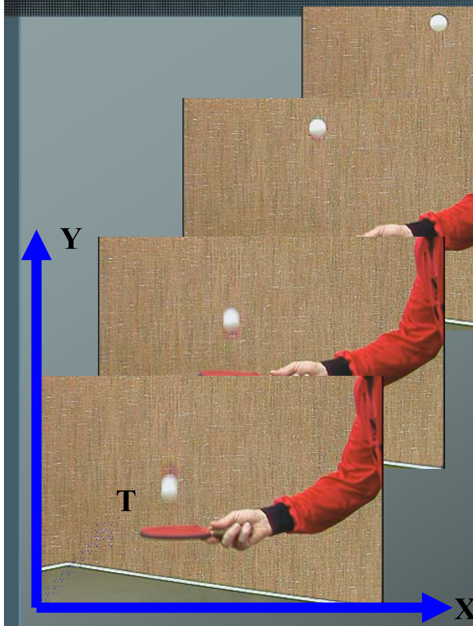
7. Motion

Computer Vision

Zoltan Kato

http://www.inf.u-szeged.hu/~kato/

Visual motion (=video)



Terminology

- **Scene flow**: 3-D velocities of scene points: Derivative of rigid transformation between views with respect to time
- **Motion field**: 2-D projection of scene flow
- **Optical flow**: Approximation of motion field derived from apparent motion of image points

Motion Analysis Problems

- **Correspondence**: Which elements of a frame correspond to which elements in the next frame?
- **Reconstruction** :Given a number of correspondences, and possibly the knowledge of the camera's intrinsic parameters, how to recovery the 3D motion and structure of the observed world
- Other problems:
  - **Motion Segmentation**: what are the regions the the image plane corresponding to different moving objects?
  - **Motion Understanding**: lip reading, gesture, expression, event…
- Main Difference between Motion and Stereo
  - Correspondence: the disparities between consecutive frames are much smaller due to dense temporal sampling
  - Reconstruction: the visual motion could be caused by multiple motions (instead of a single 3D rigid transformation)

The Motion Field of Rigid Objects

- Motion:
  - **3D Motion (R,T):**
    - camera motion (static scene)
    - or scene (object) motion
    - Only one rigid, relative motion between the camera and the scene (object)
  - **Image motion field:**
    - 2D vector field of velocities of the image points induced by the relative motion.
- Data: Image sequence
  - **Many frames**
    - captured at time t=0, 1, 2, …
  - **Basics: only consider two consecutive frames**
    - We consider a reference frame and its consecutive frame
  - **Image motion field**
    - can be viewed disparity map of the two frames captured at two consecutive camera locations (assuming we have a moving camera)

Motion Field (MF)

- MF assigns a velocity vector to each pixel in the image.
- These velocities are INDUCED by the RELATIVE MOTION between the camera and the 3D scene
- MF can be thought as the *projection* of the 3D velocities on the image plane.

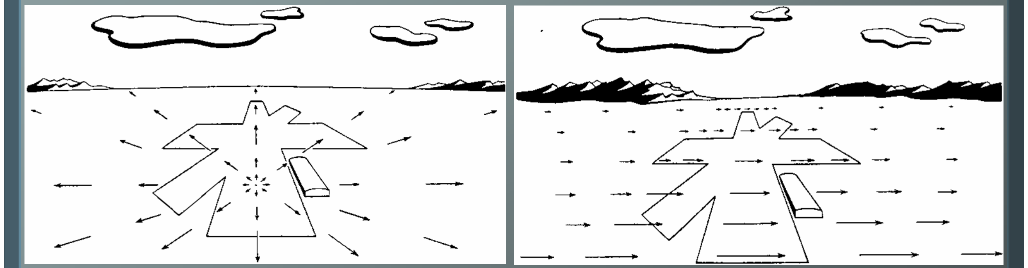
The Information from Image Motion

- Structure from Motion
  - Apparent motion is a strong visual clue for 3D reconstruction
    - 3D motion between observer and scene + structure of the scene
  - Wallach O'Connell (1953): Kinetic depth effect

    http://www.biols.susx.ac.uk/home/George_Mather/Motion/KDE.HTML
    - Motion parallax: two static points close by in the image with different image motion; the larger translational motion corresponds to the point closer by (smaller depth)
- Recognition by motion only
  - Biological visual systems use visual motion to infer properties of 3D world with little a priori knowledge of it.
    - Johansson (1975): Light bulbs on joints

      http://www.biols.susx.ac.uk/home/George_Mather/Motion/index.html
- Visual Motion = Video !
  - Surveillance (Human Tracking and Traffic Monitoring)
  - Video Coding and Compression: MPEG 1, 2, 4, 7…
  - HCI using Human Gesture (video camera)
  - …

Examples of Motion Fields I



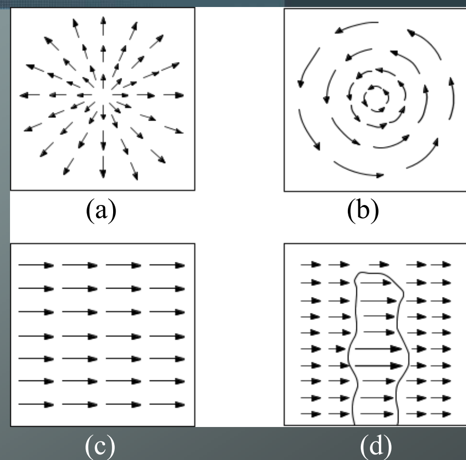
(a) (b)

(a) Motion field of a pilot looking straight ahead while approaching a fixed point on a landing strip.

(b) Pilot is looking to the right in level flight.

Examples of Motion Fields II

- (a) Translation perpendicular to a surface.
- (b) Rotation about axis perpendicular to image plane.
- (c) Translation parallel to a surface at a constant distance.
- (d) Translation parallel to an obstacle in front of a more distant background.

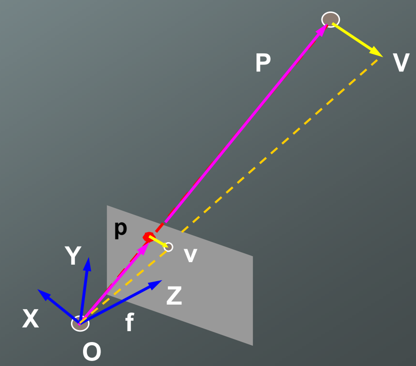


The Motion Field of Rigid Objects

- Notations
  - $P = (X,Y,Z)^T$: 3-D point in the camera reference frame
  - $p = (x,y,f)^T$: the projection of the scene point in the pinhole camera
- Relative motion between P and the camera
  - $T= (T_x,T_y,T_z)^T$: translation component of the motion
  - $\omega=(\omega_x, \omega_y,\omega_z)^T$: the angular velocity
- How to connect this with stereo geometry (with R, T)?
- Image velocity v= ?

$$\mathbf{p} = \frac{f}{Z}\mathbf{P}$$

$$\mathbf{V} = -\mathbf{T} - \boldsymbol{\omega} \times \mathbf{P}$$



The Motion Field of Rigid Objects

- Notations
  - $P = (X,Y,Z)^T$: 3-D point in the camera reference frame
  - $p = (x,y,f)^T$: the projection of the scene point in the pinhole camera
- Relative motion between P and the camera
  - $T= (T_x,T_y,T_z)^T$: translation component of the motion
  - $\omega=(\omega_x, \omega_y,\omega_z)^T$: the angular velocity
- How to connect this with stereo geometry (with R, T)?

$$\mathbf{V} = -\mathbf{T} - \boldsymbol{\omega} \times \mathbf{P}$$

$$\mathbf{P} - \mathbf{P}' = \mathbf{V} = -\mathbf{T} - \begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix} \mathbf{P}$$

$$\mathbf{P}' = \begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix} \mathbf{P} + \mathbf{T}$$

$$\mathbf{R} = \begin{bmatrix} \cos\beta\cos\gamma & -\cos\beta\sin\gamma & \sin\beta \\ \sin\alpha\sin\beta\cos\gamma + \cos\alpha\sin\gamma & -\sin\alpha\sin\beta\sin\gamma + \cos\alpha\cos\gamma & -\sin\alpha\cos\beta \\ -\cos\alpha\sin\beta\cos\gamma + \sin\alpha\sin\gamma & \cos\alpha\sin\beta\sin\gamma + \sin\alpha\cos\gamma & \cos\alpha\cos\gamma \end{bmatrix}$$

Basic Equations of Motion Field

- Take the time derivative of both sides of the projection equation
  - The motion field is the sum of two components
    - Translational part
    - Rotational part
  - Assume known intrinsic parameters

$$\mathbf{v} = \frac{f}{Z^2}(Z\mathbf{V} - V_z\mathbf{P})$$

$$\mathbf{V} = -\mathbf{T} - \boldsymbol{\omega} \times \mathbf{P} \qquad \mathbf{p} = \frac{f}{Z}\mathbf{P}$$

$$\begin{pmatrix} v_x \\ v_y \end{pmatrix} = \frac{1}{f}\begin{pmatrix} xy & -(x^2+f^2) & fy \\ y^2+f^2 & -xy & -fx \end{pmatrix}\begin{pmatrix} \omega_x \\ \omega_y \\ \omega_z \end{pmatrix} + \frac{1}{Z}\begin{pmatrix} -f & 0 & x \\ 0 & -f & y \end{pmatrix}\begin{pmatrix} T_x \\ T_y \\ T_z \end{pmatrix}$$

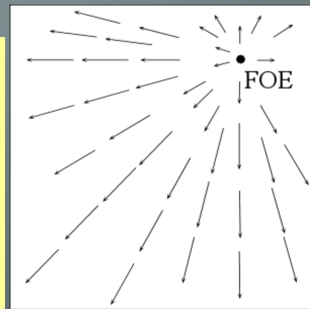
Rotation part: no depth information — Translation part: depth Z

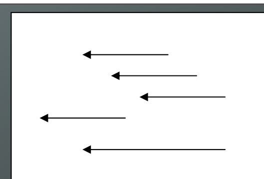
Translational flow field (ω=0)

$$\mathbf{p}_o = (x_0, y_0) = \left(\frac{T_x}{T_z} \cdot f, \frac{T_y}{T_z} \cdot f \right)$$

$$T_z \neq 0: \quad u_{tr} = (x - x_0)\frac{T_z}{Z}, \quad v_{tr} = (y - y_0)\frac{T_z}{Z}$$

$$T_z = 0: \quad u_{tr} = -f\frac{T_x}{Z}, \quad v_{tr} = -f\frac{T_y}{Z}$$





- Where $\mathbf{p_0}$ is the
  - focus of expansion (FOE) if $\mathbf{T_z<0}$
  - focus of concentration (FOC) if $\mathbf{T_z>0}$
- When $\mathbf{T_z=0}$ ➔ All motion field vectors are parallel to each other and inversely proportional to depth !

Translational flow field (ω=0)

- Pure Translation (ω =0)
- Radial Motion Field ($\mathbf{T_z \neq 0}$)
  - Vanishing point $\mathbf{p_0 = (x_0, y_0)^T}$:
    - motion direction
  - FOE (focus of expansion)
    - Vectors away from p0 if $\mathbf{T_z<0}$
  - FOC (focus of contraction)
    - Vectors towards p0 if $\mathbf{T_z>0}$
  - Depth estimation
    - depth inversely proportional to magnitude of motion vector v, and also proportional to distance from p to p_0
- Parallel Motion Field ($\mathbf{T_z=0}$)
  - Depth estimation:
    - depth inversely proportional to magnitude of motion vector v

$$\begin{pmatrix} v_x \\ v_y \end{pmatrix} = \frac{1}{Z}\begin{pmatrix} -f & 0 & x \\ 0 & -f & y \end{pmatrix}\begin{pmatrix} T_x \\ T_y \\ T_z \end{pmatrix}$$

$$\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \frac{f}{T_z}\begin{pmatrix} T_x \\ T_y \end{pmatrix}$$

$$\begin{pmatrix} v_x \\ v_y \end{pmatrix} = \frac{T_z}{Z}\begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix}$$

$$Z = \frac{T_z}{|\mathbf{v}|}\sqrt{(x - x_0)^2 + (y - y_0)^2}$$

$T_z=0$

$$\begin{pmatrix} v_x \\ v_y \end{pmatrix} = -\frac{f}{Z}\begin{pmatrix} T_x \\ T_y \end{pmatrix}$$

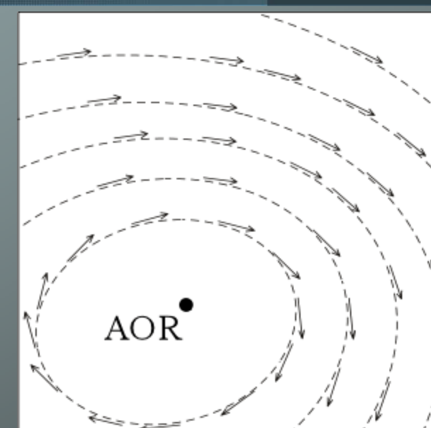
$$Z = \frac{f}{|\mathbf{v}|}\sqrt{T_x^2 + T_y^2}$$

Pure Translation: Properties of the MF

- $\mathbf{p_0}$ (FOE) is
  - the vanishing point of the direction of translation.
  - the intersection of the ray parallel to the translation vector and the image plane.
- $\mathbf{T_z=0}$ ➔
  - MF is **PARALLEL**.
  - length of the MF vectors is inversely proportional to depth $\mathbf{Z}$.
- $\mathbf{T_z \neq 0}$ ➔
  - MF is **RADIAL** with all vectors pointing towards (or away from) a single point $\mathbf{p_0}$.
  - length of the MF vectors is inversely proportional to depth $\mathbf{Z}$.
  - length is also directly proportional to the distance between $\mathbf{p}$ and $\mathbf{p_0}$.

Rotational flow field (T=0)

$$\text{AOR} = \left(\frac{\omega_x}{\omega_z} f, \frac{\omega_y}{\omega_z} f \right)$$



- AOR is the point where the rotation axis pierces the image plane.
- rotational flow field is quadratic in image coordinates.

Rotational flow field (T=0)

- Pure Rotation (**T=0**)
  - **Does not carry 3D information**
- Motion Field (approximation)
  - Small motion
  - A quadratic polynomial in image coordinates **$(x,y,f)^T$**
- Image Transformation between two frames (accurate)
  - Motion can be large
  - Homography (3x3 matrix) for all points
- Image mosaicing from a rotating camera
  - 360 degree panorama

$$\begin{pmatrix} v_x \\ v_y \end{pmatrix} = \frac{1}{f}\begin{pmatrix} xy & -(x^2+f^2) & fy \\ y^2+f^2 & -xy & -fx \end{pmatrix}\begin{pmatrix} \omega_x \\ \omega_y \\ \omega_z \end{pmatrix}$$

$$\mathbf{P'} = \mathbf{RP}$$

$$\mathbf{p'} = \frac{f'}{Z'}\mathbf{P'} \qquad \mathbf{p} = \frac{f}{Z}\mathbf{P}$$

$$\mathbf{p'} \cong \mathbf{Rp}$$

Moving Plane

- Planes are common in the man-made world
- Motion Field (approximation)
  - Given small motion

$$\mathbf{n^T P} = d \;\Rightarrow\; \frac{(n_x x + n_y y + n_z f)}{f} Z = d$$

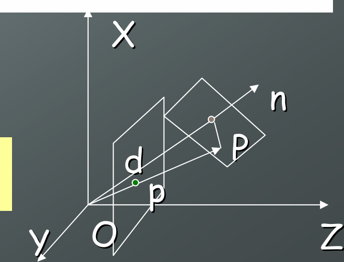
$$\begin{pmatrix} u \\ v \end{pmatrix} = \frac{1}{f}\begin{pmatrix} xy & -(x^2+f^2) & fy \\ y^2+f^2 & -xy & -fx \end{pmatrix}\begin{pmatrix} \omega_x \\ \omega_y \\ \omega_z \end{pmatrix} + \frac{1}{Z}\begin{pmatrix} -f & 0 & x \\ 0 & -f & y \end{pmatrix}\begin{pmatrix} T_x \\ T_y \\ T_z \end{pmatrix}$$

  - The MF vectors are given by a quadratic polynomial in image coord.
  - Their coeffs. a1 to a8 (only 8 !) are functions of n, d, t and w.
  - The same coeffs. can be obtained with a different plane and relative velocity.

$$u = \frac{1}{fd}(a_1x^2 + a_2xy + a_3fx + a_4fy + a_5f^2)$$
$$v = \frac{1}{fd}(a_1xy + a_2y^2 + a_6fy + a_7fx + a_8f^2)$$

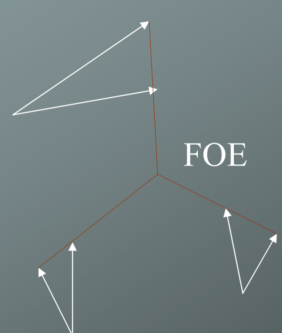
- Image Transformation between two frames (accurate)
  - Any amount of motion (arbitrary)
  - Homography (3x3 matrix) for all points
- Image Mosaicing for a planar scene
  - Aerial image sequence
  - Video of blackboard

$$\mathbf{p'} \cong \mathbf{Ap}$$



Motion Parallax

- The difference in motion between two very close points does not depend on rotation.
- Can be used at depth discontinuities to obtain the direction of translation.



Motion Parallax

1. The relative motion field of two instantaneously coincident points
   - Does not depend on the rotational component of motion
   - Points towards (away from) the vanishing point of the translation direction
2. The motion field of two frames after rotation compensation
   - only includes the translation component
   - points towards (away from) the vanishing point p0 (the instantaneous epipole)
   - the length of each motion vector is inversely proportional to the depth, and also proportional to the distance from point p to the vanishing point p0 of the translation direction
   - Question: how to remove rotation?

Motion Parallax: 1. Relative MF

- At instant t, three pairs of points happen to be coincident
- The difference of the motion vectors of each pair cancels the rotational components
- … and the **relative motion field** points in (towards or away from) the VP of the translational direction

$$\mathbf{u} = (u, v) = \mathbf{u}_{tr} + \mathbf{u}_{rot}$$

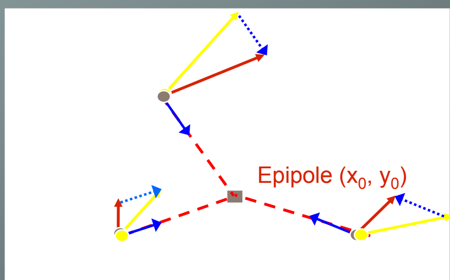
At points p_1 and $p_2 = (x, y)$ we have

$$\mathbf{u}_{1,rot} = \mathbf{u}_{2,rot}$$

$$\Delta u_{tr} = u_{1,tr} - u_{2,tr} = (x - x_o)\left(\frac{1}{Z_1} - \frac{1}{Z_2}\right)$$

$$\Delta v_{tr} = v_{1,tr} - v_{2,tr} = (y - y_o)\left(\frac{1}{Z_1} - \frac{1}{Z_2}\right)$$

$$\frac{\Delta v}{\Delta u} = \frac{y - y_0}{x - x_0}$$



Motion Parallax: 1. Relative MF

- Vector component perpendicular to translational component is only due to rotation ➔
  - rotation can be estimated from it.

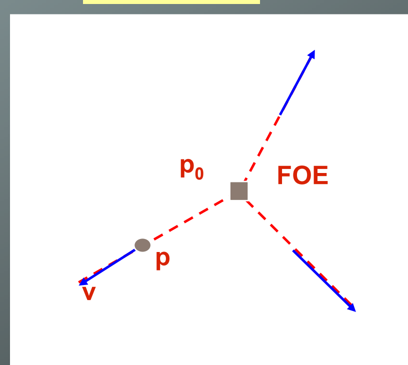
$$\mathbf{u}^{\perp}{}_{tr} = \frac{(y - y_0, x - x_0)}{\|(y - y_0, x - x_0)\|}$$

$$\mathbf{u} \cdot \mathbf{u}^{\perp}{}_{tr} = \frac{1}{\|(y - y_0, x - x_0)\|}(y - y_0)u_{rot} - (x - x_0)v_{rot}$$

Motion Parallax: 2. Rotation Compensation

- Question: how to remove rotation?
  - Rotation compensation can be done by image warping after finding three (3) pairs of coincident points
- After compensation, MF
  - only includes the translation component
  - points towards (away from) the vanishing point $\mathbf{p_0}$ (the **instantaneous epipole**)
  - the length of each motion vector is inversely proportional to the depth,
  - and also proportional to the distance from point $\mathbf{p}$ to the vanishing point $\mathbf{p_0}$ of the translation direction (if $T_z \neq 0$)

$$\frac{v_y^T}{v_x^T} = \frac{y - y_0}{x - x_0}$$



$$|\mathbf{v}| = \frac{T_z}{Z}\sqrt{(x - x_0)^2 + (y - y_0)^2}$$

Motion Estimation Techniques

- Prazdny (1981), Burger Bhanu (1990), Nelson Aloimonos (1988), Heeger Jepson (1992):
  - Decomposition of flow field into translational and rotational components.
    - Translational flow field is radial (all vectors are emanating from (or pouring into) one point),
    - rotational flow field is quadratic in image coordinates.
  - Either search in the space of rotations: remainig flow field should be translational.
    - Translational flow field is evaluated by minimizing deviation from radial field:

$$(-v, u) \cdot (x - x_0, y - y_0) = 0$$

  - Or search in the space of directions of translation:
    - Vectors perpendicular to translation are due to rotation only