Optical Flow Estimation

Goal: Introduction to image motion and 2D optical flow estimation.

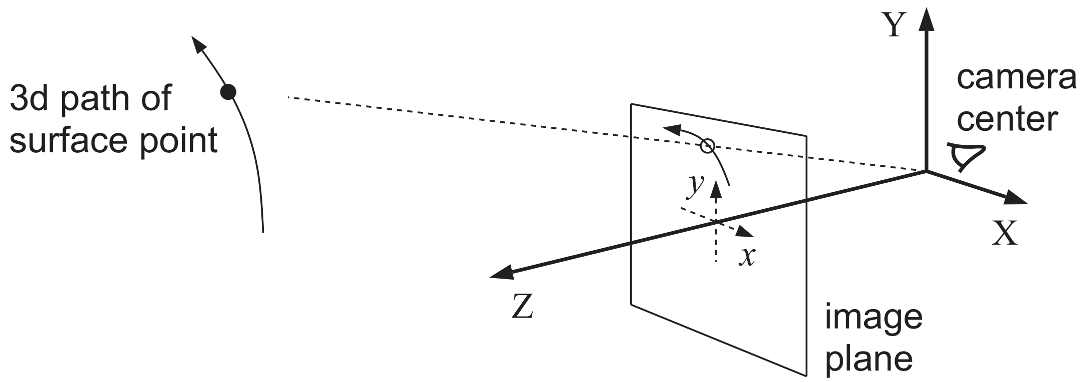
Motivation:

- Motion is a rich source of information about the world:
  - segmentation
  - surface structure from parallax
  - self-motion
  - recognition
  - understanding behavior
  - understanding scene dynamics
- Other correspondence / registration problems:
  - stereo disparity (short and wide baseline)
  - computer-assisted surgery
  - multiview alignment for mosaicing or stop-frame animation

Readings: Fleet and Weiss (2005)

Matlab Tutorials: motionTutorial.m

Introduction to Image Motion



A 3D point $\vec{\mathbf{X}}$ follows a space-time path $\vec{\mathbf{X}}(t)$. Its velocity is $\vec{\mathbf{V}}$ is

$$\vec{\mathbf{V}} = \frac{d\vec{\mathbf{X}}(t)}{dt} = \left(\frac{dX(t)}{dt}, \frac{dY(t)}{dt}, \frac{dZ(t)}{dt}\right)^T .$$

Perspective projection (for nodal length f) of the 3D path onto the image plane produces a 2D path,

$$\vec{\mathbf{x}}(t) \;=\; (x(t),\, y(t)) \;=\; \left(\frac{fX(t)}{Z(t)}, \frac{fY(t)}{Z(t)}\right) ,$$

the instantaneous 2D velocity of which is

$$\begin{aligned}\vec{\mathbf{u}} \;&=\; \left(\frac{dx(t)}{dt}, \frac{dy(t)}{dt}\right)^T \\ &=\; \frac{f}{Z(t)}\left(\frac{dX(t)}{dt}, \frac{dY(t)}{dt}\right)^T \;-\; \frac{f}{Z^2(t)}\frac{dZ(t)}{dt}(X(t),\, Y(t))^T .\end{aligned}$$

Definition: **2D Motion Field** – 2D velocities for all visible points.

Optical Flow Field – Estimate of the 2D motion field.

Optical Flow

Two Key Problems:

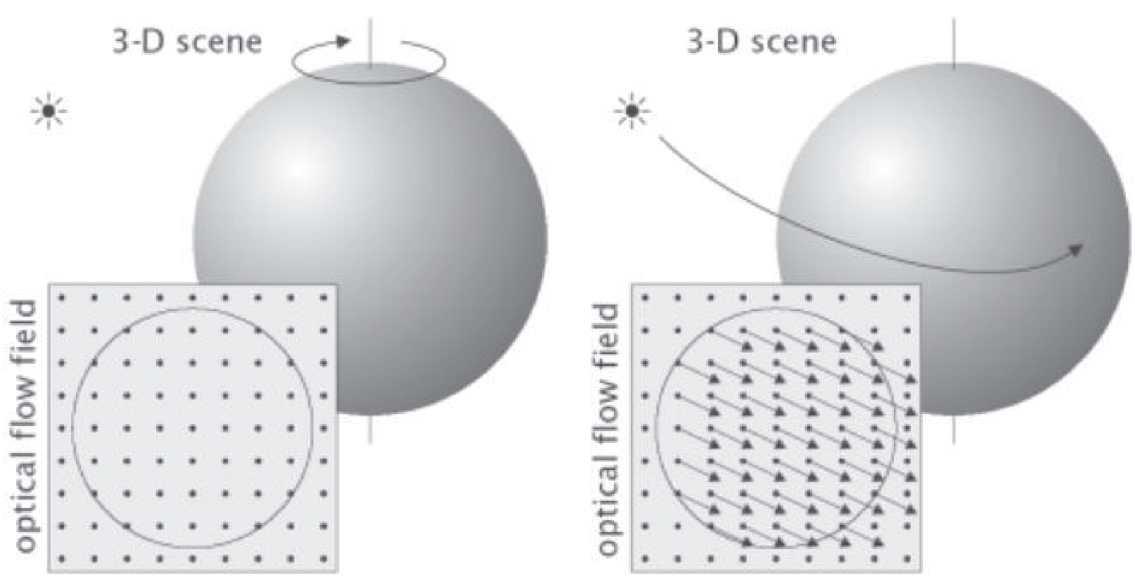
1. Determine what image property to track.
2. Determine how to track it

Brightness constancy: More precisely, let's track points of constant brightness, assuming that surface radiance is constant over time:

$$I(x, y, t + 1) = I(x - u_1, y - u_2, t) .$$

Brightness constancy is often assumed by researchers, and often violated by Mother Nature; so the resulting optical flow field is sometimes a very poor approximation to the 2D motion field.

For example, a rotating Lambertian sphere with a static light source produces a static image. But a stationary sphere with a moving light source produces drifting intensities (figure from Jahne et al, 1999).



Gradient-Based Motion Estimation

Let $f(x)$ denote a 1D greylevel function of spatial position, and assume $f(x)$ is just translated by d between time 1 and time 2:

$$f_2(x) = f_1(x - d) \ .$$

We can express the shifted signal as a Taylor expansion of f_1 about x:

$$f_1(x - d) \ = \ f_1(x) - d\, f_1'(x) + O(d^2 f_1'') \ ,$$

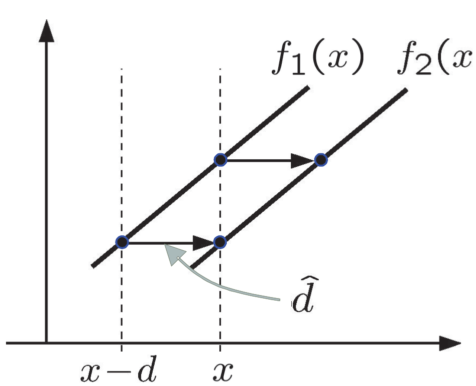
in which case the difference between the two signals is given by

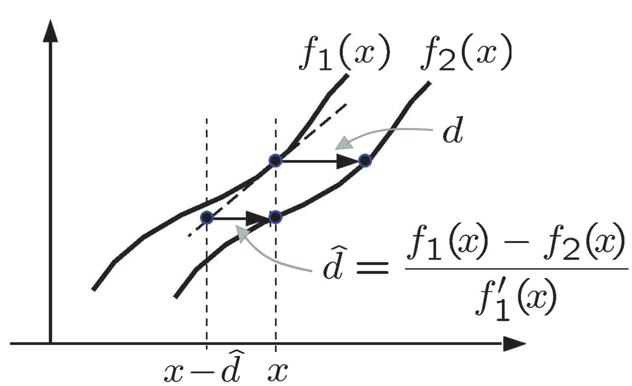
$$f_2(x) - f_1(x) \ = \ -\, d\, f_1'(x) + O(d^2 f_1'') \ .$$

Then, a first-order approximation to the displacement is

$$\hat{d} \ = \ \frac{f_1(x) - f_2(x)}{f_1'(x)} \ .$$

For linear signals the first-order estimate is exact.





For nonlinear signals, the accuracy of the approximation depends on the displacement magnitude and the higher-order signal structure.

Gradient Constraint Equation

In two spatial dimensions, the first-order approximation is:

$$f(x+u_1, y+u_2,\ t+1) \approx \\ f(x,y,t) + u_1 f_x(x,y,t) + u_2 f_y(x,y,t) + f_t(x,y,t) \quad (1)$$

Taking the image difference with times t and $t+1$, then yields:

$$u_1\, f_x(x,y,t) + u_2\, f_y(x,y,t) + f_t(x,y,t) \ = \ 0\ .$$

Written in vector form, with $\vec{\nabla} f \equiv (f_x,\ f_y)^T$:

$$\vec{\mathbf{u}}^T \vec{\nabla} f(x,y,t) + f_t(x,y,t) \ = \ 0\ .$$

When the duration between frames is large, it is sometimes more appropriate to use only spatial derivatives in the Taylor series approximation in (1). Then one obtains a different approximation

$$\vec{\mathbf{u}}^T \vec{\nabla} f(x,y,t) + \Delta f(x,y,t) \ = \ 0$$

where $\Delta f(x,y,t) \ = \ f(x,y,t+1) - f(x,y,t)$.

Brightness Conservation

One can also derive the gradient constraint equation directly from brightness conservation.

Let $(x(t),\ y(t))$ denote a space-time path along which the image intensity remains is constant; i.e., the time-varying image f satisfies

$$f(x(t), y(t), t) = c$$

Taking the total derivative of both sides gives

$$\frac{d}{d\,t}\, f(x(t), y(t), t) = 0$$

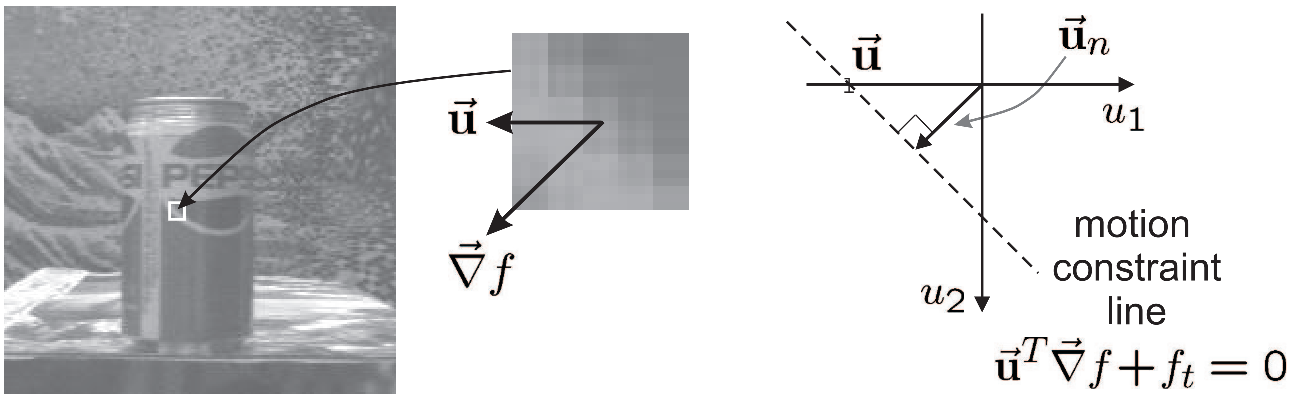
The total derivative is given in terms of its partial derivatives

$$\begin{aligned}
\frac{d}{d\,t}\, f(x(t), y(t), t) &= \frac{\partial f}{\partial x}\frac{dx}{dt} \;+\; \frac{\partial f}{\partial y}\frac{dy}{dt} \;+\; \frac{\partial f}{\partial t}\frac{dt}{dt} \\
&= f_x\, u_1 \;+\; f_y\, u_2 \;+\; f_t \\
&= \vec{\mathbf{u}}^T \vec{\nabla} f + f_t \\
&= 0
\end{aligned}$$

This derivation assumes that there is no aliasing, so that one can in principle reconstruct the continuous underlying signal and its derivatives in space and time. There are many situations in which this is a reasonable assumption. But with common video cameras temporal aliasing is often a problem with many video sequences, as we discuss later in these notes.

Normal Velocity

The gradient constraint provides one constraint in two unknowns. It defines a line in velocity space:



The gradient constrains the velocity in the direction normal to the local image orientation, but does not constrain the tangential velocity. That is, it uniquely determines only the normal velocity:

$$\vec{\mathbf{u}}_n = \frac{-f_t}{||\vec{\nabla} f||} \frac{\vec{\nabla} f}{||\vec{\nabla} f||}$$

When the gradient magnitude is zero, we get no constraint!

In any case, further constraints are required to estimate both elements of the 2D velocity $\vec{\mathbf{u}} = (u_1, u_2)^T$.