Lecture 20: Optic Flow

Reading assignment: TMB2 7.2

- If, as we walk forward, we recognize that a tree appears to be getting bigger, we can infer that the tree is in fact getting closer.

- J. J. Gibson emphasized that it does not need object recognition to make such inferences. Much information about the position, orientation and overall movement of the organism relative to its environment can be "picked up" by "low-level systems"
Optic flow is the vector field representing velocity on the retina of points corresponding to particular points of the environment.

Gibson observed that optic flow is rich enough to support the inference of where collisions may occur and the time until contact/adjacency.

Controversially, Gibson talks of direct perception of these visual affordances that can guide behavior without invocation of "high-level processes" of object recognition.

We accept the importance of such affordances but stress that neural processing is required to extract them from retinal signaling.
Using a planar retina, we analyze monocular information.

\[ \xi = \frac{ax}{z} \quad \text{and} \quad \eta = \frac{by}{z} \quad (1) \]

where \( a \) and \( b \) are positive scale constants.

For simplicity, we set \( a = b = 1 \).

The retinal displacement is based only on relative motion:

The same flow will result from the organism moving with some velocity in a stationary environment; or from the environment as a whole moving with the opposite velocity about a stationary organism.

Let the relative velocity of the organism = \((u,v,w)\)

\[ x/ u = \text{left}; \quad y/ v = \text{up}; \quad z/ w = \text{direction of gaze} \]
Assume $v = 0$ (no vertical motion) but do not assume $u = 0$
(i.e., motion need not be along the direction of gaze)

The spatial coordinates of a point $P$ will change from
$(x,y,z)$ at time $0$ to $(x-ut, y, z-wt)$ at time $t$.

Since $\xi = x/z$ and $\eta = y/z$ \hspace{1cm} (1)
the retinal coordinates take the form

$$\begin{align*}
(\xi(t), \eta(t)) &= (\frac{x-ut}{z-wt}, \frac{y}{z-wt}) \quad (2)
\end{align*}$$

Extrapolating back to $-\infty$:

$$\begin{align*}
(\xi(-\infty), \eta(-\infty)) &= (\frac{u}{w}, 0) \quad (3)
\end{align*}$$
i.e., all the flow trajectories appear to radiate from a single point \( \left( \frac{u}{w}, 0 \right) \), the **focus of expansion (FOE)**

Since \( \eta(-\infty) = 0 \), this point is on the horizon, the retinal projection of the horizon point toward which the organism is locomoting.

---

**Note 1:** The retinal displacements of objects moving with different relative velocities have different FOEs

**Note 2:** The above analysis applies only to linear motion (translations).
Recall:

\[(\xi(t), \eta(t)) = (\frac{x-ut}{z-wt}, \frac{y}{z-wt}) \quad (2)\]

\[(\xi(-\infty), \eta(-\infty)) = (\frac{u}{w}, 0) \quad (3)\]

we see that the trajectory (2) is in fact a straight line whose slope is:

\[\frac{(\xi(t) - \xi(-\infty))/ (\eta(t) - \eta(-\infty))}{yw/(x - ut)w - u(z - wt)} = \frac{yw}{xw - uz} \quad (4)\]

The trajectory is straight because this is independent of t.
The line is thus vertical (slope = $\infty$) only if

$$xw - uz = 0$$
$$\frac{x}{u} = \frac{z}{w}$$
i.e., just in case the corresponding surface point is on a collision course with the organism.
Since the organism has non-zero width, even an object on a non-vertical optic flow line may be on collision course.

Let the organism have width 2k.
Let the direction of locomotion coincide with the direction of gaze so that x is constant: \( u = 0 \) and \( x \equiv x_0 \)

There will be eventual collision of the organism with the texture element at \((x_0, y_0, z_0)\) just in case \(|x_0| < k\).

When \( u = 0 \), (2) tells us that
\[
ξ(t) = \frac{x_0}{z_0 - wt} \tag{6}
\]
and thus the corresponding retinal velocity is
\[
\dot{ξ}(t) = \frac{wx_0}{(z_0 - wt)^2} = \frac{w}{x_0} ξ(t)^2 \tag{7}
\]
\[
\dot{\xi}(t) = \frac{w x_0}{(z_0 - wt)^2} = \frac{w}{x_0} \xi(t)^2 \tag{7}
\]

and hence
\[
x_0 = \frac{w \xi(t)^2}{\dot{\xi}(t)} \tag{8}
\]

and so the condition for collision is
\[
\left| \frac{w \xi(t)^2}{\dot{\xi}(t)} \right| < k \tag{9}
\]

which (given the value w) depends only on the optic flow.
For a given $\xi(t)$

the closer the texture element, the larger is its $\dot{\xi}(t)$. That is, for a given visual angle

the further the element, the smaller its associated flow.

The condition $\frac{w_\xi(t)^2}{\dot{\xi}(t)} < k$ becomes:

Collision occurs if $|\dot{\xi}(t)| > \frac{W}{k} \xi(t)^2$  \hspace{1cm} (11)

— the smaller is $\dot{\xi}(t)$, the less likely is a collision.
As noted earlier, a problem glossed over in Gibson's writings is that of the actual computation of the optic flow from the changing retinal input.

We here offer the MATCH algorithm which is:
- played out over a number of interacting layers
- each with parallel interaction of local processes

where the retinal input is in the form of
two successive "snapshots"

and the problem is to match up corresponding features in
these two frames.
This is like stereopsis, BUT:

- In optic flow the two images are separated in time, and so depth is expressed as time until contact or adjacency.

- There are only two eyes, but there are many times: the initial algorithm for matching a pair of frames can be improved when the cumulative effect of a whole sequence can be exploited.

The correspondence / stimulus-matching problem is to match up each pair of features in the two frames that correspond to a single feature in the external world.
The problem still persists, in a somewhat modified form, with continuous processing:

**The Aperture Problem**

Flow as seen through the aperture

True Flow

cf. Movshon et al. on the transition from MSTS to MT.
The MATCH algorithm (Prager and Arbib) makes use of two consistency conditions:

**Feature Matching:**
Where possible, the optic flow vector attached to a feature in Frame 1 will come close to bringing it in correspondence with a similar feature in Frame 2.

**Local Smoothness:**
Since nearby features will tend to be projections of points on the same surface, their optic flow vectors should be similar.
"In the style of the brain", a retinotopic array of local processors

- make initial estimates of the local optic flow

- then repeatedly pass messages back and forth to their neighbors in an iterative process

- to converge upon a global estimate of the flow.

Iterations are needed if a correct global estimate is to be obtained:
The algorithm works quite well in giving a reliable estimate of optic flow within 20 iterations.
**Prediction**

If, having matched Frame $n$ to Frame $n+1$ we try to match Frame $n+1$ to $n+2$ it is reasonable to assume that to a first approximation the optic flow advances a feature by roughly the same amount in the two interframe intervals.

If we use repetition of the previous displacement to initialize the optic flow computation of the two new frames, only 4 or 5 iterations, rather than the original 20, are required, and the quality of the match on real images is improved.
Bringing in World Knowledge

To survive in a world in which we move relative to moving objects our brain must be so structured that the features activated by stimuli from a certain object must

1. not only remain activated while the object remains a relevant part of our environment
2. but must change in a way that is correlated with the movement of the object relative to the effectors

[cf. the Slide-Box Metaphor, §2.1]
[cf. experiments by Shepard and Metzler 1971 on "mental rotation", §5.2]

Back to Optic Flow: High level motion models (e.g., recognition of rotation) can further improve prediction:

With sufficient high-level knowledge, repeated iterations of low-level relaxation are replaced by one-step "confirmation".

Note: There are more "feedback" axons from visual cortex to lateral geniculate than there are "direct" axons from retina to lateral geniculate. [HBTN N: Thalamus]
An Evolutionary Perspective

The MATCH algorithm is based on two consistency conditions:
- feature matching and
- local smoothness.

We can design edge-finding algorithms which exploit the breakdown of consistency conditions to find edges in two different ways,
- on the basis of occlusion/disocclusion, and
- on the basis of optic flow "discontinuity" (i.e., a high value for the gradient).

To the extent that the estimate of edges by these two processes is consistent, we have cooperative determination of the edges of surfaces within the image.
As good edge estimates become available, the original basic algorithm can be refined:

Instead of having "bleeding" across edges (blue curve)

\[ \text{dynamically change the neighborhood of a point shown by red dividers} \]

so that the matching of features or the conformity with neighboring flow can be based almost entirely upon features on the same side of the currently hypothesized boundary — but not entirely to yield the "correct map" (shown in black).

The "dividers" correspond to the idea of a line process introduced later and independently by:


**VLSI implementation:**

We thus see an "evolutionary design process":

- The basic algorithm (1)
- provides new information which can then be exploited in cooperative segmentation algorithms (2),
- but once the segmentation information is available, the original algorithm can be refined by the introduction of segmentation-dependent neighborhoods (3).

This evolutionary design process is not simply an engineering observation, but gives us some very real insight into the evolution of the brain:
- An evolutionarily more primitive system allows the evolution of higher-level systems; but then
- return pathways evolve which enable the lower-level system to evolve into a more effective form.

Evolution was a key concept of 19th century thought. Hughlings Jackson viewed the brain in terms of levels of increasing evolutionary complexity.

Jackson argued that damage to a "higher" level of the brain caused the patient to use "older" brain regions in a way disinhibited from controls evolved later, to reveal behaviors that were more primitive in evolutionary terms.
Our Figure, then, offers a Jacksonian analysis:

- hierarchical level
- return pathways
- evolutionary interaction
- evolutionary degradation under certain lesions exemplified in a computationally explicit model based on cumulative refinement of parallel interaction between arrays.

Thus evolution not only yields new brain regions connected to the old, but yields reciprocal connections which modify those older regions.

Self-study: The subsection of 7.2 entitled "Machine Vision for Motion Information."

Read

in relation to the cooperative interaction of multiple processes.
Shape-From-X
It is also possible to generate depth maps directly from monocular color or intensity images. These methods, are known as shape-from-X, where X may be shading, texture, or contour.

The general idea behind shape-from-shading, for example, is to assume a model of the image formation process which describes how the intensity of a point in the image is related to the geometry of the imaging process, and the surface properties of the object(s).

This model can be used with two constraints on the underlying surface to recover the shape:

1. that each image point corresponds to at most one surface orientation (the uniqueness constraint); and
2. that orientations of the underlying surface vary smoothly almost everywhere (the smoothness constraint).

Read
M. Bertero, T. Poggio and V. Torre, 1988, Ill-Posed Problems in Early Vision, Proceedings of the IEEE 76: 869-889;
HBTNN: Regularization Theory and Low-Level Vision for the general mathematical framework of "Regularization Theory" using constraints to solve an ill-posed inverse problem.