Motion and Flow II
Structure from Motion

\[ u = u_{tr} + u_{rot} \]

\[ u_{tr} = \frac{1}{Z} (\hat{z} \times (t \times r)) \]

\[ u_{rot} = \frac{1}{F} (\hat{z} \times (r \times ([?] \times r))) \]
Passive Navigation and Structure

The system moves with a rigid motion with translational velocity 
\[ \mathbf{t} = (U, V, W)^T \] and rotational velocity \( \omega = (\alpha, \beta, \gamma)^T \).

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
\[
\dot{R} = -t - \omega \times R
\]
\[
\dot{X} = -U - BZ + \gamma Y
\]
\[
\dot{Y} = -V - \gamma X + \alpha Z
\]
\[
\dot{Z} = -W - \alpha X + \beta X
\]

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

\[
u = -U + xW + \alpha xy - \beta (x^2 + 1) + \gamma y = \frac{u_{tr}}{Z} + u_{rot}
\]
\[
v = -V + yW + \alpha (y^2 + 1) - \beta xy - \gamma x = \frac{v_{tr}}{Z} + v_{rot}
\]
in vector notation: \( \mathbf{r} = \frac{1}{Z} \mathbf{R} \cdot \mathbf{z}_0 \), where \( Z = \mathbf{R} \cdot \mathbf{z}_0 \)

\[
\dot{r} = \frac{1}{\mathbf{R} \cdot \mathbf{z}_0} (\mathbf{z}_0 \times (t \times \mathbf{r})) + \mathbf{z}_0 \times (\mathbf{r} \times (\omega \times \mathbf{r}))
\]

Scaling ambiguity: We can compute the translation only up to a scale factor \((Kt, KZ)\) give the same flow as \((t, Z)\).
\[
\mathbf{u}_{tr} = \left( (x-x_0)\frac{W}{Z}, (y-y_0)\frac{W}{Z} \right)
\]

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).

\[
\left( \frac{\alpha}{\gamma} f, \frac{\beta}{\gamma} f \right)\] 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 \( t_1, t_2 \) and two surfaces \( Z_1, Z_2 \)

\[
\begin{align*}
\mathbf{t}_1 &= (U_1, V_1, W_1) \\
\mathbf{t}_2 &= (U_2, V_2, W_2) \\
\mathbf{u} &= \frac{-U_1 + xW_1}{Z_1} \\
\mathbf{v} &= \frac{-V_1 + yW_1}{Z_1} \\
\mathbf{u} &= \frac{-U_2 + xW_2}{Z} \\
\mathbf{v} &= \frac{-V_2 + yW_2}{Z} \\
(-U_1 + xW_1)(-V_2 + yW_2) &= (-U_2 + xW_2)(-V_1 + yW_1) \\
U_1V_2 - xV_2W_1 - yU_1W_2 + xyW_1W_2 &= U_2V_1 - xV_1W_2 - yU_2W_1 + xyW_2W_1
\end{align*}
\]

must hold for all \( x \) and \( y \)

\[
\begin{align*}
U_1V_2 &= U_2V_1 \\
V_2W_1 &= V_1W_2 \rightarrow U_1 : V_1 : W_1 = U_2 : V_2 : W_2 \rightarrow \mathbf{t}_2 = k\mathbf{t}_1 \text{ and } Z_2 = kZ_1 \\
U_1W_2 &= U_2W_1
\end{align*}
\]

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

A least squares formulation
\[
\iint \left( u - \frac{xW - U}{Z} \right)^2 + \left( v - \frac{yW - V}{Z} \right)^2 \, dx \, dy \to \min
\]

Substitute \( a = -U + xW \quad b = -V + yW \)
\[
\iint \left( u - \frac{a}{Z} \right)^2 + \left( v - \frac{b}{Z} \right)^2 \to \min
\]

Step 1: Minimize with respect to \( Z \). (Find the length of \( \mathbf{u}_{tr} \) for which \( d \) would be minimized.)
\[
\left( u - \frac{a}{Z} \right) \frac{a}{Z^2} + \left( v - \frac{b}{Z} \right) \frac{b}{Z^2} = 0 \quad \quad Z = \frac{a^2 + b^2}{ua + vb}
\]

Substitute back
\[
\iint \frac{(ub - va)^2}{a^2 + b^2} \, dx \, dy \to \min
\]
Step 2: Differentiate with respect to $U$, $V$, $W$, set expression to zero.

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

\[ \int \int (-V + yW)K \, dx \, dy = 0 \]
\[ \int \int (-U + xW)K \, dx \, dy = 0 \]
\[ \int \int (-yU + xV)K \, dx \, dy = 0 \]

3 linearly dependent equations $(U \cdot I + V \cdot II + W \cdot III = 0)$
The Rotational Case

\[
\iint (u - u_{\text{rot}})^2 + (v - v_{\text{rot}})^2 \to \min
\]

\[
u - \alpha xy + \beta (x^2 + 1) - \gamma y = 0
\]

\[
v - \alpha (y^2 + 1) + \beta xy + \gamma x = 0
\]

\[
\begin{pmatrix}
xy & -(x^2 + 1) & y \\
(y^2 + 1) & -xy & -x
\end{pmatrix}
\begin{pmatrix}
\alpha \\
\beta \\
\gamma
\end{pmatrix} =
\begin{pmatrix}
u \\
v
\end{pmatrix}
\]

In matrix form

\[
A \cdot \omega = u
\]

\[
\omega = (A^T A)^{-1} A^T u
\]
The General Case

Minimization of epipolar distance

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

or, in vector notation

\[
\int ((t \times r)(\dot{r} - \omega \times r))^2 \, dr \rightarrow \min
\]
Motion Estimation Techniques

- Linearization (Tsai Huang 1984, Longuet-Higgins 1981) in the discrete case
  \[ x'Ex = 0, \text{ where } E = T_x R \]
  \[ x'Ex = a^T e \]
  with \( a = (x', y', x, y, y, x, y, 1) \)
  1. LS minimization \( \sum (a_i^T e)^2 \rightarrow \text{min solve for } E. \)
  2. Obtain from \( E \) translation and rotation using SVD.


Optical flow difficulties

- The aperture problem
- Depth discontinuities
Translational Normal Flow

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.

\[ u_n = \frac{u_{tr} \cdot n}{Z} \]
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: $\mathbf{v}_{cp}(t_i)$ perpendicular to translational flow field defined by $t_i$

$$v_{cp}(t_i) = \hat{z} \times u_{tr}(t_i) = \hat{z} \times (\hat{z} \times (t_i \times r))$$

The components of flow along $v_{cp}(t_i)$ amount to

$$\hat{\mathbf{r}} \cdot \frac{\mathbf{v}_{cp}}{\|v_{cp}\|} = \frac{1}{\|v_{cp}\|} \left( \frac{1}{Z} (t \times t_i) \cdot \mathbf{r} + (\omega \times \mathbf{r}) \cdot (t_i \times \mathbf{r}) \right)$$

Thus the translational component is separated by a line into positive and negative values

$$(t \times t_i) \cdot \mathbf{r} = 0$$

The rotational component is separated by a second-order curve into positive and negative values

$$(\omega \times \mathbf{r}) \cdot (t_i \times \mathbf{r}) = 0$$

Pattern with positive areas, negative areas, and some undefined areas
copoint vectors
translational component

rotational component
Coaxis vector fields

Coaxis vectors: $v_{ca}(\omega_r)$ perpendicular to rotation

$$v_{ca}(\omega_r) = \hat{z} \times u_{rot}(\omega_r) = \hat{z} \times (r \times (\omega_r \times r))$$

The components of flow along $v_{ca}(\omega_i)$ amount to

$$\dot{r} \cdot \frac{v_{ca}}{\|v_{ca}\|} = \frac{1}{\|v_{ca}\|} \left( (\omega \times \omega_i) \cdot r + \frac{1}{Z} (t \times r) \cdot (\omega_i \times r) \right)$$

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

$$(t \times r) \cdot (\omega_r \times r) = 0$$

and the rotational component is separated by a line

$$(\omega \times \omega_r) \cdot r = 0$$

Intersection of patterns provides the FOE.
Three coaxis vector fields

(a) (b) (c)
: Negative
: Positive
: Don't know
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:

\[ u_n = u \cdot n = \frac{1}{Z} u_{tr} \cdot n + u_{rot} \cdot n \]

\[ \hat{Z} = \frac{u_n - u_{rot} \left( \hat{t} \right) \cdot n}{u_{tr} \left( \hat{t} \right) \cdot n} \]
Visual Space Distortion

\[ \hat{Z} = Z \cdot D, \quad D = \frac{u_{tr}(\hat{t}) \cdot n}{[u_{tr}(t) - u_{rot}(d?)] \cdot n} \]

- 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:
  \[ u_n = \frac{1}{Z} u_{tr} \cdot n + u_{rot} \cdot n \]

- The error function to be minimized:
  \[ \Theta = \sum_{R} \sum_{i} W_i (\hat{u}_n - u_n)^2 \]

- Global parameters: \( \hat{t}, \hat{\tau} \)
- Local parameter: \( \hat{Z} \)
Error function evaluation

• Given a translation candidate \( \hat{t} \), each local depth can be computed as a linear function of the rotation \( \hat{\theta} \).

• 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)