

1. Consider the “parking lot surveillance problem,” where a computer vision system is used to watch a large parking lot overnight and set off an alarm if an individual is detected who remains on foot in the parking lot for more than 3 minutes.

(a) Explain clearly why a design based on the Marr paradigm will very likely cost more and perform less well than one based on the newer paradigms of active or animate vision.

(b) How consistent is the Goad algorithm with the Marr paradigm? Give it a letter grade A-F on consistency with the Marr approach, explaining your reasoning.

(c) The active/animate approach is not always better than the Marr approach. Describe a specific vision application for which a Marr approach is likely to succeed as well as or better than an active/animate approach. Also, point out some general attributes of a vision application that would make the Marr approach well suited.

(a) Some problems with the Marr approach:

1. Recovery is expensive in time and resources, and unnecessary. You don't need to recover the shapes and locations of all objects, just find moving, vertically-oriented objects that might be humans on foot.

2. Passive vision is inefficient. If you have fixed cameras or cameras sweeping a field of view with a fixed passive sweep pattern, you would need many cameras to cover the parking lot sufficiently to be able to find and then track suspect objects. If you have foveal cameras with large peripheral fields of view you can saccade to suspects and track them easily with only a few cameras.

3. If a camera platform could move around, say on a track suspended over the parking lot, less cameras would be needed than if the camera was passively fixed.

(b) Goad algorithm has some attributes inconsistent with the Marr paradigm:

1. Rather than recover the object's detailed shape, it uses a polyhedral model in which any object is represented just as a set of connected vertices. This is not recovery..

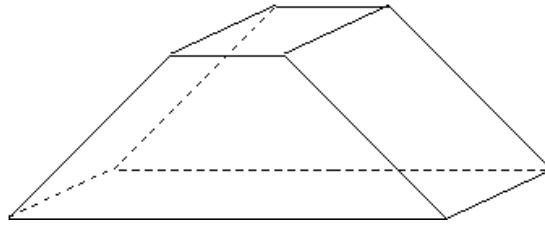
2. Information that would be required for recovery is ignored, eg. color, texture.

3. Decision-making is active, ie. next view is planned.

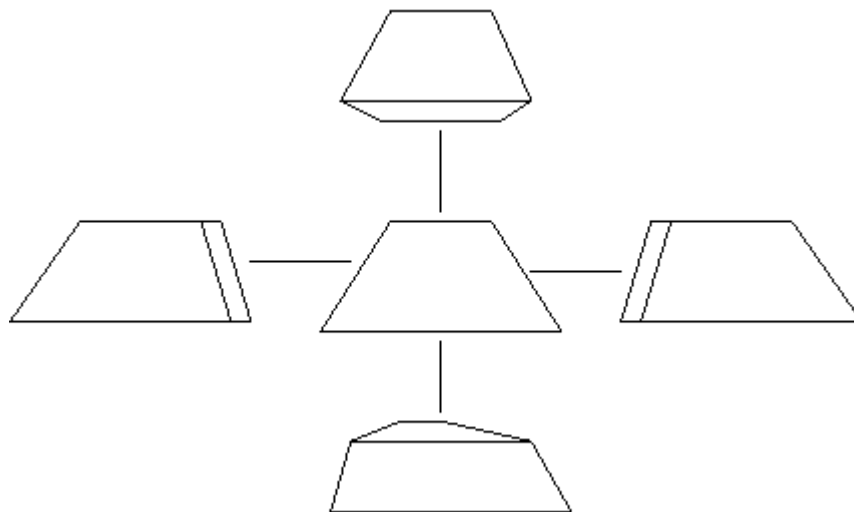
A fair grade at consistency with Marr might be anything D+ or lower.

(c) Marr approach works best for 2-D static scenes with many small highly detailed regions of interest which require recovery. For instance, if you have a satellite image and want to label all pixels semantically: road, field, house, car, tree, river, etc. Here there would be no advantage to active or animate vision, and the extra complexity of fovea and gaze control and camera view planning would be an increased cost without reward compared to a passive, recovery, Marr system.

2.



(a) Adjacent aspects differ by just a single face. Starting from the aspect revealing only the front trapezoidal face, there are four aspects with two faces corresponding to the front plus any other face but the back. This part of the aspect graph looks like



(b) There is one set of topologically equivalent characteristic views with 1 face, one with 2 faces, and two with 3 faces. The 3 faces can be all adjacent, eg. top, front and one side are visible, or they not all adjacent, as when the left side, top and right side are visible. There is one with four faces (both sides, top and either front or back). So the answer is 5.

(c) Call the six faces f, b, l, r, t, o for front, back, left, right, top and bottom.

1. Aspects with one face: f, b, l, r, t, o, =6
2. Aspects with two faces: fl, fr, ft, fo; bl, br, bt, bo; lt, lo, rt, ro = 12
3. Aspects with three faces: frt, flt, fro, flo; brt, blt, bro, blo; tlr = 9
4. Aspects with four faces: fltr, bltr =2

So total number of nodes in aspect graph = 6+12+9+2=29.

3. You are in car driving in the x -direction, you point your camera out the window in the z -direction and take a picture of a stationary scene (no independent object motion in the scene) at time t . The brightness $E_1(x,y)$ recorded on this image is found to $E_1(x,y) = \cos^2 x$, that is, its image plane brightness does not vary in the vertical y -direction, and varies in the horizontal direction with x as cosine-squared. Then one second later at time

$t+1$ you take another image $E_2(x,y)$, and find that at $x=\pi/4$, $y=1$ the image has brightness $E_2(\pi/4,1)=1/4$.

(a) Find the normal component of optical flow at the point $x=\pi/4$, $y=1$.

(b) The normal component and the total optical flow at the point $x=\pi/4$, $y=1$ at time t are the same. Explain why.

(a) $\partial E/\partial x = 2 (\cos x) (-\sin x)$ so at $x=\pi/4$ this is $\partial E/\partial x = -1$. Also, $\partial E/\partial y = 0$ everywhere. Now estimate $\partial E/\partial t$ at $(x,y)=(\pi/4,1)$ by the forward difference

$$(E(t+\Delta t)-E(t))/(\Delta t) = (1/4 - \cos^2(\pi/4))/(1) = -1/4$$

So the normal component is

$$v' = \frac{-E_t}{\|\nabla E\|^2} \nabla E = \frac{-1/4}{1} \begin{bmatrix} -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/4 \\ 0 \end{bmatrix} \text{ mm/s}$$

(b) The component of the optical flow orthogonal to the image gradient is missing when the normal component is calculated. But in our case, since the car is driving in the x direction and the scene is still, there is no y -component of optical flow, and $v=v'$.