Motion Perception

- The barberpole illusion shoes that the perception of motion is not straightforward. The barberpoles rotate to the right, but the perception of motion is vertically upwards. This is because there may not be enough information to determine the motion unambiguously.
- 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 called the *aperture problem*. At the endpoints of the bar the motion will seem to be unambiguous. But the observations at the endpoints have to propagate to the other points on the bar. How is this done? How far can information at unambiguous points be propagated?
- Consider a rotating ellipse. This has no ambiguous points. It is perceived either as: (i) a non-rigid rotating ellipse, or (ii) a rigid circle rotating in 3D. But, surprisingly, it is not seen as a rigidly rotating ellipse unless the aspect ratio is very big (because now it appears to have endpoints).
- Nakayama performed a series of experiments which studied the effects of occluding the endpoints of the bars (so that the ends are not visible), or having isolated dots move near the bar, and other variants. This showed that perception is very subtle and that the motion of the bar could be *captured* by the motion of the dots. Many of these effects could be modelled by the motion coherence theory (see Yuille and Grzywacz handout) which claims that motion is perceived by combining the local measurements with a prior that the motion is slow and smooth.

Short-range and long-range motion

- The human eye receives input images as a continuous stream in time $I(\vec{x}, t)$, where t is continuous. Typical videos have 24 frames per second (recent movies, e.g., the Hobbit, tried 48 frames per second but moviegoers complained that this looked weird). Curiously dogs may not be able to see motion at 24 frames per second, but humans can. Humans can even perceive motion with far fewer frames per second.
- Short-range motion is defined to be situations where the image frames are so close together that it is difficult to distinguish them from continuous frames. This will means that the intensity is differentiable (as we will discuss later). In such cases we have the aperture problem.
- Long-range motion occurs there is a significant difference between neighboring image frames. In this case the images are not differentiable. Instead we have to solve a correspondence problem between features in the two images. This is similar to the binocular stereo correspondence problem, except it does not have an epipolar line constraint.
- For short-range motion, we assume that $I(\vec{x}, t) = F(\vec{x} - \vec{v}t)$ where $\vec{v}$ is the motion and we assume that the motion is locally rigid in the image plane with intensity $F(.)$. Differentiating $I(\vec{x}, t)$ with respect to $\vec{x}$ and t gives: $\vec{\nabla} I(\vec{x}, t) = \vec{\nabla} F(\vec{x} - \vec{v}t)$ and $\frac{\partial I(\vec{x},t)}{\partial t} = -\vec{v} \cdot \vec{\nabla} F(\vec{x} - \vec{v}t)$. This yields the *optical flow equation* $\vec{v} \cdot \vec{\nabla} I(\vec{s}, t) + \frac{\partial I(\vec{x},t)}{\partial t} = 0$. In other words, we can directly measure the motion component in the direction of the image gradient, but we do not know the perpendicular component.

Short-range

- When implementing a model on an image lattice – and with discrete time frames – we must approximate the derivatives by differences. I.e. $\frac{\partial I(\vec{x},t)}{\partial t} \approx \frac{I(\vec{x},t+\Delta)-I(\vec{x},t)}{\Delta}$ where Δ is the difference between two adjacent time frames. (The true derivative would be to take the limit as $\Delta \mapsto 0$).
- This approximation is okay of Δ is small (relative the to rate of change of $I(.,.)$). But it becomes problematic if the image $I(\vec{x},t)$ is a rapidly changing function of $\vec{x}$. This can be reduced by smoothing the images with respect to $\vec{x}$ and yields an algorithm to estimate the optical flow (Black and Anandan) which first estimating the optical flow on the smoothed image and then makes a correction for the true image.
- This is because the optical flow equation can be derived by matching points between image frames so that $I(\vec{x}+\Delta\vec{v}(\vec{x}),t+\Delta) = I(\vec{x},t)$. We do a Taylor series expansion, about $\vec{v}=0$, to express $I(\vec{x}+\Delta\vec{v}(\vec{x}),t+\Delta) = I(\vec{x}) + \Delta\vec{v}(\vec{x})\cdot\vec{\nabla}I(\vec{x},t) + \Delta\frac{\partial I(\vec{x},t)}{\partial t}$. This yields the optical flow equation $\vec{v}(\vec{x})\cdot\vec{\nabla}I(\vec{x},t) + \frac{\partial I(\vec{x},t)}{\partial t} = 0$.
- We can proceed to make a series of expansion on smoothed images $G(\vec{x};\sigma) * I(\vec{x},t)$, where $G(\vec{x};\sigma)$ is a Gaussian with variance σ^2. Starting with large σ, i.e. with a smoothed image, we can estimate an optical flow $\vec{v}^*_\sigma$ by minimizing an energy function (see next slide). Then we can do a Taylor series expansion at a smaller scale (smaller σ) around $\vec{v}^*_\sigma$. I.e. we assume $\vec{v} = \vec{v}^*_\sigma + \Delta\vec{v}$, where $\Delta\vec{v}$ is a small correction (so the Taylor expansion is likely to be a valid approximation even if $I(\vec{x},t$ is not smooth).

Short-range (2)

- The advantage of using the optical flow equation (with or without smoothing) is that it takes the velocity variable $\vec{v}$ out of the argument of the function $I(\vec{x}, t)$. This is very helpful for short-range optical flow.
- First recall that the optical flow equation is unable to specify the full optical flow (because only the component in the image gradient direction is known). So Horn and Schunk proposed a smoothness constraint. Namely, estimate the velocity which minimizes $E[\vec{v}(\vec{x})] = \int \{\vec{v}(\vec{x}) \cdot \vec{\nabla} I(\vec{x}, t) + \frac{\partial I(\vec{x},t)}{\partial t}\}^2 d\vec{x} + \lambda \int \frac{\partial \vec{v}(\vec{x})}{\partial \vec{x}} \cdot \frac{\partial \vec{v}(\vec{x})}{\partial \vec{x}} d\vec{x}$.
- The first term imposes the optical flow equation while the second term biases it an optical flow which is spatially smooth, i.e. where $\frac{\partial \vec{v}(\vec{x})}{\partial \vec{x}}$ is small. Horn and Schunk discretized this function (i.e. replaced derivatives by differences) which results in a quadratic energy function which is convex and can be minimized by steepest descent (note: if the optical flow equation is not used then $\vec{v}$ would be an argument of the intensity $I(\vec{x}, t)$ and the resulting formulation would be highly non-convex).
- Note: this can be discretized to yield the standard energy formulation for Markov Random Fields. These is a data term (the first term) which depends only on the local velocity $\vec{v}(\vec{x})$ and a prior term (the second) which depends on the local velocities.

Short-range (3)

- Horn and Schunck used a regularization/prior which penalized the first order derivatives. This can be augmented in two ways: (I) Penalize higher order derivatives, which imposes more "rigid" smoothing, by adding terms such as $\int \frac{\partial^2 \vec{v}(\vec{x})}{\partial \vec{x}^2} \cdot \frac{\partial^2 \vec{v}(\vec{x})}{\partial \vec{x}^2} d\vec{x}$. (II) Penalize the velocity itself by adding a term $\int \vec{v}(\vec{x}) \cdot \vec{v}(\vec{x}) d\vec{x}$.
- These types of smoothness and slowness functions have been effective for modeling human perception of optical flow (Yuille and Grzywacz, Weiss and Simoncelli, etc). Slowness alone is able to account for some perceptual phenomena whole slowness combined with smoothness accounts for others. Indeed, it could be shown (Yuille and Grzywacz) that these theories could even account for the fact that motion capture decreases with distance in agreement with Nakayama's experiments (this is a technical result which involves showing that these problems could be expressed in terms of the Green's functions of differential operators).
- There is no need, of course, for the regularization terms to be quadratic in $\vec{v}$. It is highly desirable that the energy is convex, but this can be ensured by using L_1 norms instead of L_2.
- These types of optical flow models are now highly effective after much engineering (e.g., modifying the data term). They are, of course, being replaced by deep neural networks provided supervised datasets are available (not easy for optical flow), see Carlo Tomasi. An alternative (Zhe Ren et al) is *unsupervised optical flow* where the regularization term is used as the loss function (c.f. Smyrnakis and Yullle).

Long-range motion

- Minimal mapping theory (Ullman) formulates long-range motion as finding the correspondence between points/dots $\{\vec{y}_a\}$ in the first image to points $\{\vec{x}_i\}$ in the second (all the dots are indistinguishable).
- There is a correspondence variable $\{V_{ai}\}$ so that $V_{ai} = 1$ if the point at $\vec{y}_a$ in the first image is matched to the point $\vec{x}_i$ in the second image. $V_{ai} = 0$ otherwise, and we impose conditions that all dots in the first image must be matched to exactly one dot in the second image (can be relaxed if the number of dots in the two images is different).
- Minimal mapping theory finds the V^* which minimizes $E(V) = \sum_{a,i} V_{ai}|\vec{y}_a - \vec{x}_i|$. This can be solved by linear programming. This makes a slowness assumption which tries to match dots which are in similar positions (i.e. where $|\vec{y}_a - \vec{x}_i|$ is small).
- An alternative model (Yuille and Grzywacz) gives better fot to human experiments but also relates more closely to the short-range motion theories. It suggests to minimizing $E[V, \vec{v}] = \sum_{a,i} V_{ai}\{\vec{y}_a - \vec{x}_i - \vec{v}(\vec{x}_i)\}^2 + \lambda \int \vec{v}\cdot\vec{x}d\vec{x} + \mu \int |\frac{\partial \vec{v}(\vec{x})}{\partial \vec{x}}|^2 + \nu \int |\frac{\partial^2 \vec{v}(\vec{x})}{\partial \vec{x}^2} \cdot \frac{\partial^2 \vec{v}(\vec{x})}{\partial \vec{x}^2}|^2 d\vec{x}$.
- The idea is that if point at $\vec{y}_a$ is matched to $\vec{x}_i$, then this gives a local estimate of displacement $\vec{y}_a - \vec{x}_i$. We hence seek to match the points so that the interpolated optical flow (interpolated by the slowness and smoothness constraints) is as slow and smooth as possible (i.e. minimizes the energy $E[V, \vec{v}]$. Hence, from this perspective the only difference between short- and long-range data is the data term.

Long-range motion and ideal observers

- Barlow and Tripathy studied human perception of long-range motion. They used an experimental setup where the first image consisted of a set of random dots. The second image was generated by taking a subset of the dots in the first image and moving them horizontally by a fixed amount (randomly chosen) and filling up the rest of the image with randomly placed dots.
- They then tested human performance at several visual tasks, e.g., judging whether the dots moved to the right or to the left, and they compared performance to an *ideal observer model* that knew how the two images had been generated (statistically). Not surprisingly, humans performed much worse (by many orders of magnitude to the ideal observer).
- But the slow-and-smooth gives a much better fit to human observers (H-J Lu and A.L. Yuille). Presumably because slow-and-smooth is a prior assumption that applies to many images, while there is no reason to believe that human observers have any knowledge of the statistics of the stimuli in Barlow and Tripathy's experiments (or have the ability to learn them).
- So human observers are probably adapted to the statistics of the environment, not to the statistics of the data they are shown in a research laboratory.

Long-range motion and computer vision

- Long-range matching is not of great interest to computer vision researchers at present.
- Typically researchers work with images where the time frames are fairly close together so the optical flow assumptions are valid (particularly after smoothing). But there are cases where this is not valid, e.g., consider a game of tennis, where the spectators move very little, the players move faster, and the tennis ball moves extremely fast. Smoothing the images (e.g., Black and Anandan) will yield good motion for the spectators and the players, but will work badly for the tennis ball. It moves too quickly and is too small, so blurring the image will tend to remove it.
- An alternative strategy for objects moving rapidly is to detect a small set of *interest points*. If this set is small, and the interest points have attributes, then the correspondence problem is fairly unambiguous and can be solved by nearest neighbour methods.

Long-range motion and shape matching

- Theories of long-range matching can be applied to matching shapes of objects, where the velocity is replaced by the displacement between points on the two shapes, e.g., Rangarajan and Chui for digits. A variant (Belongei and Malik) extract *shape context features* from the shapes and then matches them using a cost function similar to minimal mapping, but with a similarly terms depending on their similarity of appearance.
- More recent models, e.g., Jiayi Ma et al., have models that express the displacement between shapes as linear sums of radial basis functions (e.g, the formulation by Yuille and Graywacz) and give very good performance on a range of tasks including shape matching and detecting shapes in images.