MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Vision Memo.
A.I. Memo. No. 122. March 1967.

Remarks on Correlation Tracking

Marvin L. Minsky.

1. Theoretical discussion:

The problem is to track the motion of part of a field of view. Let us assume that the scene is a two-dimensional picture in a plane perpendicular to the roll axis. (These simplifying assumptions, of course, are a main problem in estimating how the system works in real life). So we can think of the picture as a function $f(x,y)$ in some plane.

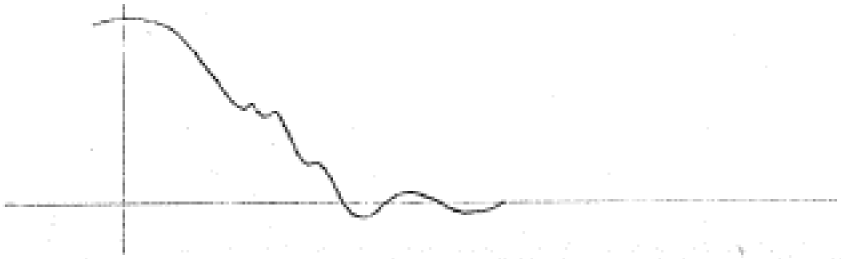
Now suppose that at time t_o the scene is $f_o(x,y)$ and at some time later it has moved, and is $f_t(x,y)$. Suppose also that the scene has not changed, but has only been moved rigidly in the plane. Then an elegant mathematical way to estimate this motion is to compute the cross-correlation of the original and current picture. First let us review a basic simple mathematical fact. Given any function $f(x)$ and any displacement Δ, it is true that

$$\int_{-\infty}^{\infty} f(x)\ f(x) \geq \int_{-\infty}^{\infty} f(x)\ f(x+\Delta)$$

(when the expressions are meaningful) and we have $>$ rather than $=$ except under the most peculiar conditions, i.e., when the pattern is perfectly periodic. (This fact is a consequence of the triangle inequality $a^2 + b^2 \geq c^2$, slightly generalized). In fact, if one considers

$$\phi\ (\Delta) = \int f(x)\ \ f(x + \Delta)$$

one always gets a graph like



with its maximum at 0 and symmetrical on both sides. Now, consider some small number d, called the "delay", and consider the formula

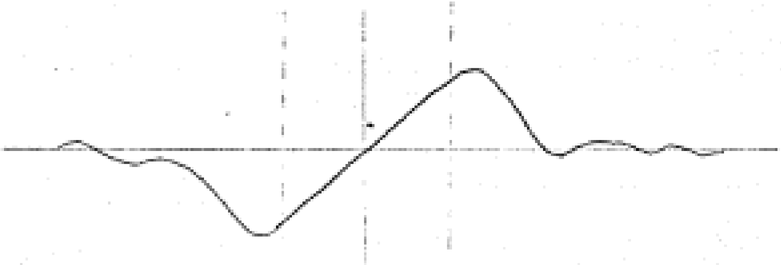
$$g(h) = \int [\ f(x+d) - f(x-d)\ ]\ f(x+h)$$

If we look at the values of this for various values of h, we find:

$$g(d) = \int f(x+d)\ f(x+d) - \int f(x-d)\ f(x+d) = \phi(o) - \phi(2d) > 0$$
$$g(o) = \int f(x+d)\ f(x) - \int f(x-d)\ f(x) = \phi(d) - \phi(d) = 0$$
$$g(-d) = \int f(x+d)\ f(x-d) - \int f(x-d)\ f(x-d) = \phi(2d) - \phi(o) < 0$$

(because $\int_{-\infty}^{\infty} f(x+a)\ f(x+a+b) = \int_{-\infty}^{\infty} f(x)\ f(x+b)$ in general).

and assuming everything is nice and smooth, we can expect to get a curve like this:



yielding a "discrimination curve" which is just what we want, because in the "linear" range it computes the displacement of f(x) from f(x+h) for us! Now it is of practical importance for us to know how large is the "linear" range because

(1) it tells us how large a step a device could cope with and

(2) it incidentally tells us how accurate or coarse the computation of the integrals must be - this bears on the size of optical slit - or whatever is used in practice.

Now we can find a lot about this by (theoretically) considering a very small displacement d - in fact, if d is allowed to go to zero, we get a sort of differential analysis: we divide by d to keep things from blowing up, and we get:

$$\lim_{d \to o} \frac{1}{d} \int [\, f(x + d) - f(x - d)] \; f(x + h)$$

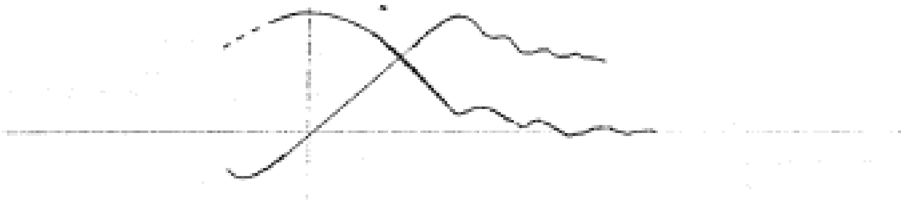
$$= 2 \int \frac{f(x + d) - f(x - d)}{2d} \quad f(x + h)$$

$$= 2 \int f'(x) \; f(x + h)$$

$$= 2 \, \phi' \; (h)$$

by some analysis involving $\frac{d\phi(h)}{dh} = \int f'(x)f(x + h) + f'(x)f(x + h)$,
and the fact that $\phi(h) = \phi(-h)$.

This says that as the delay d goes to zero, the "discrimination curve" approaches the derivative of the auto-correlation function $\phi(h)$



and this suggests that there is no serious risk if we take d to be of the order of ½ the width of the main peak of the auto-correlation function. In terms of the picture-function f(x,y), this width is a measure of the "grain" of the scene-texture; it is probably the most important ("first-order") parameter for characterizing scenes in terms suitable for correlation-based instruments.

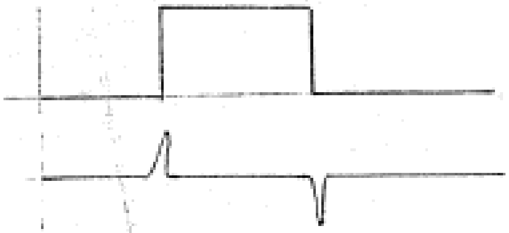
What is the intuitive significance of the tracking function?

$$\int f'(x) \; f(x + h) \;\; = \phi'(h)$$

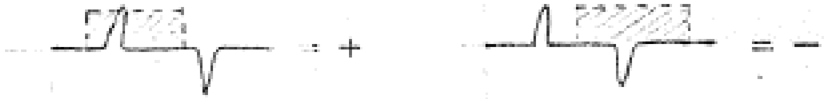
Two simple examples show what is happening:

1. f(x) =

f'(x) =



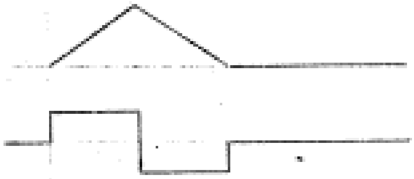
so if h is positive, or negative we get



and here, for a square wave, we get simply the <u>sign</u> of the displacement! on the other hand,

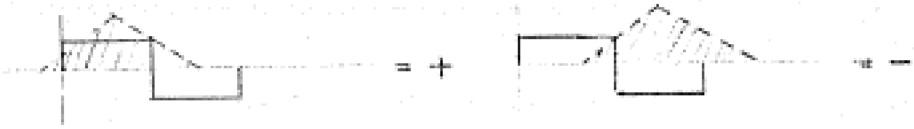
2. f(x) =

f'(x) =



h positive

h negative



and we get something more proportional to the amount of the displacement. Under the assumption that the lengths of regions of positive and negative derivatives are generally equal, we can approximate this "analogue" situation by differentiating and "clipping" the results; the error in this depends very much on the situation and sometimes the "digitized" result will be better, usually worse, than the analogue.

2. <u>Polar coordinates and axial tracking</u>.

Now let us consider a more practical situation; we consider the problem of tracking, still with the same assumptions, but assuming the scene is on a sphere so that we want to compute pitch, yaw, and roll. Now let us replace f(x,y) by F(r,θ) and consider a polar integral

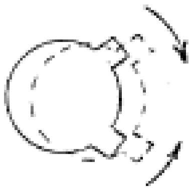
$$\int_0^\infty K(r) \int_0^{2\pi} Q(r,\theta)\, J(\theta)\, d\,\theta dr$$

where K and J are "hardware" functions we have to consider. Suppose first that $J(\theta) = 1$ and $K(r)$ falls off fast enough to make the integral converge (that is, a "finite radial angle shaded slit." The slit width is concealed in the rest of the mathematics). Then if the scene is simply rotated, we use for $Q(r,\theta)$

$$Q(r, \theta) = F^1(r,\theta)\, F(r,\theta + \alpha)$$

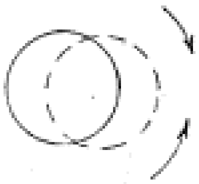
and this gives us an estimate of the phase, α , just as in the first section.

If the scene suffers a <u>translation</u>, in pitch or yaw, the effect on this averages out to zero because we get opposite effects in each hemisphere:



but this is a statistical cancellation and is an error that must be reckoned on! Thus there will be "cross talk" if the picture is not symmetric, in a suitable average sense.

Now, to get the yaw-component, we use $J(\theta) = \sin\theta$ and get the effects above to add!



$+$ (advance) $\sin \frac{\pi}{4} = +$

$-$ (retard) $\sin \left(-\frac{\pi}{4}\right) = +$

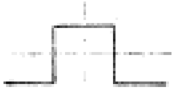
In this case the contribution of the pitch component averages to zero because for it we get $\cos(\theta) = J(\theta)$:

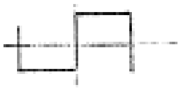
$$+\ \text{(advance)}\ \cos\left(\frac{\pi}{4}\right) = +$$

$$-\ \text{(retard)}\ \cos\left(-\frac{\pi}{4}\right) = -$$

and these cancel.

Thus it is possible to resolve the three components roll ($J(\theta) = 1$), pitch, ($J(\theta) = \cos\theta$), and yaw ($J(\theta) = \sin\theta$) by changing the "kernel" of the integral. One could even approximate this digitally by using

$\cos\theta \equiv$ 

$\sin\theta \equiv$ 

as approximations, but this would increase the "cross talk."

3. The Assumptions

It remains to discuss the assumptions:

1. Stereo changes mean that throughout, we never really have

   $$f(x + h)$$

   but we always have a new picture

   $$g(x + h) \approx f(x + h) + \text{error}$$

2. It is safe to say that because the situation is real, there is no trouble with the analytic mathematics. One has to worry only about the photon and other noise involved in empirical evaluation of the integrals and derivatives.

3. Although the auto-correlation function is always symmetrical

   $$\int f(x)\, f(x + h) = \int f(x)\, f(x - h)$$

   this is only approximately true for a new scene

   $$\int g(x)\, f(x + h) \neq \int g(x)\, f(x - h)$$

   and a systematic drift during a changing scene could cause a cumulative error.

4. The most serious problem is: To what extent will a theoretical calculation based on a good smooth, convex maximum of autocorrelation function be realistic? It will be valuable in calculating the underlying basic random-variable drift of the system, but this theory has to be supplemented by a special-case analysis of the worst "deterministic" patterns known or suspected to cause trouble.

We should do some experiments using correlation on very simple arrays. Even, say, 16 points around a circle should yield interesting results!