

The Computational Problem of Structure from Motion (SFM)

Structure from Motion (SFM)

- People spend a large portion of their time in locomotion. Every time we change the position of our eyes or head, retinal motion is produced.
 - dominant stimulus for normal everyday vision.
- Role for biological vision
 - Find direction of heading (navigation and stabilization)
 - Cues to depth and shape (compare with stereo)
 - Avoid obstacles when moving
 - Find other objects approaching/leaving (hunt/escape)
 - Figure-ground segmentation / object segmentation (similar velocities and /or depths)
 - Controlling eye movements (saccades and smooth pursuit)



Importance of motion

- Motion Blind patient
 - Difficulty to cross roads: - “When I am looking at the car first, it seems far away. But then when I want to cross the road, suddenly it is very near.”
 - Notable: Defect selective to visual motion perception (e.g., the patient could recognize cars)
 - Difficulty to pour tea or coffee, because fluid appears to be frozen
- Applications: Visual Motion = Video !
 - Understanding video content (conveyed via motion cues)
 - Surveillance (Human Tracking and Traffic Monitoring)
 - Virtual and Augmented Reality. Digital Movies.



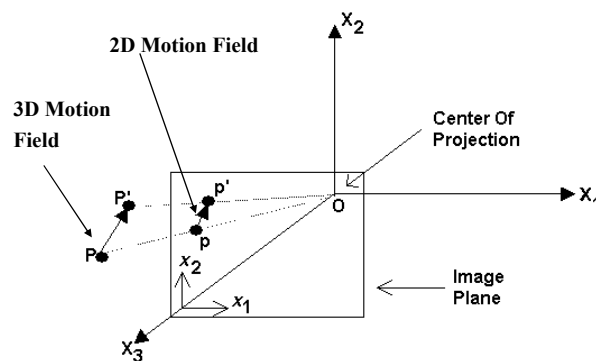
Problem Statement

- In technical terms, SFM amounts to: Given multiple views of a scene, to recover the rigid transformation between any two views and the structure of the imaged scene.
- Formal analyses have shown that purely retinal or image motion contains sufficient information to do so (e.g., Longuet-Higgins & Prazdny, 1980).
- The extraretinal theory assumes that information about the eye movement itself — either from the eye command signal or from proprioceptive feedback — is taken into account
 - not covered in this course.

Scope of SFM

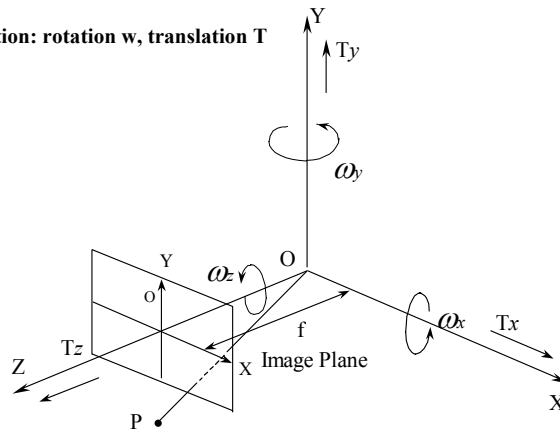
- Two Subproblems
 - Correspondence / 2D motion field: Which elements of a frame correspond to which elements in the next frame?
 - 3D Motion & Reconstruction: Given a number of correspondences, and possibly the knowledge of the camera's intrinsic parameters, how to recovery the 3-D motion and structure of the observed world
- The Third Subproblem, and Fourth....
 - Motion Segmentation: what are the regions of the image plane corresponding to different moving objects?
 - Motion Understanding: lip reading, gesture, expression, event...

2D Motion Field



3D Motion of Camera

3D Motion: rotation w , translation T



Main Difference between Motion and Stereo

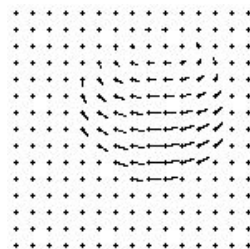
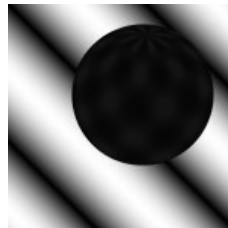
Stereo	Motion
Disparity	Motion field
Displacement – (dx, dy)	Differential concept – velocity (v_x, v_y), i.e. time derivative ($dx/dt, dy/dt$)
Consecutive frame sufficiently far apart to guarantee large enough disparity	Consecutive frame sufficiently close to approximate differential conditions

2D Motion Estimation

- Three prevalent approaches to computing 2D Motion Field:
 - gradient techniques
 - » obtain 2D image flow / optical flow from spatial and temporal image intensity derivatives.
 - token/feature matching or correlation
 - » similar to the methods covered in stereo
 - velocity sensitive filters
 - » frequency domain models of motion estimation
 - » not covered in this course.

Optical Flow

In the case of small image motion, the 2-D motion estimation problem is cast as the problem of estimating the time-varying derivatives of the image brightness pattern, usually called optical flow.



Gradient Techniques

- Suppose there is a grey-level gradient of 2 units per pixel in some small region of the image. The grey-levels in the region might be:

23	25	27	29	31
23	25	27	29	31
23	25	27	29	31

----->
g.l. gradient is 2 units/pixel this way

In the next frame, the values are

17	19	21	23	25
17	19	21	23	25
17	19	21	23	25

----->
Image has moved 3 pixels this way.
g.l.'s at corresponding pixel increases by 6

Gradient Techniques

- The change in grey-level at corresponding pixels is 6 units. If the spatial grey-level gradient is 2 units per pixel, and we shift the values by 3 pixels, then the total change is $2 \times 3 = 6$.

$$\text{distance moved per frame} = \frac{\text{Change in g.l. at corresponding pixels}}{\text{Spatial grey-level gradient}}$$

- Advantage: easy to measure these spatial and temporal gradients.
- Big assumption - the constant intensity assumption - that the change in grey-level is entirely due to motion. In other words, the grey-levels "move along with" the motions of the images of physical features. This is often not true
 - for instance specular reflections and highlights do not behave in this way, and the illumination can change
 - but it is a good enough assumption to make a reasonable starting point for estimating the flow.

Gradient Techniques

- Assume $E(x,y,t)$ is a continuous and differentiable function of space and time
- Suppose the brightness pattern is locally displaced by a distance dx, dy over time period dt .
 - we assume that as the time varying image evolves, the image brightness of points don't change as they move in the image
 $E(x,y,t) = E(x + dx, y + dy, t + dt)$
 - In other word, the total derivative of E w.r.t. time is zero.

$$\frac{dE}{dt} = 0$$

Gradient Techniques

- Expanding $E(x + dx, y + dy, t + dt)$ in a Taylor series & divide by dt :

$$\frac{dE}{dt} = 0 = \frac{\delta E}{\delta x} \frac{dx}{dt} + \frac{\delta E}{\delta y} \frac{dy}{dt} + \frac{\delta E}{\delta t}$$

- Can write as $dE/dt = E_x v_x + E_y v_y + E_t = 0 = \nabla E^T (v_x, v_y) + E_t = 0$.
 - The ∇E and E_t are derivatives measured from the image
 - (v_x, v_y) are the unknown optic flow components $(dx/dt, dy/dt)$ in the x and y directions, respectively
- This equation is known as the Brightness Constancy Equation (BCE) -- valid only if intensity change is due entirely to motion

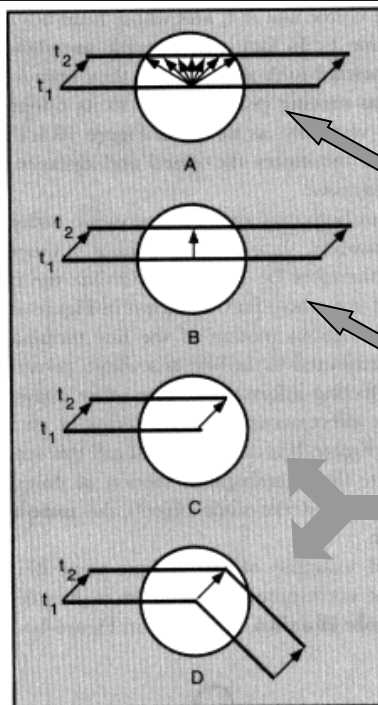
Aperture Problem

```

-----
| 23 | 25 | 27 | 29 | 31 |
|-----|
| 23 | 25 | 27 | 29 | 31 |
|-----|
| 23 | 25 | 27 | 29 | 31 |
|-----|

```

- Beside the constant brightness assumption, there is a further problem.
 - In the example above, a motion along the y-axis will not have any effect on the grey levels
 - We only have information about motion along the x-axis in this case, or in general along the grey-level gradient.
 - This component of motion is sometimes called the edge motion; it is at right-angles to the local edge direction.
- This effect is a manifestation of what is known as the aperture problem, because it is a result of considering a local region of the image only.



The aperture problem.

The motion of a continuously moving straight line behind an aperture is ambiguous because any point along it at time t_1 can correspond to any point at time t_2 (A).

In fact, observers perceive perpendicular motion (B) in the absence of any unique points.

When unique points are visible (C and D), the actual motion will be perceived. Unique point heuristic.

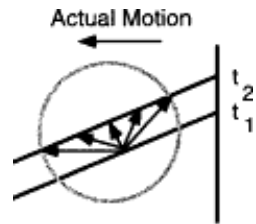
Aperture problem

The spatial and temporal derivatives at a point in the image only provide a linear constraint on the optic flow

$$(\nabla E)^T (v_x, v_y) + E_t = 0$$

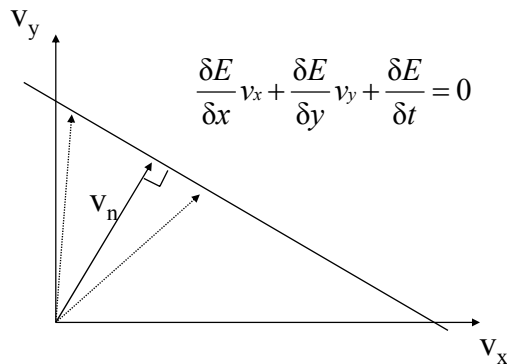
$$\frac{(\nabla E)^T (v_x, v_y)}{\|\nabla E\|} = -\frac{E_t}{\|\nabla E\|}$$

$$(v_x, v_y) \cdot (n_x, n_y) = v_n = -\frac{E_t}{\|\nabla E\|}$$



Aperture problem

Motion constraint line



Constraining the BCE

- Three common approaches:

- impose smoothness constraint on flow (H+S)

- Solution: iterative scheme minimizing departure of (u,v) from smoothness.



- assume velocity is locally constant (L+K)

- Solution: Least square of two variables (u,v) from NxN Equations in the NxN window



- assume local parametric model (eg affine, $u^T = Ap^T + b$).

- Solution: Least square of 6 variables (A,b) from NxN

- All methods combine information from a larger region. Other approaches also exist.

Horn and Schunck's algorithm

H+S assumptions: Flow field are varying smoothly within a small patch

Thus we shall try to minimize a measure of departure from smoothness

$$E_c^2 = \left(\frac{\partial v_x}{\partial x} \right)^2 + \left(\frac{\partial v_x}{\partial y} \right)^2 + \left(\frac{\partial v_y}{\partial x} \right)^2 + \left(\frac{\partial v_y}{\partial y} \right)^2$$

The error in the optical flow constraint equation should be small too

$$E_b = (\nabla E)^T (v_x, v_y) + E_t$$

Solve using calculus of variation

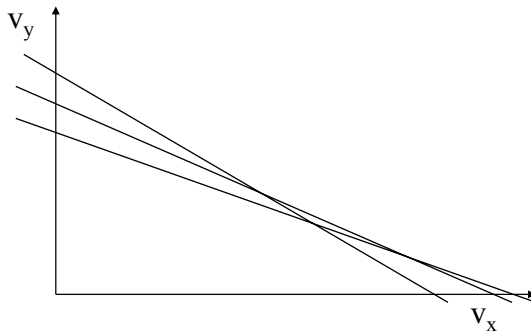
$$\min \iint (\alpha^2 E_c^2 + E_b^2) dx dy$$

Known as regularization approach.

big α : put emphasis on regularity
small α : put emphasis on the optical flow constraint equation

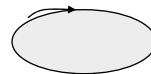
Horn and Schunck's algorithm

- Also requires that neighborhoods have edges with different orientations
 - If the constraint lines in a neighborhood are nearly parallel (i.e., the gradient directions are all similar), then the location of the best fitting (v_x, v_y) will be very sensitive to errors in estimating gradient directions.



Horn and Schunck's algorithm

- Many variations on the method of Horn & Schunck have been proposed. *(See course website for links where you can download different algorithms and sequences.)*
- One of its chief problems: smoothing blurs what ought to be sharp jumps in the flow field, at the edges of objects at different depths for instance.
- This can be addressed by
 - allowing explicit discontinuities in the vector field, e.g. discontinuous regularization.
 - combine vectors along closed contours (e.g. Hildreth 84) so that the problem of propagating the flow into large uniform regions is avoided.



Interlude: A regularization approach

- Many vision problems such as stereo reconstruction of visible surfaces and recovery of optic flow are instances of ill posed problems.
- A problem is well posed when its solution (Hadamard's definition):
 - exists
 - is unique, and
 - depends continuously on its initial data
- Any problem that is not well posed is said to be ill posed
- The optic flow problem is to recover both degrees of freedom of motion at each image pixel, given the spatial and temporal derivatives of the image sequence
 - but any solution chosen at each pixel that locally satisfies the motion constraint equation can be used to construct an optic flow field consistent with the derivatives measured
 - therefore, the solution is not unique - how to choose one?

A regularization approach

- Solution - add a priori knowledge that can choose between the solutions
- Formally, suppose we have an ill posed problem of determining z from data y expressed as
 - $Az = y$, where A is a linear operator (e.g., projection operation in image formation)
- We must choose a quadratic norm $\| \cdot \|$ and a so-called stabilizing functional $\| Pz \|$ and then find the z that minimizes:
 - $\|Az - y\|^2 + \lambda \|Pz\|^2$
 - λ controls the compromise between the degree of regularization and the closeness of the solution to the input data (the first term).
- T. Poggio, V. Torre and C. Koch, Computational vision and regularization theory, Nature, 317, 1984.

A regularization approach

- A regularization approach for optic flow:

-- the first term is $\| \frac{dE}{dt} \|^2 = \| (\nabla E)^T (v_x, v_y) + E_t \|^2$

-- the second term enforces a smoothness constraint on the optic flow field :

$$E_c^2 = \left(\frac{\partial v_x}{\partial x} \right)^2 + \left(\frac{\partial v_x}{\partial y} \right)^2 + \left(\frac{\partial v_y}{\partial x} \right)^2 + \left(\frac{\partial v_y}{\partial y} \right)^2$$

- The regularization problem is then to find a flow field that minimizes both terms.
- This minimization can be done over the entire image using various iterative techniques (calculus of variation; not needed).

The Lucas & Kanade Algorithm

- Assume motion field at given time, $v(x; y)$, is constant over a block of pixels B
- Not unreasonable for small blocks in homogeneous motion regions
- Error in BCE over block is then

$$\varepsilon = \sum ((\nabla E)^T (v_x, v_y) + E_t)^2$$

- Estimated motion at centre of block is then the flow which minimises ε . Typical least square problem.
- Repeat procedure for all point

L&K Algorithm: Example

3x3 Block B

Assume: derivative computed using forward difference. Spatial derivatives computed using frame at t_0 only.

23 25 27 29 31	17 19 21 23 25
23 25 27 29 31	17 19 21 23 25
23 25 27 29 31	17 19 21 23 25
23 25 27 29 31	17 19 21 23 25
23 25 27 29 31	17 19 21 23 25
at time t_0	at time t_1

- Error in BCE over 3x3 block B is

$$\varepsilon = \sum_{x, y \in B} ((\nabla E)^T (v_x, v_y) + E_t)^2$$

$$\begin{bmatrix} \frac{\partial E}{\partial x} & \frac{\partial E}{\partial y} \\ \vdots & \vdots \end{bmatrix} \begin{bmatrix} v_x \\ v_y \end{bmatrix} = \begin{bmatrix} -E_t \\ \vdots \end{bmatrix}$$

$$\begin{bmatrix} 2 & 0 \\ 2 & 0 \\ \vdots & \vdots \end{bmatrix} \begin{bmatrix} v_x \\ v_y \end{bmatrix} = \begin{bmatrix} 6 \\ \vdots \end{bmatrix}$$

9 equations

What is the solution?
What happens to SVD?

First Order (Affine) Methods

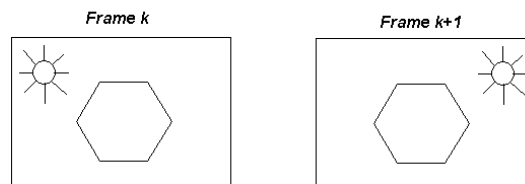
- L+K algorithm assumes constant velocity within region - only valid within small regions.
- For larger regions variation of motion with position is more complex. E.g. moving plane gives rise to quadratic motion field.
- Improved performance can be achieved by integrating optical flow estimates over larger regions using higher order motion models.
- Example - first order (affine) model, ie

$$v_x = a_1 + a_2 x + a_3 y \quad v_y = a_4 + a_5 x + a_6 y$$
 - For given region, fit above model to optical flow estimates using method of least squares.
- Assumptions: 1) region is planar; 2) region is small so that 2nd order flow variation can be ignored.

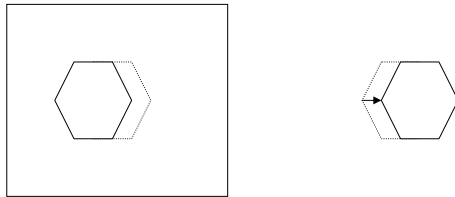
Optical Flow vs Motion Field

- The optical flow approximates the motion field. Assumptions required are:
 - no photometric distortion
 - no occlusion problem occurs.
 - Lambertian surfaces
 - pointwise light sources at infinity
- Error of this approximation will be small at points with high spatial gradient
- Error exactly zero for translational motion or for any rigid motion such that the illumination direction is parallel to the angular velocity

Non constant illumination

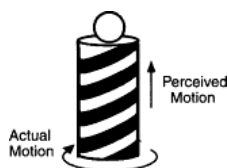


Occlusion Problem

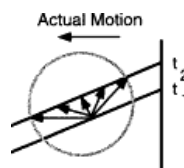


- It is impossible to find correspondence in the to be covered and uncovered regions. An example: Barber-pole illusion.
- Thus strictly speaking, the motion field estimates are accurate only if the changes in intensity over time are due to texture and surface markings on the moving surfaces, rather than depth discontinuities occluding each other.

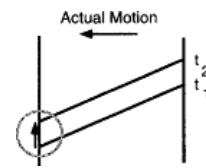
Barber-pole Illusion



A. Barberpole Illusion



B. Possible Motions in the Center



C. Possible Motion at the Edge

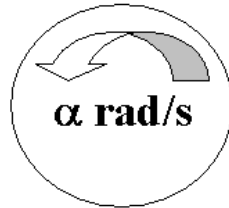
If the stripes were made up of many red dots on a white background so that correspondences over time were unambiguous, the barberpole illusion should disappear.

Notice that the ends of the stripes at the bottom edge of the barberpole are actually moving leftward. Because they are true endpoints, they specify the true motion of the pole.

Why do these bottom endpoints not determine the perceived motion of the whole barberpole? The answer appears to be that they are simply overwhelmed by the larger number of ends along the sides.

If the barber-pole is fat and short (so that there are less endpoints), then the perceived motion is horizontal!

Insufficient spatial image gradient



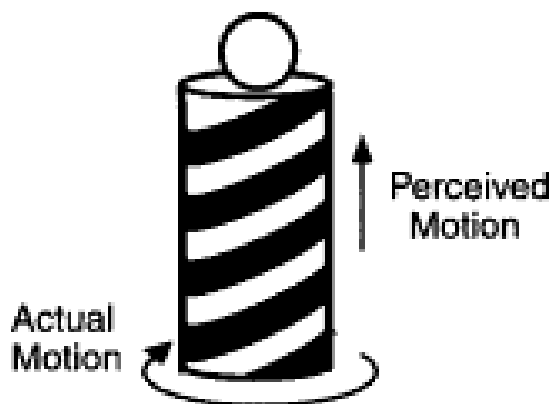
Flow vs Correspondence

- Gradient based methods only work when the 2D motion is “small” so that the derivatives can be reliably computed
 - example of large “motion”: stereo pair, action movie.
 - to cope with “large” 2D motions, one has to employ multiresolution methods
- Preferably, correspondence algorithms should be used to compute 2D motion when the motion is “large”.
 - correlation
 - feature tracking

Summary: Computing 2D Image Motion

- We wish to estimate the 2-D motion field. What we observe is the brightness variation on the image.
- The estimate obtained from the brightness variation is so-called apparent motion or optical flow.
- The link between the optic flow & the 2-D motion field is not always obvious. eg1

Barber-pole Illusion



The Brightness Constancy Equation (BCE)

$$\frac{dE}{dt} = 0 = \frac{\delta E}{\delta x} \frac{dx}{dt} + \frac{\delta E}{\delta y} \frac{dy}{dt} + \frac{\delta E}{\delta t}$$

$$(\nabla E)^T (v_x, v_y) + E_t = 0$$

$$(v_x, v_y) = (dx/dt, dy/dt)$$

= the components of the motion field

Only normal flow (optical flow projected along the normal direction) can be determined.

1533
to
1587



29563
to
29633



2100
to
2145



47098
to
47133



48998
to
49018



48682
to
48708



3746
to
3764



3764
to
3777



End of 2D Motion Estimation

Reference:

- Chapter 8, Textbook
- Handout (Motion and Depth Perception)