CPS331 Lecture: Vision last revised October 10, 2018

Objectives:

1. To note reasons why vision is a particularly challenging AI problem
2. To overview the three levels of visual processing

Materials:

1. Peephole demonstration with text and a picture
2. Projectable of structure of the human eye
3. Projectable of Sun and Triangle Illusions (based on Tanimoto figures 10.4(e), (f)
4. Projectable of Cawsey figure 6.1
5. Image Editor project (used in the past in CPS112 and 122) with parrots image for edge detection demonstration
6. Projectables of Winston (2e) figures 10.11, 10.12, 10.13
7. Projectable of partial line drawing showing ambiguity when only part seen
8. Projectable of complete line drawing to be labeled
9. Projectable of Nilsson (1998) Figure 6.16
10. Projectables of progressive labeling of #6 using Waltz procedure
11. Projectables of two interpretations of line drawing with shading
12. Handout problem with possible vertex labelings and figure to label, plus filled in version to project

I. Introduction

A. Much of what we know comes to us through one of two means: natural language, and vision. (Of course, natural language can, in turn, be perceived through either reading or listening - but we are here distinguishing it from strictly visual material).

1. Textbooks typically include both natural language text and figures.
2. Class lectures typically include both spoken natural language and projected graphics.
3. TV and movies typically make use of both spoken language and sight.
4. We say "A picture is worth 1000 words".

B. Of course, the way we perceive these two sources of information is really quite different.

1. How would you contrast the way we handle natural language and visual information?

   ASK

2. A key difference is that natural language is basically sequential, while vision is basically parallel/holistic.

   Demo: Peephole with text, picture.

C. Vision seems to be closely connected with the idea of "understanding" as distinct from mere textual manipulation.

Consider the way we use "vision" terms in conversation.

1. "Visionary"

2. "The big picture"

3. To understand something is to "see" it.

D. Actually, there are some amazing things that happen in the process of visual perception.

PROJECT Diagram of Eye

1. There are two types of visual perceptors in our eyes - rods and cones.

   a) Rods are more sensitive than cones, but have less acuity and do not perceive color.

   b) Cones give us detail vision and color perception, but are less sensitive.

      The cones are concentrated in the central part of our retinas (fovea)

SHOW ON DIAGRAM

[Hence if you stare at a star at night it may seem to disappear, because the cones in the center of your field of vision are less sensitive to light].

2. Moreover, where the optic nerve enters our retina (slightly off center), there is a region in which there are neither rods nor cones -

   SHOW ON DIAGRAM

3. Physically, then, our field of vision should consist of a sharply focused full-color region in the center surrounded by a more blurry region lacking in color, with a large "blind spot". However, what we actually perceive is a uniform field of vision. That is, the brain somehow "fills in" the physical picture.

4. The way in which the brain fills in the raw data leads to some situations in which we see things that are not really there - optical illusions.

   Example: PROJECT Sun and Triangle Illusions

E. A Symbolic Approach to Vision has proved to be a difficult challenge for AI

1. Perhaps this is because there does not seem to be a good "match" between the essentially parallel nature of vision and digital computers, which are inherently sequential machines.

2. Perhaps, too, the difference between mere syntactic manipulation and "understanding" (as, e.g, argued by Searle) is related to this as well.

3. Minsky story

4. Many visual tasks - such as face recognition - have been more successfully handled by approaches such as deep learning.

5. Nonetheless, we will spend some time on the <u>symbolic</u> approach today.

II. Components of Symbolic Computer Vision

A. Cawsey divides visual processing into three levels

PROJECT Cawsey Figure 6.1

1. Low-level processing starts with the raw data and does various pre-processing operations on it, often resulting in a line drawing (often called the primal sketch)

2. Medium level processing seeks to discover the regions that are present in the image, and determine characteristics such as their distance and orientation.

   (We will look at an algorithm shortly - the Waltz algorithm - which is a form of medium-level processing and represents and interesting application of the use of constraint propagation in search. Indeed, it was the place where this approach was first utilized.)

3. High-level processing seeks to determine what is actually being seen -perhaps by describing it in terms of some sort of symbol structure.

B. The raw data for vision is typically represented as a 2-dimensional array of intensity values.

1. Pixel concept

   a) Resolution - e.g. standard TV is 704 x 480; 1080p HDTV is 1920 x 1080.

   b) For gray scale images, each pixel assumes a brightness value in the range of 0 (totally black) to 255 (totally white). For color, three values are used corresponding to red, green, and blue intensities (which correspond to the colors the three kinds of cones in our eyes are most sensitive to.)

2. This seems to correspond to the way our own vision system works, since our retinas are essentially a 2-dimensional array of rod and cone cells, though not at all uniformly laid out as we have already noted.

3. Various mathematical operations can be performed on this array. These are generally local operations performed on a group of neighboring pixels.

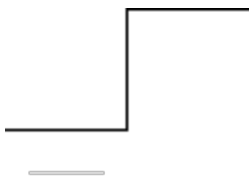
   a) Smoothing - to eliminate the effects of noise and unnecessary detail (e.g. texture)

   b) Edge detection.

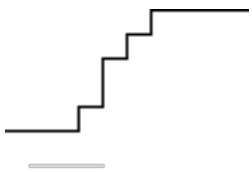
4. The following is a simplified explanation of how this might work out.

   a) If we consider one dimensions of the intensity function, an edge corresponds to a place where the intensity is changing rapidly.

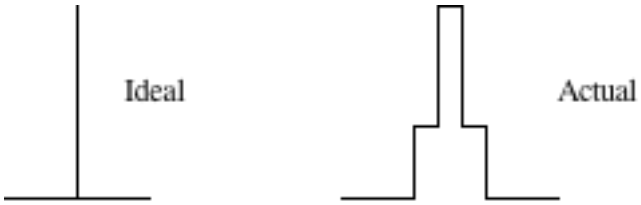
      (1)Ideally, this might look like:



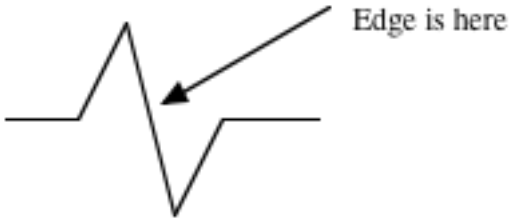
      (2)In practice, due to smoothing, it might end up more like this:



b) If we were to take the derivative (rate of change) to this, we would get:



c) If we took the derivative again (second derivative), we would get an S-shaped curve that crosses through 0 at the edge:



5. The effect of smoothing and taking the second derivative can be combined into a single filter that is called a "sombrero filter" because of its resemblance to a Mexican sombrero hat.

   PROJECT Winston (2e) figure 10-11

   When a sombrero filter is applied to an image represented as an array of intensity values, the zeroes (or the place where the filtered value changes from positive to negative) correspond to the edges.

   (This is the opposite of the example we looked at earlier ("parrots" image), where the most intense points (white lines) corresponded to the edges. This example used a simpler filtering technique.)

6. In living creatures, low-level processing is typically done in the retina, before the data is sent to the brain. Many low-level operations can be done by local filtering operations.

   It turns out that living creatures are particularly sensitive to edges, and there is evidence that, in fact, the eyes of living creatures use a filtering technique that is similar to the sombrero filter.

   PROJECT Winston (2e) figures 10.12, 10.13

7. As you can see, work in computer vision at this level typically involves mathematical techniques that have a distinctly different "feel" from symbolic AI. (This is one place where calculus rather than discrete math is the natural mathematics)

C. A key step in medium-level processing is segmentation - breaking the image into regions that correspond to various objects/backgrounds in the image, and then learning things about those regions.

1. This can use edges found be edge detection to delineate regions, or it can be done more directly from the raw data.

2. Segmentation can also make use of visual intensity, color, or texture

3. At this level, living creatures obtain further information about the image by using stereo vision - the fact that almost all living creatures have two eyes.

   a) The nearer something is to us, the greater will be the difference between its positions in a pair of images from two eyes (or cameras)

   b) Of course, the challenge in stereo vision is determining which parts of the two images correspond (as was discussed in Cawsey)

4. Living creatures also obtain information from motion

   a) This provides another way of learning about distance, since objects that are farther away change position more slowly with movement than do objects near us.

b) Living creatures are often especially sensitive to motion in terms of drawing attention to an object (as anyone who has owned a cat knows!)

D. The final step is converting the image into a high level description in terms of objects occurring in a scene and their relationship to one another.

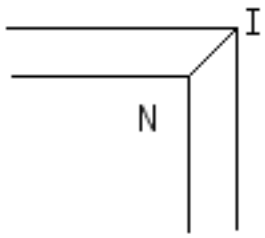
1. Cawsey discusses some approaches to this. However, there is no definitive way of doing this.

2. Perhaps this is not surprising given that often visual data is something that we "can't put into words" either.

III. Constraint Propagation in Vision

A. We are going to look at an AI technique that seems similar to what actually occurs as a part of human vision.

B. Consider the following line drawing: (PROJECT)



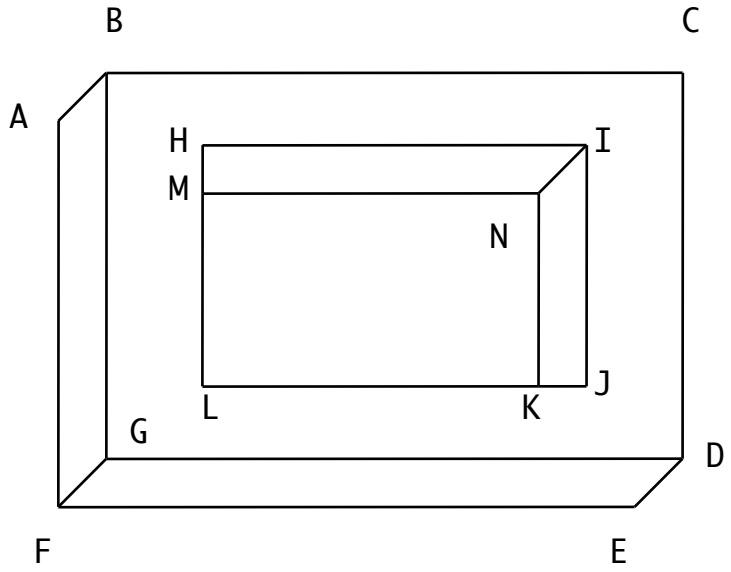
1. Is point “N” closer or farther than point “I”?

   ASK

   Be sure to get both possible answers

2. Now consider a more complete drawing of which this was a part.

   PROJECT



a) Now, the correct answer should be clear.

b) However, if you focus your attention on the portion we looked at earlier alone, does N seem to leap out of the page at you?

ASK

What's going on?

ASK

3. It appears that our visual system is particularly sensitive to the presence of edges.

   Examples?

   ASK

4. One thing that is often done in computer vision is to convert raw data to a line drawing. There is actually a fairly straight-forward way to do this computationally, which seems (based on experiments done on frogs) to be similar to what animal eyes actually do.

   DEMO: Image Editor project with parrots drawing - convert to edges

5. To interpret a line drawing, it is necessary to interpret the individual lines

C. One of the earliest uses of constraint propagation in AI was in conjunction with interpretation of line drawings. We will consider a simplified version of this procedure here. (Even the full procedure as originally developed by David Waltz has significant limitations, which subsequent work has addressed).

1. Lines occur in images for a number of reasons

   a) Actual physical edges

      (1)A boundary between an object and the background

      (2)A convex edge in the interior of the object

      (3)A concave edge in the interior of the object

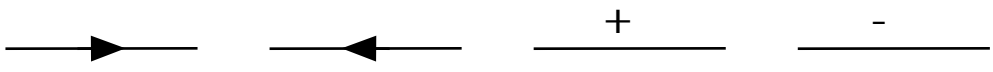
   b) Pseudo-edges

      (1)Crack

      (2)Shadows

   c) Markings on the surface of the object - which we may choose to treat as real edges

2. To interpret a line drawing, we need to interpret the lines. For simplicity, the example we will develop here will consider only lines corresponding to actual physical edges. Our ultimate goal is to interpret each line as corresponding to one of the four kinds of physical edge.

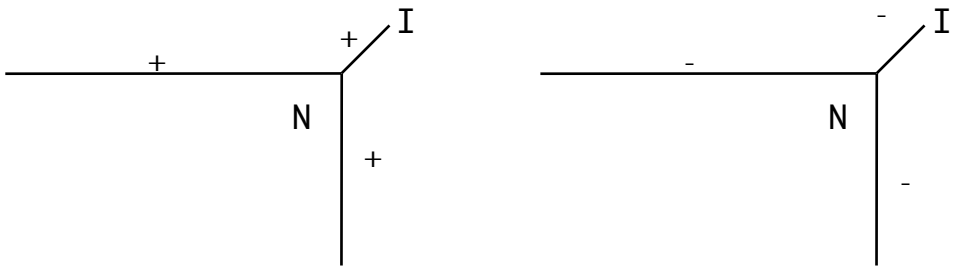


Boundary edge - object is to right when following arrow

Convex edge

Concave edge

Example: The two ways of "seeing" the relationship between points N and I correspond to two ways of interpreting the edges



The "correct" interpretation (the one on the right) is actually determined by constraint propagated from the surrounding context.

3. Waltz's procedure turns the problem of labeling lines into a problem of labeling junctions. It relies on two physical constraints to make the problem tractable:

   a) Though many combinations of line labelings at a junction are combinatorially possible, only certain combinations are physically possible.

      PROJECT: Nilsson Figure 6.16. This gives a set of of junction labelings in a restricted environment where we impose the following limitations:

      (1)No shadows or cracks are allowed; only real edges. (Thus, only the four line labels we mentioned earlier are needed.)

      (2)All junctions are formed from at most three faces. (The pyramids of Egypt are ruled out.)

      (These classes of junctions are called, respectively, V's, W's, Y's, and T's)

      (3)The viewing angle is not singular; what we see would not be drastically altered by a slight movement of position.

(4)Note that real vision systems must deal with a much more complex set of conditions, including shadows, cracks, and junctions formed from more than three faces. The set of possible labelings of junctions would be very hard to enumerate by hand, but has been done by computation.

b) Each line connects two junctions. The labelings at the two junctions must both assign the same label to the line.

4. The problem of finding a set of consistent line labelings could be viewed as a search problem, with a particular assignment of labels being a state. In this case, the goal would be to find a physically consistent set. However, the size of the search space would be very large - e.g. with just the 4 labels and 10 lines, we would have 4^10 (> 1 million) states to consider; and with 20 lines, we would have over a trillion. Waltz's procedure dramatically reduces the size of the search space by using constraint propagation, as we did in the Sudoku example. (In fact, when I first started solving Sudoku's I recognized that a way to approach a solution was to use this technique that I already knew about!)

a) At any given time in the procedure, each junction has associated with it a set of possible labelings. (Ultimately, the set associated with each junction will become a singleton.)

b) Initially, we label the outside edges as boundaries. Then we associate with a junction the set of all possible labelings for junctions of that type (L, fork, arrow etc.) which are consistent with these labeling. (In some cases, this will mean that all possibilities for a junction are available.)

c) Constraints propagated from neighbors allow us to eliminate certain elements from the set of possible labels for the junction, until we are (hopefully) eventually left with just one. Constraint is propagated as follows: whenever we change the set of labelings on a junction (e,g, by making it smaller), we examine each of its neighbors.

(1)Let J be a junction we have just learned something new about

(2)Let E be an edge connecting it to a neighbor (N).

(3)The set of labels for J determines a set of possible labelings for E. For example, if none of the possible labels for J has E labeled as a "+", then "+" is not one of the possible interpretations for E.

(4)We eliminate from N's set of labelings any label which requires an interpretation for E not allowed by J's labeling.

(5)Eventually, the set of possible labelings for J or N determines a unique label for E. At this point, we mark E.

d) Whenever the set of labelings for a junction's edges eliminates all possibilities for the junction save one, we can label the junction and the edges connected to it.

D.Example - labeling the figure we used earlier.

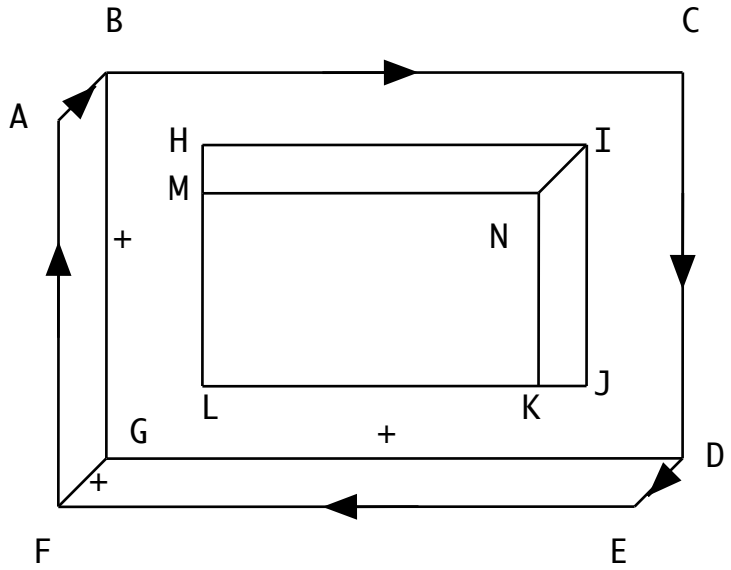
PROJECT AGAIN (Use pdf version - choose View | Single Page & View | Full Screen but <u>not</u> slideshow)

E.First, label all outside boundaries with an arrow going clockwise starting with A.

1. The W's at B, D, and F must be labeled with +'s on their third barb.

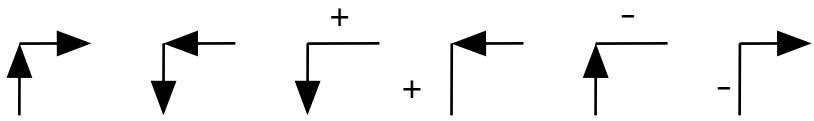
2. The Y at G therefore becomes all +'s, which is one of the possibilities for a Y.

   PROJECT

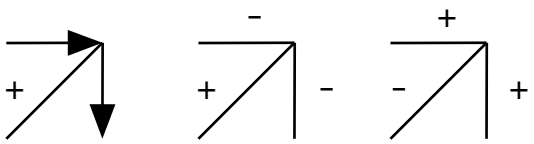


3. This finishes the outer edges. We now must plunge into the interior.

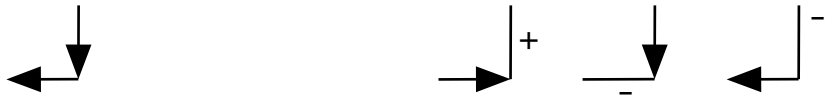
   a) If we start with H, all 6 "V" labelings are initially possible.



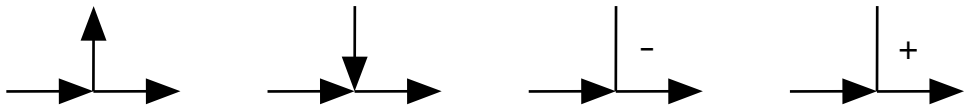
   b) Likewise, for I all 3 "W" labelings are initially possible, since each is consistent with some labeling for H.



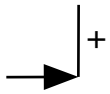
   c) At J, we can eliminate 2 of the 6 "V" labelings as inconsistent with the set of labelings for I. (There is no labeling for I that has a boundary edge going from J to I.) That leaves the following possibilities:



d) At K all 4 “T's” are possible

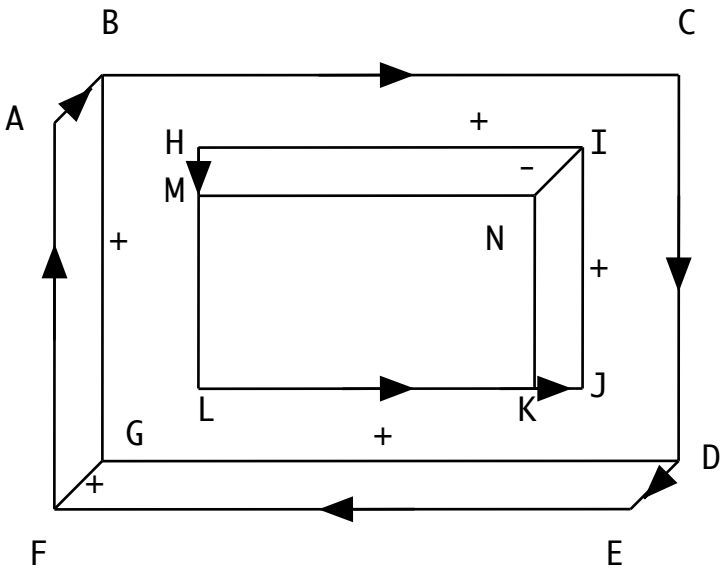


e) But now, we can propagate some constraint back to J. The two “V’s” having a boundary from J to K are out, as is the “V” having the J-K edge a -, since K constrains the J K edge to be a K to J boundary. Thus, the labeling for J has been fixed as:

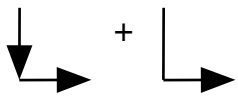


f) Further, the fixing of J's labeling forces the labeling for I, which in turn forces the labeling for H, leading to:

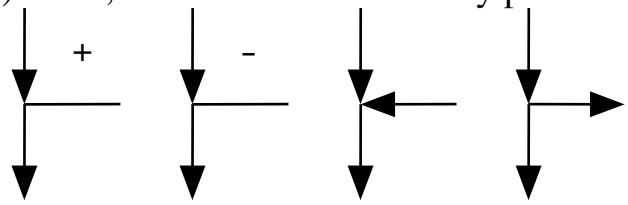
PROJECT



g) At L, only two labelings are possible consistent with K:

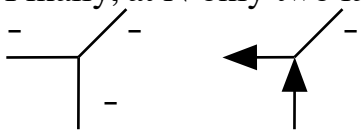


h) At M, all four "T"'s are initially possible:



i) However, since all of M's labelings have ML a boundary edge, this eliminates one possibility for L, forcing the remaining one as the only possibility.

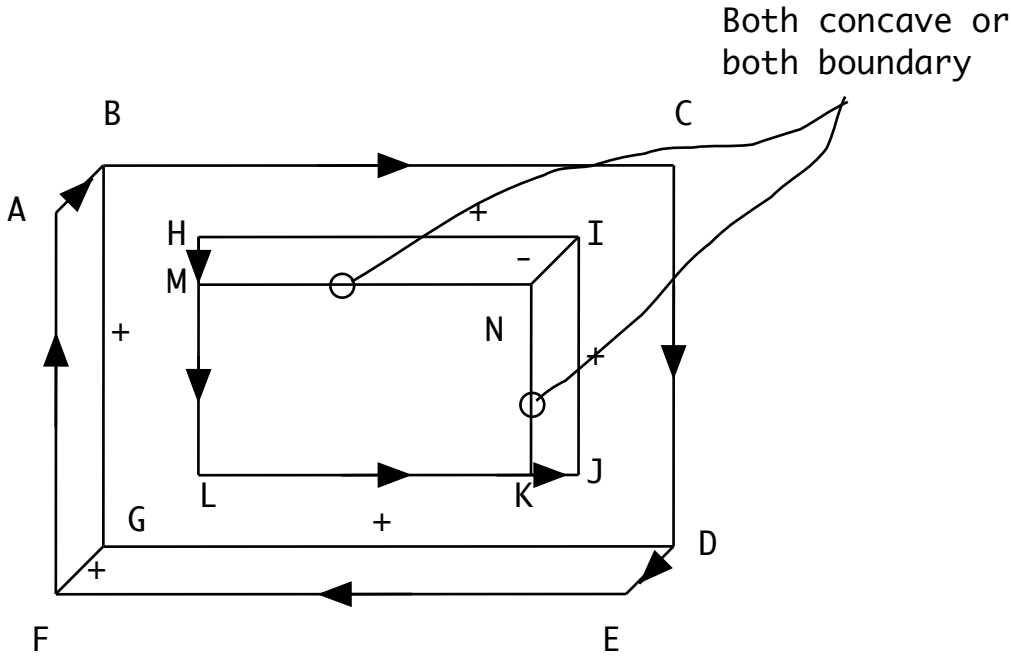
j) Finally, at N only two labelings are consistent with I:



k) This, in turn, reduces possible labelings of K and M to two each.

4. At this point, our labeling looks like this

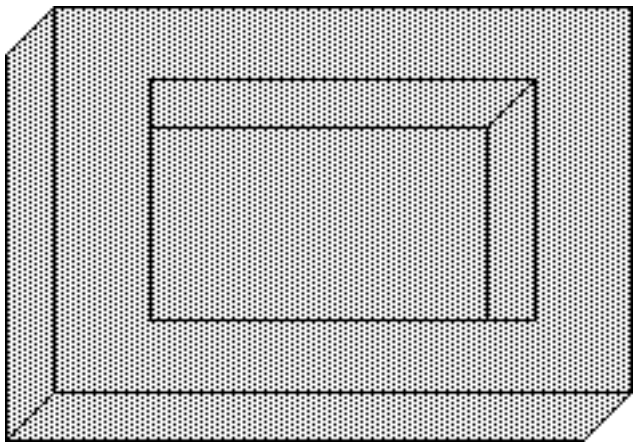
PROJECT

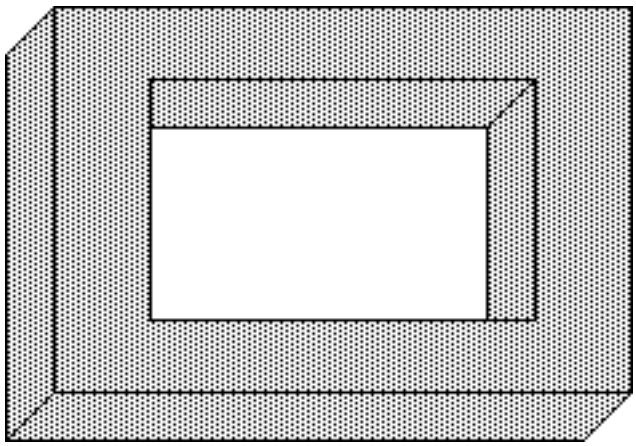


We can go no further. The Waltz procedure has uncovered a genuine a genuine ambiguity - is the center area (bounded by K,L,M,N) a solid bottom or a hole? Further data would be needed to settle this. (E.g. if this area were a different color or shading from the background, it's probably a bottom; if the same as the background, probably a hole.

To see this, consider these two different versions of our drawing with shading:

PROJECT (Use pdf version - choose View | Single Page & View | Full Screen but <u>not</u> slideshow)





F. Exercise to do in class:

Waltz's procedure problem

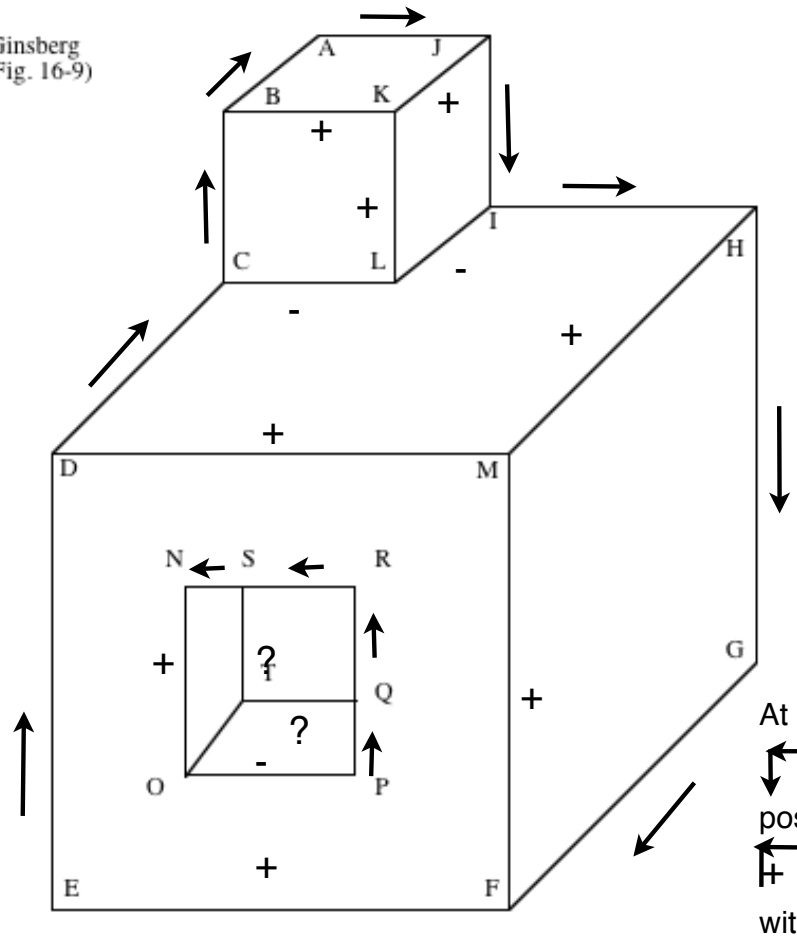
HANDOUT with possible labelings and figure to label.

Hint: Outside labeling is straightforward.

For inside, note that T's at Q and S force N-S-R-Q-P to be boundary edges. While two V's are possible at N and Q consistent with the boundaries, only the one with the + on the other barb is consistent with O (the W version for O requires the arrows going the other way.) This forces O to be three +'s, and fixes N and Q. Now all the rest are fixed.

PROJECT Filled in version

(From Ginsberg (1993) Fig. 16-9)



At N and P, both ← and ← are possible; but only ← is compatible with any possibility for O

The two innermost edges can be consistently labeled either - or ↓→ corresponding to either a solid bottom or a hole.