My first page

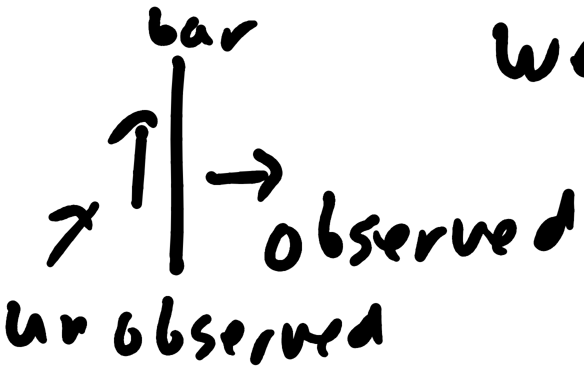
Friday, October 14, 2016 6:41 PM

Motion Lecture 1.

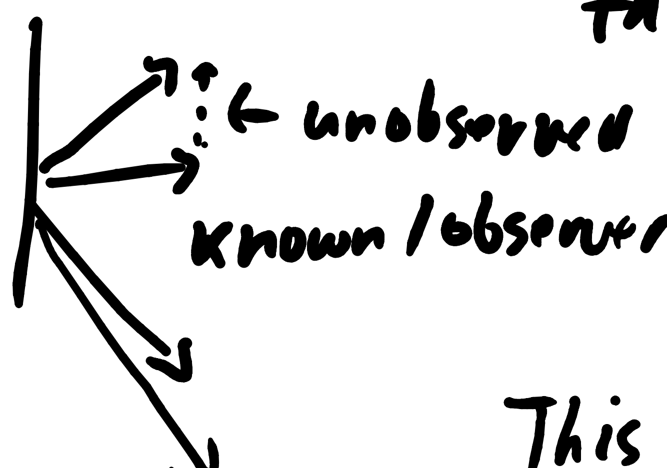
The barberpole illusion shows that perception of motion is not straightfoward. The barberpoles rotate to the right. But the perception of motion is vertically upwards.

This is because locally there is often not enough information to determine the motion unambiguously.

For example, consider a moving bar.



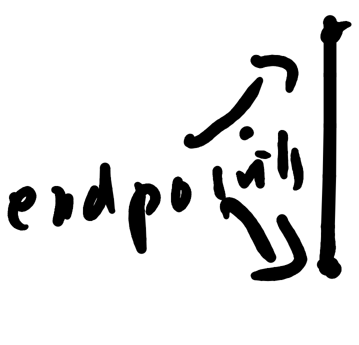
We can observe the motion in the direction perpendicular to the bar. But we cannot observe the motion along the bar.



So the local observation is consistent with many possible motions.

This is the Apertune Problem.

If the bar has visible endpoints, then we can observe their motion.

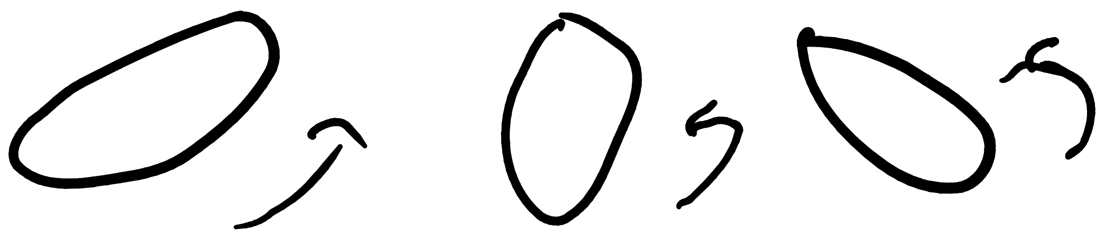


But the observations at the endpoints have to propagate to the other points on the bar.

ellipse)! to the other points on the ...
How is this done?
How far can information at unambiguous points (e.g. endpoints) be propagated? (See next lecture)

Consider a rotating ellipse. This has no unambiguous points (no endpoints)

ellipse

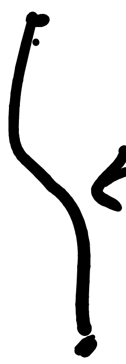


It is perceived either as:
(i) a non-rigid rotating ellipse.
or (ii) a rigid circle rotating in 3D.

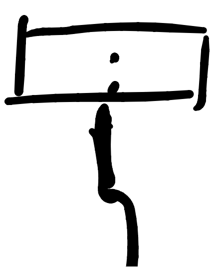
But, not seen as a rigidly rotating ellipse. Unless the aspect ratio is very big. e.g.

Because now it appears to have endpoints

Other experiments - Nakayama - on motion capture.

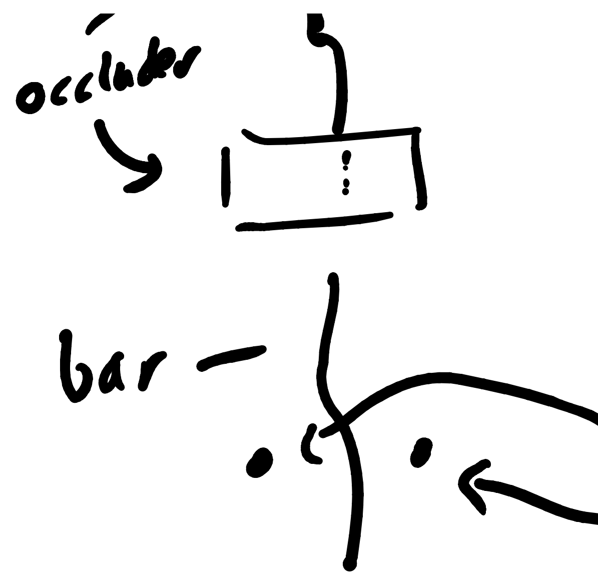


True Motion: up and down rigid.
Perceived Motion non-rigid. if endpoints are far away.



occluder

True Motion: up and down rigid
End points hidden by occluders.
Perceived Motion: non-rigid.



Perceived Motion : non-rigid.

True Motion : up and down rigid
side points have unambiguous motion.

If side points are close to bar, then they "capture" the bar and it is perceived to move rigidly.

Short Range & Long Range Motion.

The human eye receives input images as a continuous stream in time.

E.g $I(\underline{x}, t)$ — t variable is continuous

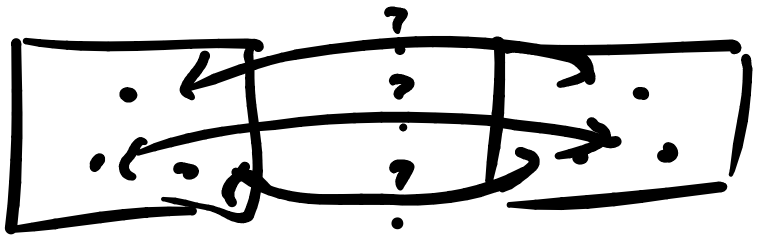
But cameras output a discrete set of image frames $I(\underline{x}, t)$, $I(\underline{x}, t+\Delta)$, $I(\underline{x}, t+2\Delta)$, ...
where Δ is the frame rate.

Typical movies have 24 frames per second but recent movies (e.g. the Hobbit) try 48 frames per second. Dogs may not be able to see motion at 24 frames per second, but humans can. Humans can also perceive motion with fewer frames per second.

Long Range Motion — few frames per second.
This means big changes between adjacent images $I(\underline{x}, t)$ and $I(\underline{x}, t+\Delta)$

images $I(\underline{x},t)$ and $I(\underline{x},t+\Delta)$.

Long Range Motion has the correspondence problem.



$I(\underline{x},t)$ $\quad$ $I(\underline{x},t+\Delta)$

which dots correspond between the two images.

From a mathematical perspective, short-range motion is approximately differentiable (because the time frames are sufficiently close together), but long-range motion is not.

Differentiable $\quad I(\underline{x},t) = F(\underline{x}-\underline{v}t)$ $\quad$ motion $\underline{v}$

$$\underline{\nabla} I(\underline{x},t) = \underline{\nabla} F(\underline{x}-\underline{v}t)$$

$$\frac{\partial I}{\partial t}(\underline{x},t) = -\underline{v}\cdot\underline{\nabla} F(\underline{x}-\underline{v}t)$$ chain rule of differentiation

Gives optical flow equation.

$$\underline{v}\cdot\underline{\nabla} I(\underline{x},t) + \frac{\partial I}{\partial t}(\underline{x},t) = 0$$

specifies motion component in direction of gradient $\underline{\nabla} I$ only.

motion component perpendicular to gradient is unknown. Aperture problem.

In practice, we must approximate derivatives by differences.

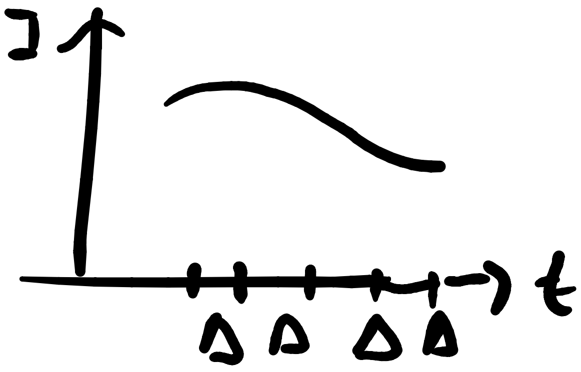
derivatives by differences

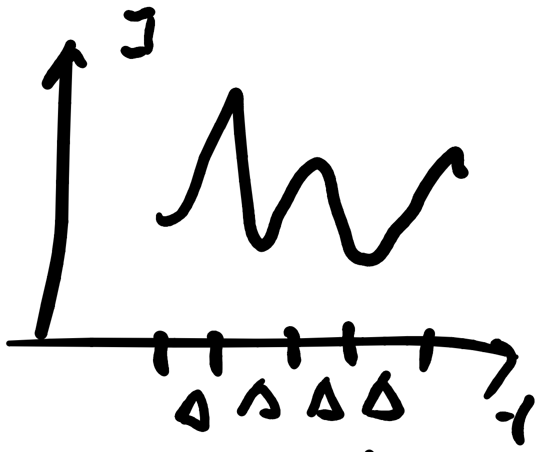
e.g. $\frac{\partial I}{\partial t}(\underline{x},t) = \lim_{\delta \to 0} \frac{I(\underline{x},t+\delta) - I(\underline{x},t)}{\delta}$

$\approx \frac{I(\underline{x},t+\Delta) - I(\underline{x},t)}{\Delta}$

This approximation is ok if Δ is small compared with the rate of change of $I(\cdot)$ with respect to time

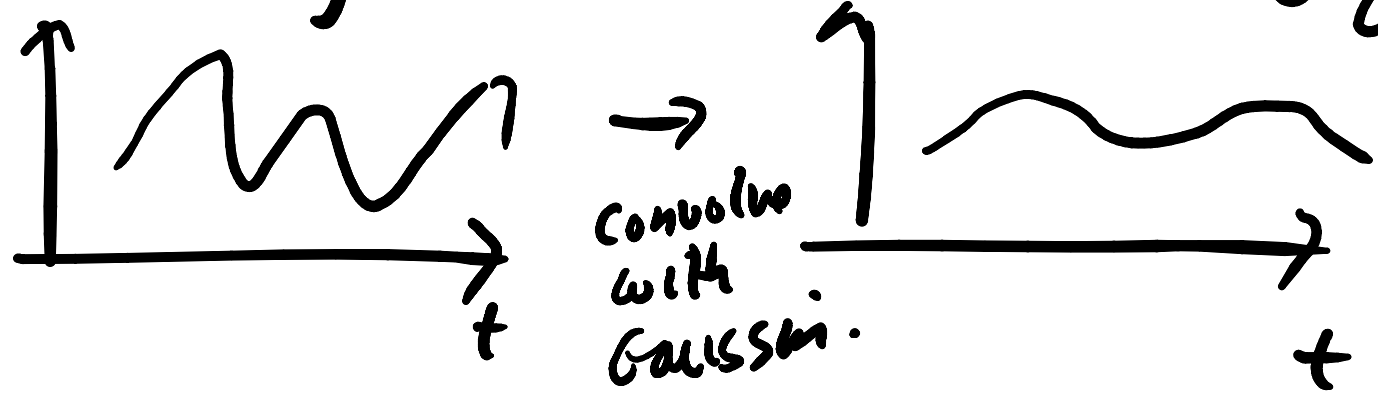
ok ✓





Not okay ✗

In practice, this effect can be reduced by smoothing the image by convolving with Gaussian



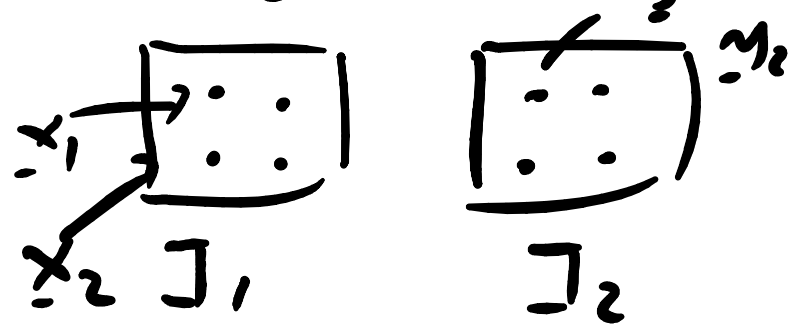
Note: this applies to both time and space.

Theory of Long Range Motion.

Minimal Mapping Theory (Ullman): Two Image Frames only.

$I_1(\underline{x}) = \{ \underline{x}_i : i = 1 \text{ to } N \}$

$I_2(\underline{x}) = \{ \underline{y}_a : a = 1 \text{ to } N \}$



Correspondence Variable

$V_{ia} \in \{0,1\}$ $\quad i = 1$ to M, $a = 1$ to N

$V_{ia} = 1$, means point $\underline{x}_i$ corresponds to $\underline{y}_a$

$V_{ia} = 0$, otherwise.

Unique match constraint:

$\sum_a V_{ia} = 1$, for $i = 1$ to M.

each point $\underline{x}_i$ is only matched to a single point $\underline{y}_a$

$\sum_i V_{ia} = 1$, for $a = 1$ to N

Determine the correspondence by minimizing

$$E_{nn}[V] = \sum_{ia} V_{ia} |\underline{x}_i - \underline{y}_a|^2$$

subject to $\sum_a V_{ia} = 1, \forall i$

$\sum_i V_{ia} = 1, \forall a$.

Find correspondence to minimize distance $|\underline{x}_i - \underline{y}_a|^2$. Hence minimal mapping.

How to minimize $E_{nn}[V]$ w.r.t. V?

$E_{nn}[V]$ is a special case of the <u>linear assignment problem</u>.

Can be solved in polynomial time by the Hungarian Algorithm, the Auction Algorithm.

<u>Note</u>: can also solve to get $\hat{V} = \arg\min_V E[V]$ by mean field theory.

$$F_{nn}[S] = \sum_{ia} S_{ia} |\underline{x}_i - \underline{y}_a|^2 + T \sum_{ia} S_{ia} \log S_{ia}$$

subject to $\sum S_{ia} = 1 \; \forall i$, $\sum S_{ia} = 1 \; \forall a$

\+ constraints $\sum_a S_{ia}=1, \forall i, \sum_i S_{ia}=1, \forall a$

$F_{mm}[S]$ is a convex function. So mean field theory will converge to the optimal solution. Detailed analysis (Kossowsky & Yuille 1993)

This gives a neuronally plausible algorithm.

<u>Short Range Motion.</u> (Horn & Schunk 1981)

$$E[\underline{v}(\underline{x})] = \int \left(\underline{v}(\underline{x}) \cdot \underline{\nabla} I(\underline{x}) + \frac{\partial I}{\partial t}(\underline{x}) \right)^2 d\underline{x} + \lambda \int \frac{\partial \underline{v}}{\partial \underline{x}} \cdot \frac{\partial \underline{v}}{\partial \underline{x}} \, d\underline{x}$$

data term

local smoothness. First order.

The smoothness function regularizes the problem so there is a unique solution.

From probabilistic perspective, the smoothness term is a prior on the velocity.

Most modern algorithms to estimate motion are derived from Horn & Schunk. (coarse to fine).

A unified approach.

<u>Minimal Mapping</u> finds correspondence in long range motion by minimizing the displacement – i.e. slow motion.

<u>Horn & Schunk.</u> solves aperture problem for [illegible] smoothness.

motion estimation by assuming smoothness.

Both slowness & smoothness are reasonable assumptions to make about motion. Can be justified by mathematical analysis or statistical studies of natural images. slow-and-smooth

Consider an alternative model for long range motion using a correspondence variable V and a velocity field $\underline{v}(\underline{x})$

$$E[V, \underline{v}] = \sum_{i,a} V_{ia} \left(\underline{y}_a - \underline{x}_i - \underline{v}(\underline{x}_i) \right)^2$$

$$+ \lambda \int \underline{v}(\underline{x}) \cdot \underline{v}(\underline{x}) \, d\underline{x}$$

$$+ \mu \int \frac{\partial \underline{v}}{\partial \underline{x}} \cdot \frac{\partial \underline{v}}{\partial \underline{x}} \, d\underline{x}$$

$$+ \nu \int \frac{\partial^2 \underline{v}}{\partial \underline{x}^2} \cdot \frac{\partial^2 \underline{v}}{\partial \underline{x}^2} \, d\underline{x}$$

if $\underline{x}_i$ matches $\underline{y}_a$ then we have velocity $\underline{v}(\underline{x}_i) = \underline{y}_a - \underline{x}_i$

slowness

first-order-smoothness

second-order smoothness.

This model gives similar results to Minimal Mapping for a special set of parameters λ, μ, ν.

But, in general, the model yields smoother prediction than minimal mapping – which agrees with human experiments. Yuille & Grzywacz Grzywacz & Yuille

<u>Note</u>: we can use the slowness & second-order smoothness for short-range motion also.

for short-range motion ...

Hence, short-range & long-range can be formulated in similar ways. Same priors, but different data terms.

$$\sum_{ia} V_{ia} \left(\underline{y}_a - \underline{x}_i - \underline{v}(\underline{x}_i) \right)^2 \quad \text{versus.} \quad \int \left(\underline{v} \cdot \underline{\nabla} I + \frac{\partial I}{\partial t} \right)^2 d\underline{x}$$

Comments: (1) Short-range motion can be discretized in space. This reduces to a Markov Random Field.

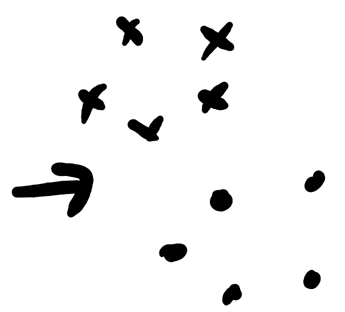
(2) The models use L^2 norms – e.g. $|\underline{v}(\underline{x})|^2$ – But this can be replaced by other norms like $|\underline{v}(\underline{x})|$ – L^1 norm.

(3) We have ignored motion discontinuities. This requires more complex models.

Are humans Ideal observers for motion?

Recall an ideal observer is a system which minimizes Bayes Risk.

Barlow & Tripathy considered long range motion where a set of dots translate rigidly to the left or the right by a fixed amount.



Time Frame 1

Time Frame 2

. . .

In addition, there are a random set of dots in both images.

Barlow & Tripathy made approximations to obtain an ideal observer model for task such as: (i) do the dots move right or left? (ii) how far do they move?

Human perception was tested. And the performance was much worse than the ideal observer model by many orders of magnitude.

Degrading Barlow & Tripathy's Ideal Observer model — by assuming that human vision has low resolution — gives better agreement with experiments.

But more detailed analysis, without making the approximations used in Barlow & Tripathy, shows that human performance is still much worse than Ideal Observer even if the model is degraded. (Lu & Yuille)

However human perfomance is very similar to the perfomance of the slow-and-smooth model desurbed earlier.

visual perception

slow-and-smooth ...

<u>Interpretation</u>: Human visual perception is ideal for the types of stimuli that they see in everyday life. And not the types of stimuli shown in scientists laboratories.

This relates to studies of human rationality. Experiments show that human decision making - in laboratory settings - is not consistent with Bayes decision theory.

But maybe humans are optimal for situations which are important to them and which they perform frequently.

E.g. Racing bookmakers are probably optimal - if not, they go bankrupt. (But there are tricks they employ to achieve this.)