qquad \mathbf{t}_2 = (U_2, V_2, W_2)$$

$$u = \frac{-U_1 + xW_1}{Z_1} \qquad v = \frac{-V_1 + yW_1}{Z_1}$$

$$u = \frac{-U_2 + xW_2}{Z} \qquad v = \frac{-V_2 + yW_2}{Z}$$

$$\frac{\dfrac{-U_1 + xW_1}{-V_1 + yW_1}}{} = \frac{-U_2 + xW_2}{-V_2 + yW_2}$$

$$(-U_1 + xW_1)(-V_2 + yW_2) = (-U_2 + xW_2)(-V_1 + yW_1)$$

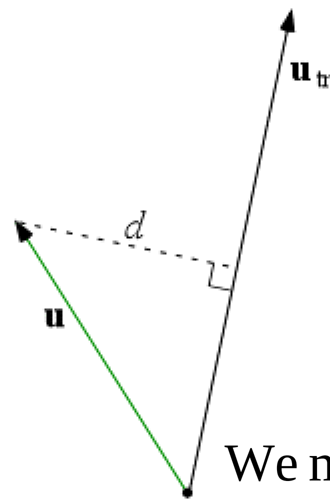
$$U_1V_2 - xV_2W_1 - yU_1W_2 + xyW_1W_2 = U_2V_1 - xV_1W_2 - yU_2W_1 + xyW_2W_1$$

must hold for all x and y

$$\begin{aligned} U_1V_2 &= U_2V_1 \\ V_2W_1 &= V_1W_2 \\ U_1W_2 &= U_2W_1 \end{aligned} \rightarrow \begin{matrix} U_1 : V_1 : W_1 = U_2 : V_2 : W_2 \\ Z_2 = KZ_1 \end{matrix} \rightarrow \mathbf{t}_2 = k\mathbf{t}_1 \text{ and } Z_2 = kZ_1$$

A translational flow field determines the motions of the camera uniquely up to a scaling factor.

The Translational Case



We minimize d^2

A least squares formulation

$$\iint \left(u - \frac{xW - U}{Z} \right)^2 + \left(v - \frac{yW - V}{Z} \right)^2 dx\,dy \rightarrow \min$$

Substitute $a = -U + xW \quad b = -V + yW$

$$\iint \left(u - \frac{a}{Z} \right)^2 + \left(v - \frac{b}{Z} \right)^2 \rightarrow \min$$

Step 1: Minimize with respect to Z. (Find the length of $\mathbf{u}_{tr}$ for which d would be minimized.)

$$\left(u - \frac{a}{Z} \right)\frac{a}{Z^2} + \left(v - \frac{b}{Z} \right)\frac{b}{Z^2} = 0 \qquad Z = \frac{a^2 + b^2}{ua + vb}$$

Substitute back

$$\iint \frac{(ub - va)^2}{a^2 + b^2} dx\,dy \rightarrow \min$$

Step 2: Differentiate with respect to *U*, *V*, *W*, set expression to zero.

Let $K = \frac{(ub - va)(ua + vb)}{(a^2 + b^2)^2}$

$$\mathrm{I} \iint (-V + yW) K \, dx \, dy = 0$$

$$\mathrm{II} \iint (-U + xW) K \, dx \, dy = 0$$

$$\mathrm{III} \iint (-yU + xV) K \, dx \, dy = 0$$

3 linearly dependent equations $(U \cdot \mathrm{I} + V \cdot \mathrm{II} + W \cdot \mathrm{III} = 0)$

The Rotational Case

$$\iint (u - u_{\mathrm{rot}})^2 + (v - v_{\mathrm{rot}})^2 \rightarrow \min$$

$$u - axy + b\left(x^2 + 1\right) - gy = 0$$

$$v - a\left(y^2 + 1\right) + bxy + gx = 0$$

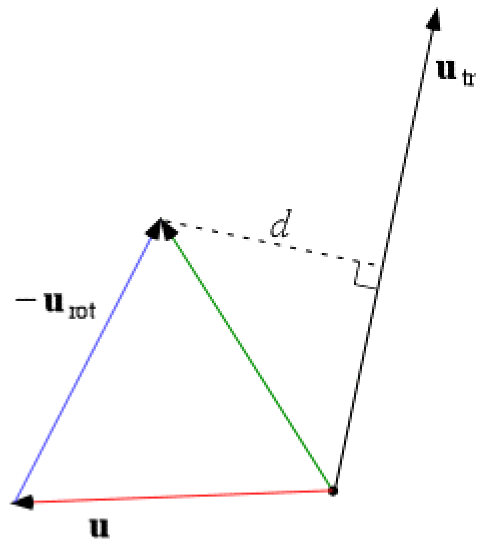
$$\begin{pmatrix} xy & -\left(x^2+1\right) & y \\ \left(y^2+1\right) & -xy & -x \end{pmatrix} \begin{pmatrix} a \\ b \\ g \end{pmatrix} = \begin{pmatrix} u \\ v \end{pmatrix}$$

In matrix form $A \cdot \mathrm{w} = \mathbf{u}$

$$\mathrm{w} = \left(A^T A\right)^{-1} A^T \mathbf{u}$$

The General Case

Minimization of epipolar distance



$$\iint \left(\begin{pmatrix} u - u_{\text{rot}} \\ v - v_{\text{rot}} \end{pmatrix} \cdot \begin{pmatrix} -v_{\text{tr}} \\ u_{\text{tr}} \end{pmatrix} \right)^2 dx\, dy \rightarrow \min$$

or, in vector notation

$$\int \left((\mathbf{t} \times \mathbf{r})(\dot{\mathbf{r}} - \mathrm{w} \times \mathbf{r}) \right)^2 d\mathbf{r} \rightarrow \min$$

Motion Estimation Techniques

- Linearization (Tsai Huang 1984, Longuet-Higgins 1981) in the discrete case

  $\mathbf{x}' E\mathbf{x} = 0$, where $E = T_{\times}\mathbf{R}$

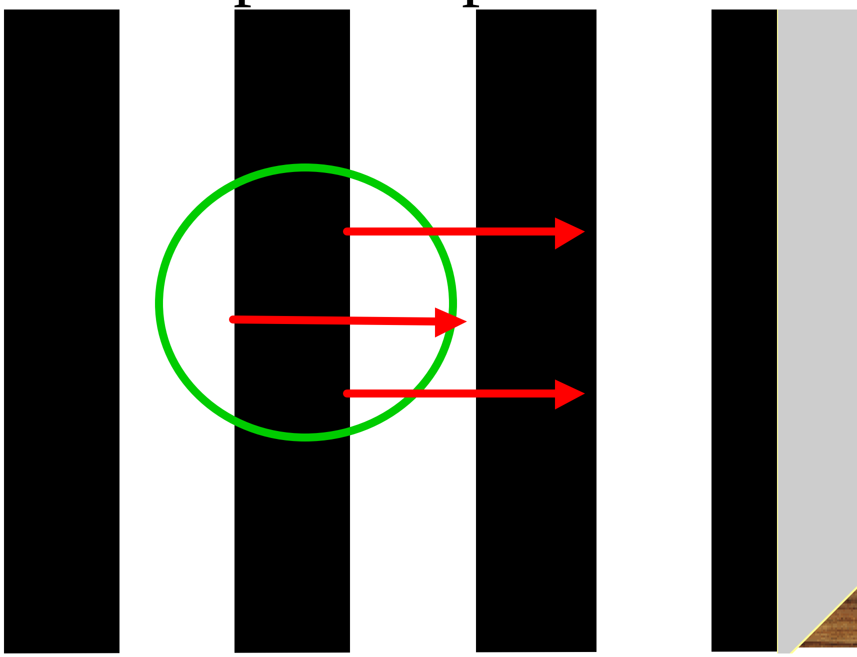
  $\mathbf{x}' E\mathbf{x} = \mathbf{a}^T e$

  with $a = (x'x, x'y, x', y'x, y'y, y', x, y, 1)$

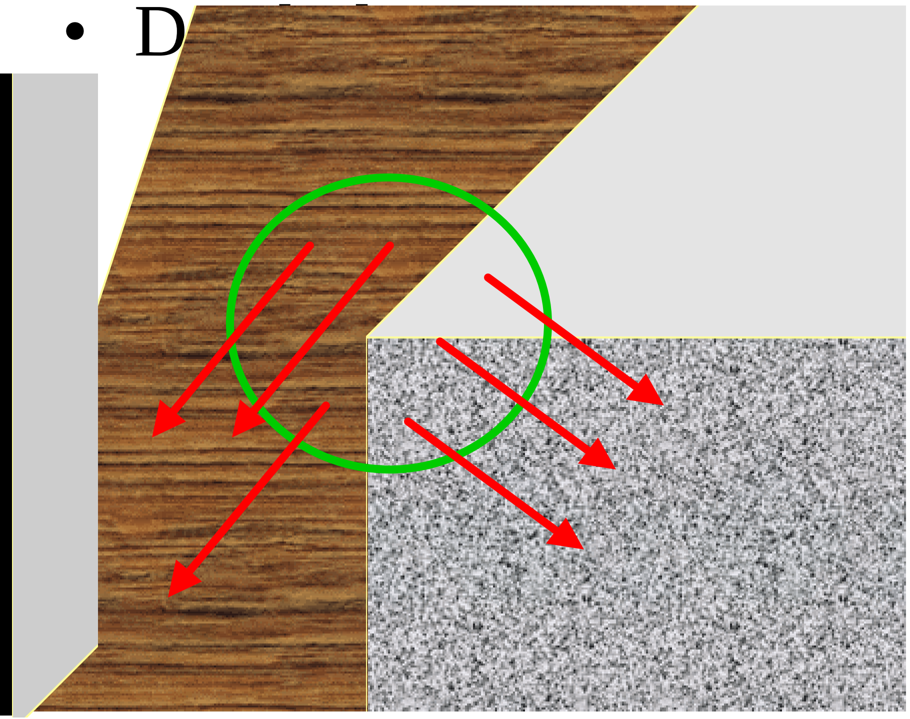
  1. LS minimization $\sum (a_i^T e)^2 \rightarrow \min$ solve for E.
  2. Obtain from E translation and rotation using SVD.
- Prazdny (1981), Burger Bhanu (1990), Nelson Aloimonos (1988), Heeger Jepson (1992): Decomposition of flow field into translational and rotational components. Translational flow field has a certain structure: All vectors are emanating from a point. Either search in the space of rotations or the space of translational directions.
- Longuet-Higgins Prazdny (1980), Waxman (1987): Parametric model for local surface patches ⌫ solve locally for motion parameters and structure

Optical flow difficulties

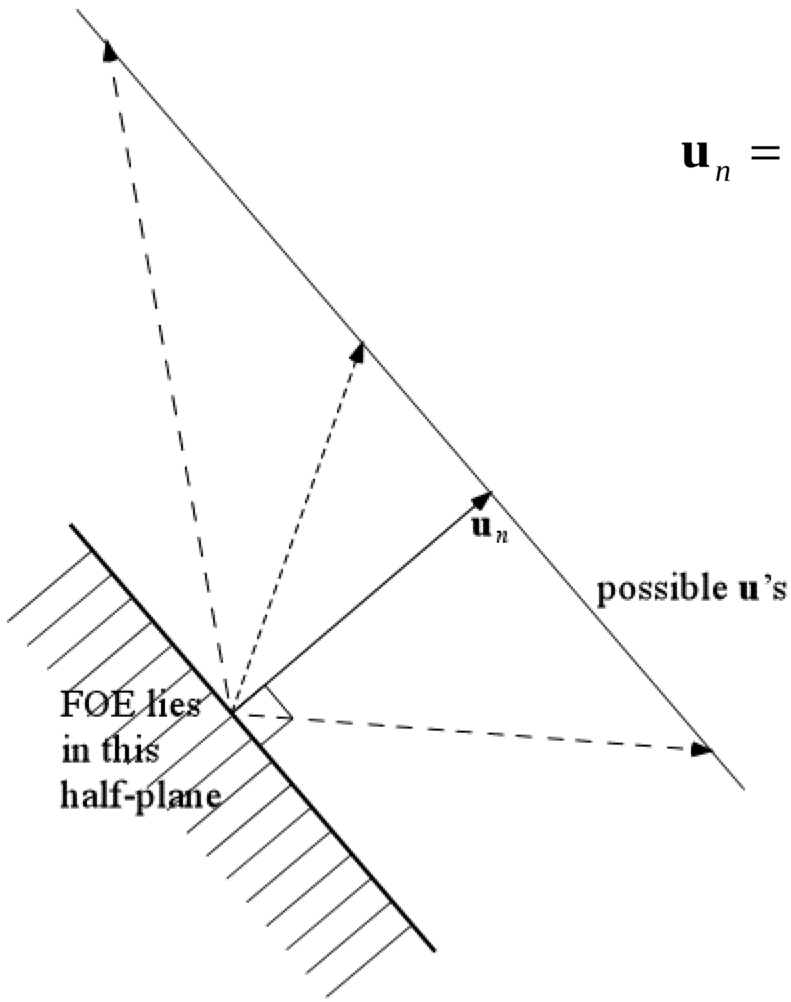
- The aperture problem



- D



Translational Normal Flow



$$\mathbf{u}_n = \frac{\mathbf{u}_{\text{tr}}}{Z} \cdot \mathbf{n}$$

- In the case of translation each normal flow vector constrains the location of the FOE to a half-plane.
- Intersection of half-planes provides FOE.

Egoestimation from normal flow

- patterns defined on the sign of normal flow along particular orientation fields

- positive depth constraint

- 2 classes of orientation fields: copoint vectors and coaxis vectors

Copoint vector fields

Copoint vectors: $\mathbf{v}_{cp}(\mathbf{t}_i)$ perpendicular to translational flow field defined by $\mathbf{t}_\mathrm{i}$

$$\mathbf{v}_{\mathrm{cp}}(\mathbf{t}_i) = \hat{\mathbf{z}} \times \mathbf{u}_{\mathrm{tr}}(\mathbf{t}_i) = \hat{\mathbf{z}} \times (\hat{\mathbf{z}} \times (\mathbf{t}_i \times \mathbf{r}))$$

The components of flow along $\mathbf{v}_{\mathrm{cp}}(\mathbf{t}_i)$ amount to

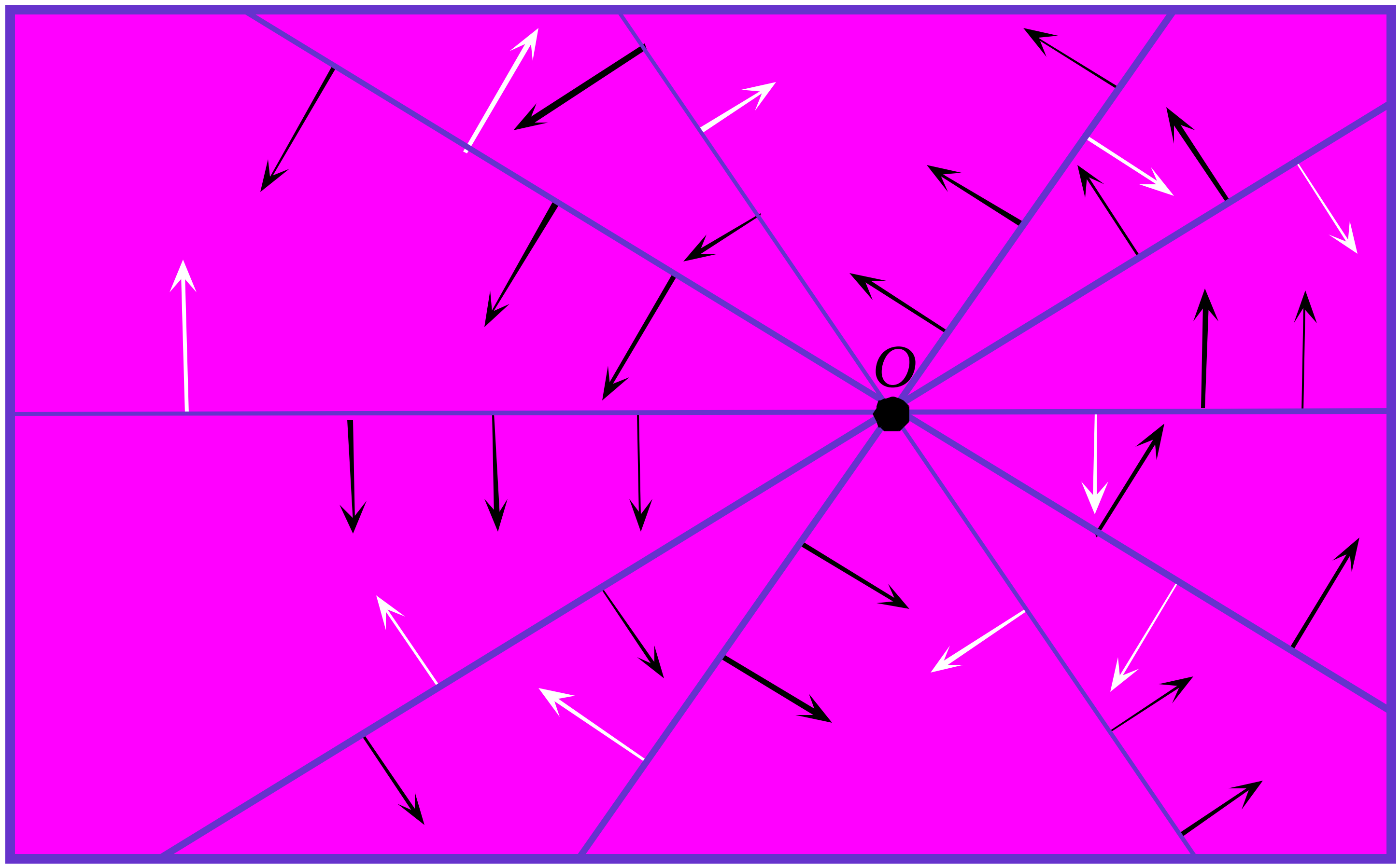
$$\dot{\mathbf{r}} \cdot \frac{\mathbf{v}_{\mathrm{cp}}}{\left\|\mathbf{v}_{\mathrm{cp}}\right\|} = \frac{1}{\left\|\mathbf{v}_{\mathrm{cp}}\right\|}\left(\frac{1}{Z}(\mathbf{t} \times \mathbf{t}_i) \cdot \mathbf{r} + (\mathrm{w} \times \mathbf{r}) \cdot (\mathbf{t}_i \times \mathbf{r})\right)$$

Thus the translational component is separated by a line into positive and negative values $(\mathbf{t} \times \mathbf{t}_i) \cdot \mathbf{r} = 0$

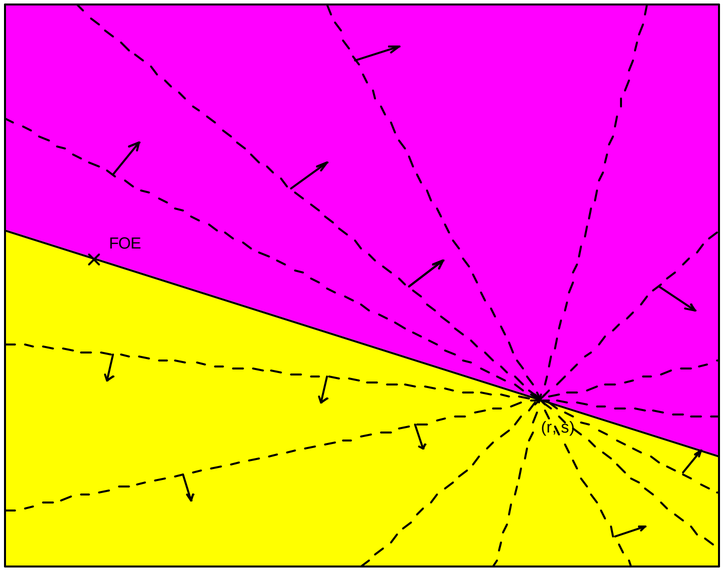
The rotational component is separated by a second-order curve $(\mathrm{w} \times \mathbf{r}) \cdot (\mathbf{t}_i \times \mathbf{r}) = 0$ into positive and negative values

Pattern with positive areas, negative areas, and some undefined areas

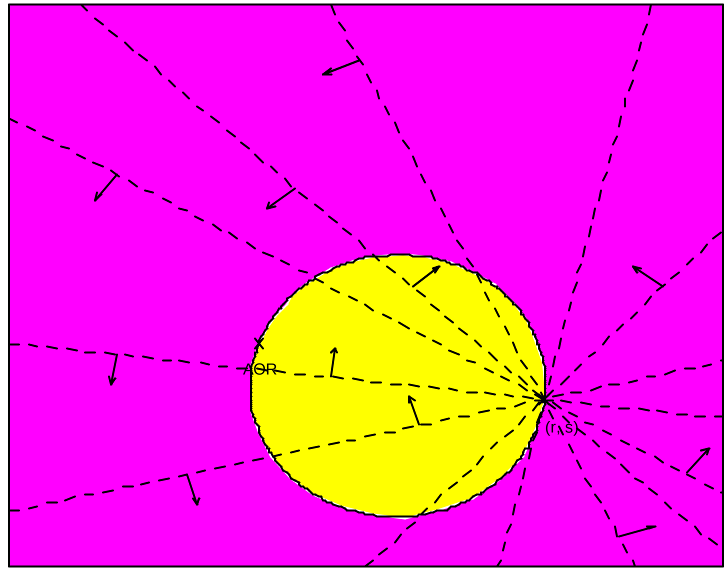
copoint vectors

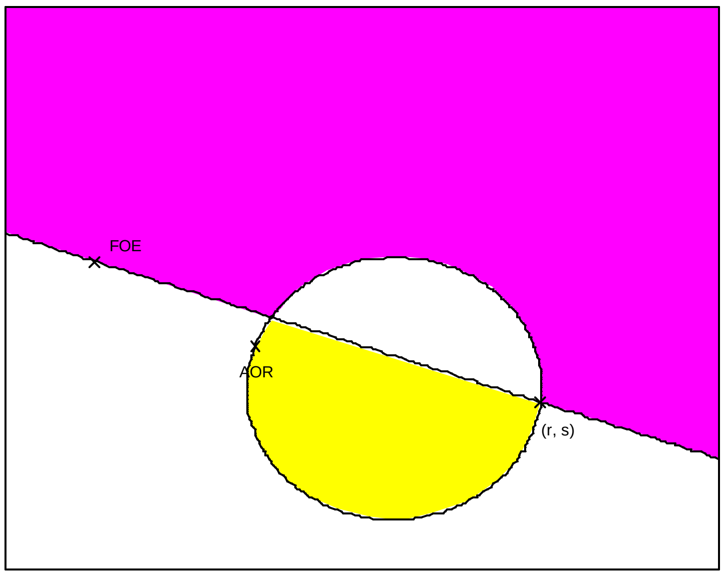


translational component



rotational component





Coaxis vector fields

Coaxis vectors: $\mathbf{v}_{\text{ca}}(\text{w}_?)$ perpendicular to rotation

$$\mathbf{v}_{\text{ca}}(\text{w}_?) = \hat{\mathbf{z}} \times \mathbf{u}_{\text{rot}}(\text{w}_?) = \hat{\mathbf{z}} \times (\times \mathbf{r} \times (\text{w}_? \times \mathbf{r}))$$

The components of flow along $\mathbf{v}_{\text{ca}}(\text{w}_i)$ amount to

$$\dot{\mathbf{r}} \cdot \frac{\mathbf{v}_{\text{ca}}}{\|\mathbf{v}_{\text{ca}}\|} = \frac{1}{\|\mathbf{v}_{\text{ca}}\|}\left((\text{w} \times \text{w}_i) \cdot \mathbf{r} + \frac{1}{Z}(\mathbf{t} \times \mathbf{r}) \cdot (\text{w}_i \times \mathbf{r}) \right)$$

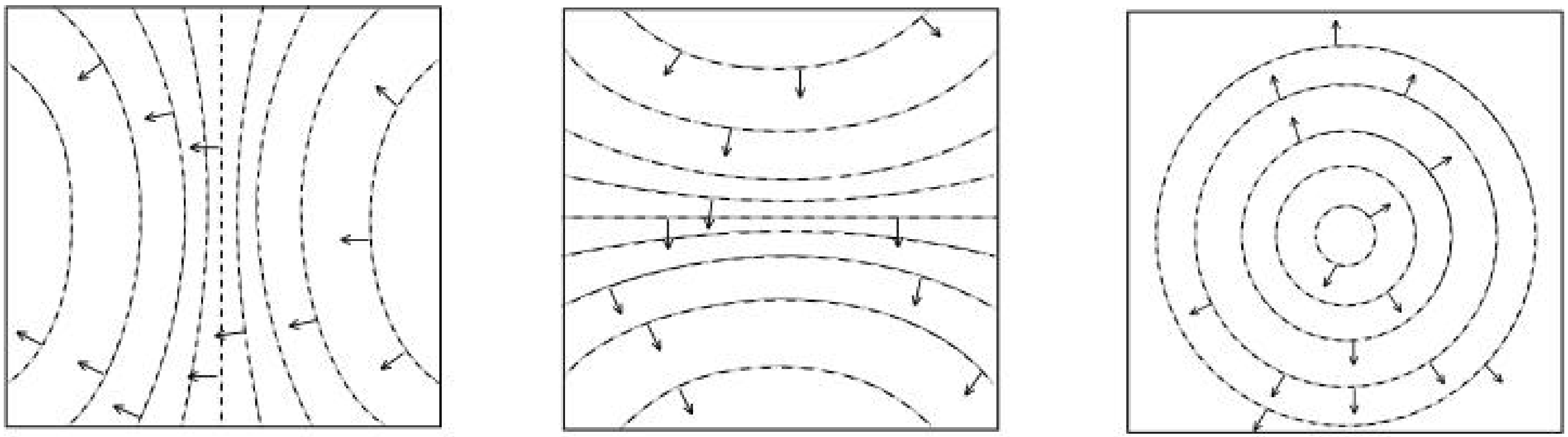
Thus the translational component is separated by a second-order curve

$$(\mathbf{t} \times \mathbf{r}) \cdot (\text{w}_? \times \mathbf{r}) = 0$$

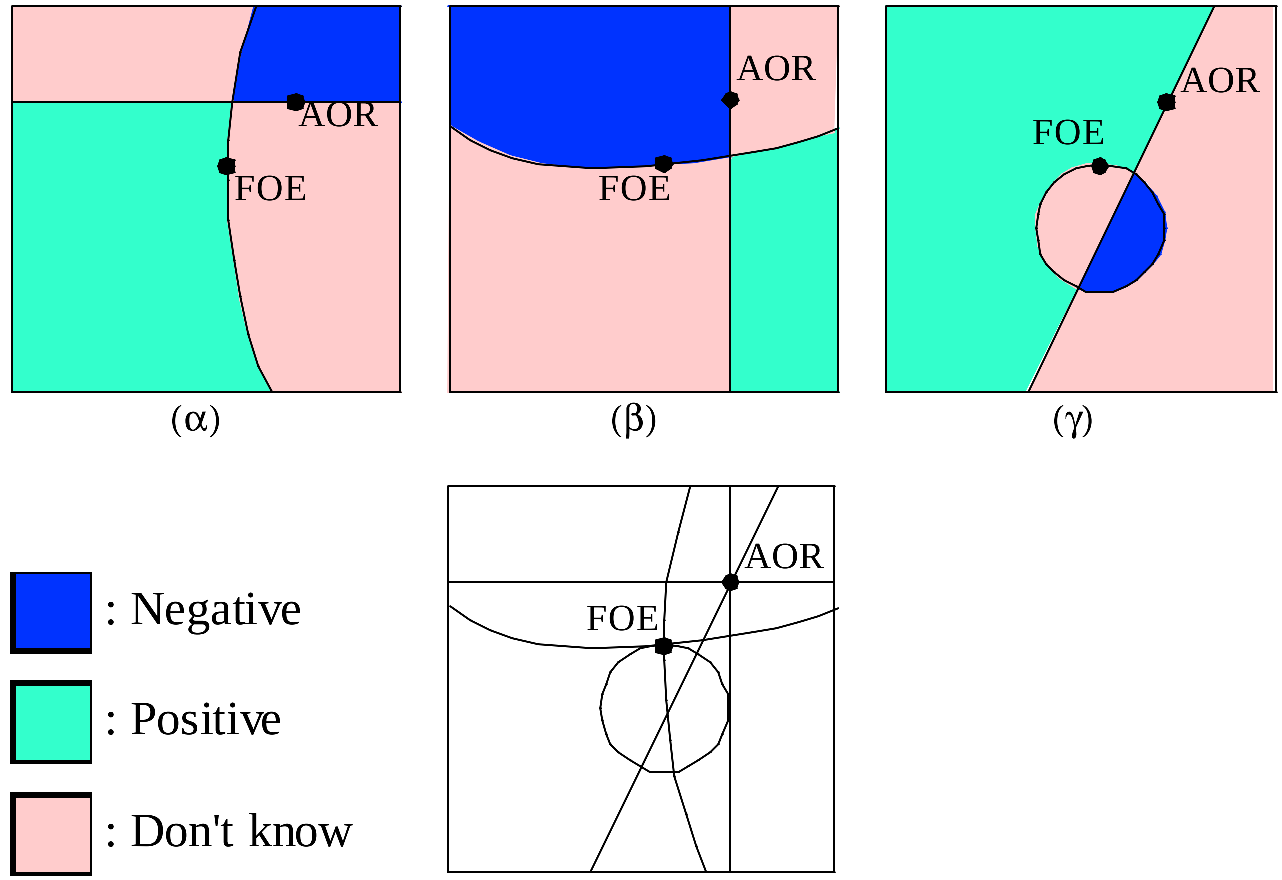
and the rotational component is separated by a line $(\text{w} \times \text{w}_?) \cdot \mathbf{r} = 0$

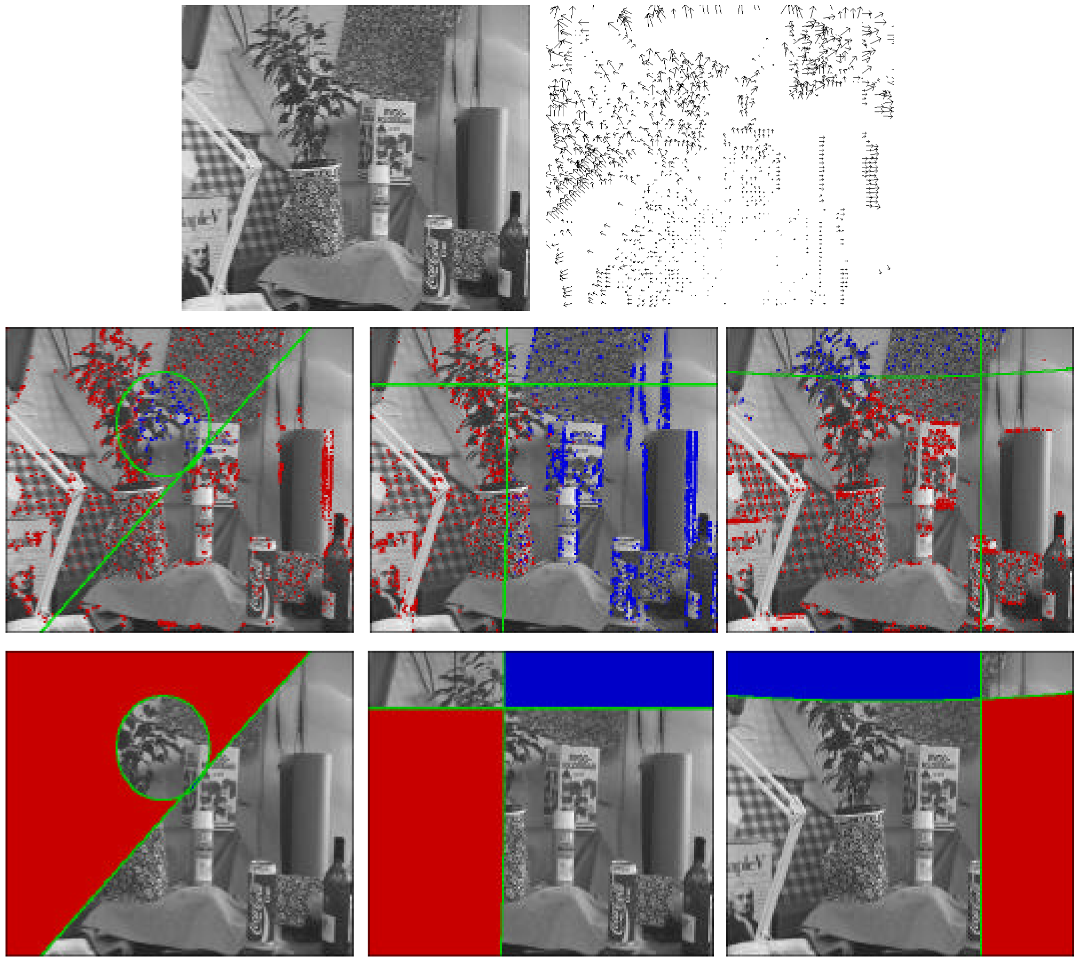
Intersection of patterns provides the FOE.

Three coaxis vector fields



(a) (b) (c)


AOR
FOE
(α)
AOR
FOE
(β)
AOR
FOE
(γ)
: Negative
: Positive
: Don't know
AOR
FOE



Depth variability constraint

- Errors in motion estimates lead to distortion of the scene estimates.
- The distortion is such that the correct motion gives the “smoothest” (least varying) scene structure.

Depth estimation

- Scene depth can be estimated from normal flow measurements:

$$u_{\mathrm{n}} = \mathbf{u} \cdot \mathbf{n} = \frac{1}{Z} \mathbf{u}_{\mathrm{tr}} \cdot \mathbf{n} + \mathbf{u}_{\mathrm{rot}} \cdot \mathbf{n}$$

$$\frac{1}{\hat{Z}} = \frac{u_{\mathrm{n}} - \mathbf{u}_{\mathrm{rot}}\left(\hat{?}\right) \cdot \mathbf{n}}{\mathbf{u}_{\mathrm{tr}}\left(\hat{\mathbf{t}}\right) \cdot \mathbf{n}}$$

Visual Space Distortion

$$\hat{Z} = Z \cdot D, \quad D = \frac{\mathbf{u}_{\mathrm{tr}}\left(\hat{\mathbf{t}}\right) \cdot \mathbf{n}}{\left[\mathbf{u}_{\mathrm{tr}}\left(\mathbf{t}\right) - \mathbf{u}_{\mathrm{rot}}\left(\mathrm{d?}\right)\right] \cdot \mathbf{n}}$$

- Wrong 3D motion gives rise to a rugged (unsmooth) depth function (surface).
- The correct 3D motion leads to the "smoothest" estimated depth.

The error function

- A normal flow measurement:

$$u_{\mathrm{n}} = \frac{1}{Z}\mathbf{u}_{\mathrm{tr}} \cdot \mathbf{n} + \mathbf{u}_{\mathrm{rot}} \cdot \mathbf{n}$$

- The error function to be minimized:

$$\Theta = \sum_{R}\sum_{i} W_i \left(\hat{u}_{\mathrm{n}} - u_{\mathrm{n}}\right)^2$$

- Global parameters: $\hat{\mathbf{t}}, \hat{?}$
- Local parameter: $\hat{Z}$

Error function evaluation

- Given a translation candidate $\hat{\mathbf{t}}$, each local depth can be computed as a linear function of the rotation $\hat{?}$.
- We obtain a second order function of the rotation; its minimization provides both the rotation and the value of the error function.

Handling depth discontinuities

- Given a candidate motion, the scene depth can be estimated and further processed to find depth discontinuities.
- Split a region if it corresponds to two depth values separated in space.

The algorithm

- Compute spatio-temporal image derivatives and normal flow.
- Find the direction of translation that minimizes the depth-variability criterion.
  - Hierarchical search of the 2D space.
  - Iterative minimization.
  - Utilize continuity of the solution in time; usually the motion changes slowly over time.

Sources:

- Horn (1986)
- http://www.cfar.umd.edu/ftp/TRs/CVL-Reports-1993/TR3064-Fermuller.ps.gz
  http://www.cfar.umd.edu/ftp/Trs/CVL-Reports-1995/TR3484-Fermuller.ps.gz (patterns on normal flow)
- http://www.cfar.umd.edu/ftp/TRs/CVL-Reports-1999/TR4000-brodsky.ps.gz (depth variability constraint)