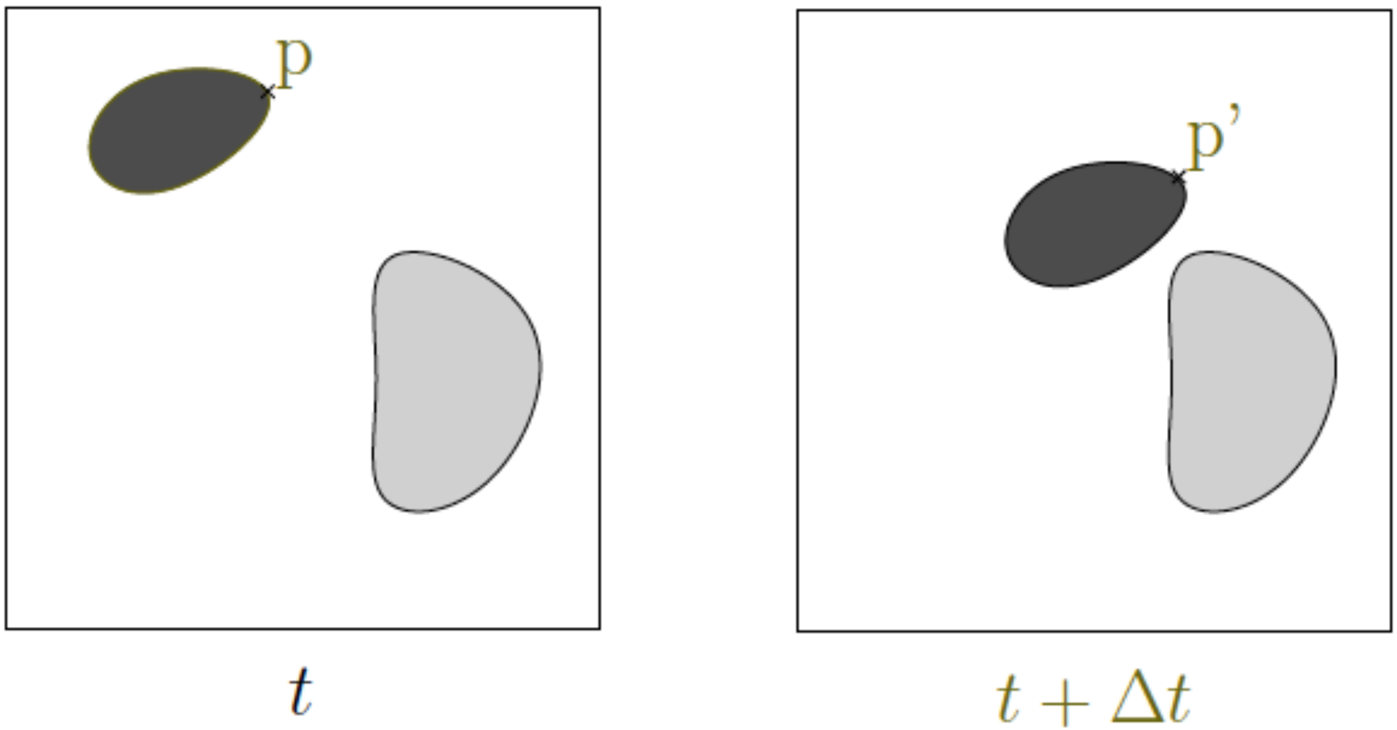
Measuring Optical Flow

Jitendra Malik

Optical flow is based on correspondence over time

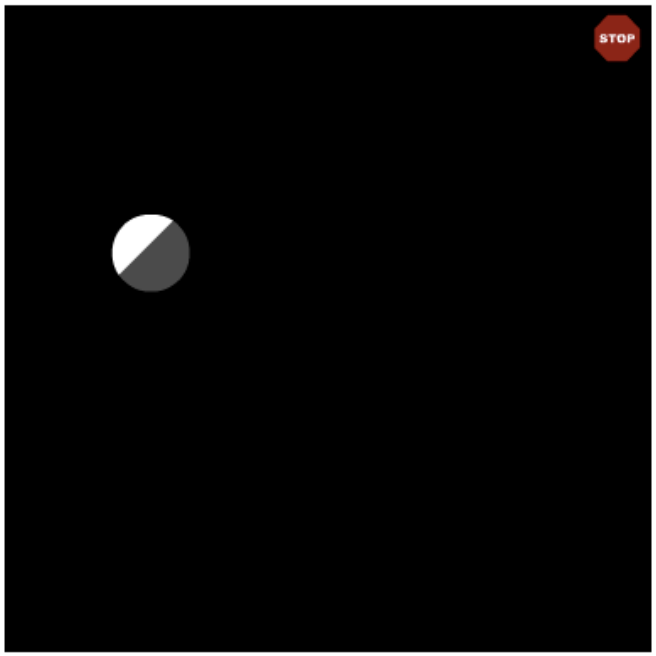


If we can solve this correspondence, then if p has coordinates (x_1, y_1, t_1) and p' has coordinates (x_2, y_2, t_2), we get $u = \frac{x_2 - x_1}{t_2 - t_1}$ and $v = \frac{y_2 - y_1}{t_2 - t_1}$

http://web.mit.edu/persci/demos/Motion&Form/master.html

TWO SQUARES

View the explanatory figure



Two moving shapes viewed through apertures reveal the problems that occur when interpreting local motions. The edge motions (a) are ambiguous by virtue of the aperture problem - they appear to move diagonally, but are in fact consistent with many other motions as well. The corner motions (b), in contrast, are unambiguously horizontal, since corners are 2D features. These features might be used to disambiguate the motion of edges. However, some features, such as the T-junction (c) are the artifactual result of occlusion. The unambiguous vertical motion of the T-junction is spurious and must somehow be discounted. Viewing the whole scene at once (d) reveals the two translating squares that caused the local motions seen through the apertures. The visual system manages to correctly discount the spurious local motions.

- **a** one edge
- **b** a corner
- **c** a T-junction
- **d** the two squares revealed
- **e** two edges from one square
- **f** two edges from the other square
- **g** an edge from either square
- **h** both squares revealed

Source: McDermott, J., Weiss, Y., Adelson, E.H. (2001). Beyond junctions: Nonlocal form contraints on motion interpretation. Perception, 30: 905-923.

Brightness constancy assumption: we assume that the brightness of a given point remains constant over a short period of time.

So we assume that $I(x_1, y_1, t_1) = I(x_2, y_2, t_2)$ for corresponding points. By differentiating we immediately get

$$\begin{aligned} \mathrm{d}I &= I_x \mathrm{d}x + I_y \mathrm{d}y + I_t \mathrm{d}t \\ &= 0 \end{aligned}$$

Dividing by $\mathrm{d}t$, this equation can be rewritten as

$$I_x \frac{\mathrm{d}x}{\mathrm{d}t} + I_y \frac{\mathrm{d}y}{\mathrm{d}t} + I_t = 0$$

$$\implies I_x u + I_y v + I_t = 0$$

We rewrite the optical flow constraint equation as

$$\begin{pmatrix} I_x & I_y \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = -I_t$$

which can in turn be rewritten in the concise form $\nabla I \cdot u = -I_t$. u now denotes what was previously the vector (u, v).

Suppose we have a neighborhood of n pixels surrounding a point p. We make the following assumption (in addition to the brightness constancy assumption, which we still assume to hold):

The optical flow is constant in a small neighborhood of the point p.

If we define $I^i = I^i(x_i, y_i, t_i)$ to be the intensity of the i-th point in the neighborhood of p, we then have the following equation

$$\begin{pmatrix} I_x^1 & I_y^1 \\ I_x^2 & I_y^2 \\ \vdots & \vdots \\ I_x^n & I_y^n \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = - \begin{pmatrix} I_t^1 \\ I_t^2 \\ \vdots \\ I_t^n \end{pmatrix}$$

This equation is of the form $Au = -b$ with

$$A = \begin{pmatrix} I_x^1 & I_y^1 \\ I_x^2 & I_y^2 \\ \vdots & \vdots \\ I_x^n & I_y^n \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} I_t^1 \\ I_t^2 \\ \vdots \\ I_t^n \end{pmatrix}$$

We solve it using linear least squares :

$$u = -(A^T A)^{-1} A^T b$$

We have that

$$A^T A = \begin{pmatrix} \sum_i (I_x^i)^2 & \sum_i I_x^i I_y^i \\ \sum_i I_x^i I_y^i & \sum_i (I_y^i)^2 \end{pmatrix} = \sum_i \nabla I^i (\nabla I^i)^T$$

where

$$\nabla I^i (\nabla I^i)^T = \begin{pmatrix} (I_x^i)^2 & I_x^i I_y^i \\ I_x^i I_y^i & (I_y^i)^T \end{pmatrix}$$

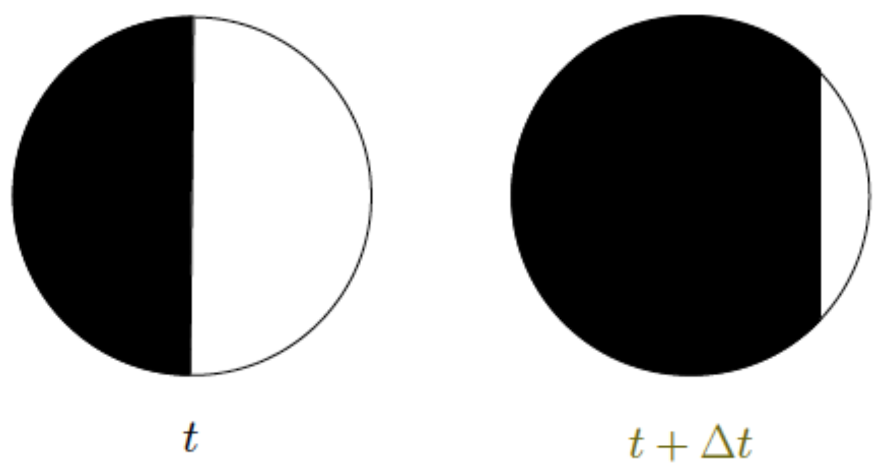
is a rank 1 matrix. The matrix $A^T A$ is called the *second moment matrix*.

$$u = -\left(\sum_i \nabla I^i (\nabla I^i)^T\right)^{-1} \begin{pmatrix} \nabla I^1 & \cdots & \nabla I^n \end{pmatrix} \begin{pmatrix} I_t^1 \\ \vdots \\ I_t^n \end{pmatrix}$$

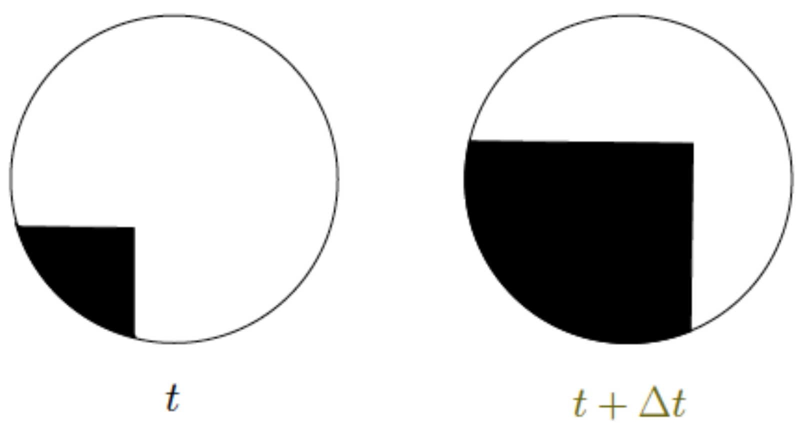
This equation involves the inverse of the second moment matrix. When does this inverse make sense? The second moment matrix is a 2×2 matrix, so it can have rank 0, 1 or 2:

- Rank 0. I is constant over the image : the second moment matrix is then the null matrix. In this case we cannot solve the equations; anyways we wouldn't expect to be able to compute the optical flow in such a case.

- Rank 1. In the example of the black square that we considered before, in the neighborhood defined by our aperture we have that for all points $I_y = 0$, and for some points $I_x \neq 0$ (the points on the border), so that $\sum \nabla I (\nabla I)^T = \begin{pmatrix} K & 0 \\ 0 & 0 \end{pmatrix}$. This matrix is of rank 1 and is not invertible, so we cannot estimate the optical flow. Here, only the normal component (i.e. the component in the direction of the image gradient) can be computed. Indeed, the equation $\nabla I \cdot u = -I_t$ shows that we can estimate the projection of u on ∇I, which is precisely the normal component.



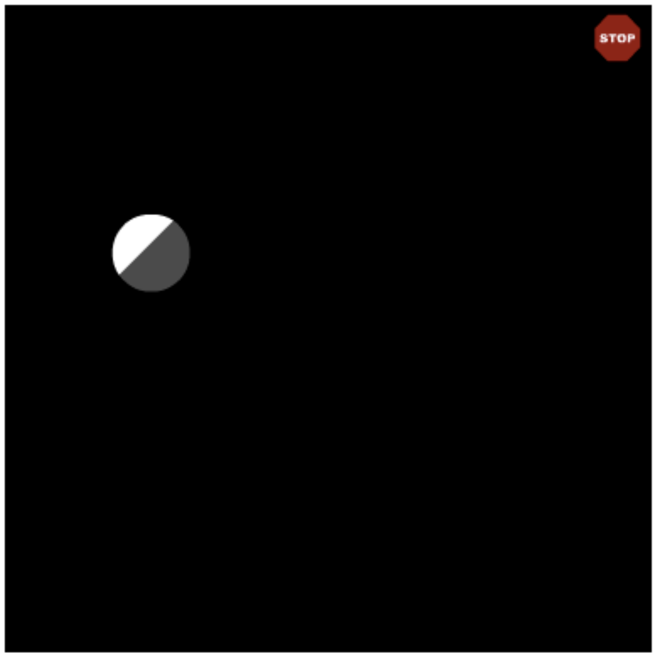
- Rank 2. The second moment matrices corresponding to the vertical edge of the corner have the form $\begin{pmatrix} K_1 & 0 \\ 0 & 0 \end{pmatrix}$, whereas those corresponding to the horizontal edge have the form $\begin{pmatrix} 0 & 0 \\ 0 & K_2 \end{pmatrix}$, so that when we add these matrices up over all points in the neighborhood, we get a rank 2 matrix that is invertible. This makes sense physically: our intuition is that if we see a corner, we can know precisely it which direction it is moving.



http://web.mit.edu/persci/demos/Motion&Form/master.html

TWO SQUARES

[View the explanatory figure]



- **a** one edge
- **b** a corner
- **c** a T-junction
- **d** the two squares revealed
- **e** two edges from one square
- **f** two edges from the other square
- **g** an edge from either square
- **h** both squares revealed

Two moving shapes viewed through apertures reveal the problems that occur when interpreting local motions. The edge motions (a) are ambiguous by virtue of the aperture problem - they appear to move diagonally, but are in fact consistent with many other motions as well. The corner motions (b), in contrast, are unambiguously horizontal, since corners are 2D features. These features might be used to disambiguate the motion of edges. However, some features, such as the T-junction (c) are the artifactual result of occlusion. The unambiguous vertical motion of the T-junction is spurious and must somehow be discounted. Viewing the whole scene at once (d) reveals the two translating squares that caused the local motions seen through the apertures. The visual system manages to correctly discount the spurious local motions.

Source: McDermott, J., Weiss, Y., Adelson, E.H. (2001). Beyond junctions: Nonlocal form contraints on motion interpretation. Perception, 30: 905-923.