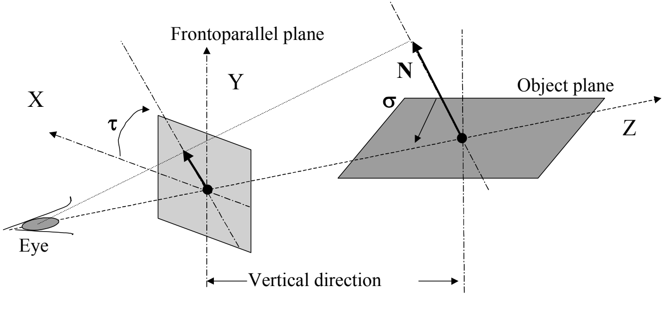
The Problem of Depth Perception

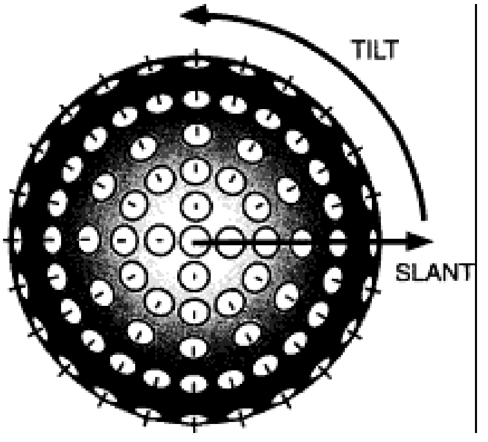
- There are two closely related problems that must be solved in perceiving the spatial arrangement of surfaces with respect to the observer.

  1. Determining depth
  2. Perceiving surface orientation: the slant and tilt of the surface with respect to the viewer's line of sight.

The Problem of Depth Perception

- Slant refers to the size of the angle between the observer's optical axis and the surface normal.
- Tilt refers to the direction of the depth gradient as projected on the frontal plane. See example.





Slant and tilt of local surface patches on a sphere.

Slant refers to the angle between the optical axis and its surface normal (the direction perpendicular to the surface), as indicated by the degree of elongation of the ellipses in this figure.
Tilt refers to the direction of the surface normal projected onto the frontal plane.

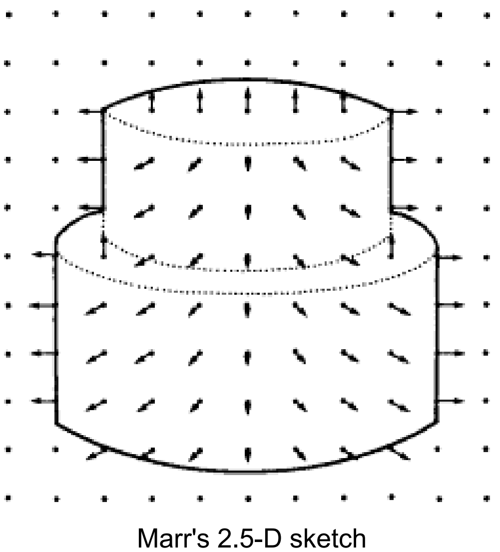
Back

The Problem of Depth Perception

- Why such encoding?
  - Partly invariant over image rotations, ie rotating about Z-axis: slant remains invariant.
  - More importantly, separate tilt from slant. Tilt can be determined accurately (from texture, motion etc) but not slant. Advantageous to separate these two.

Marr's 2.5-D Sketch

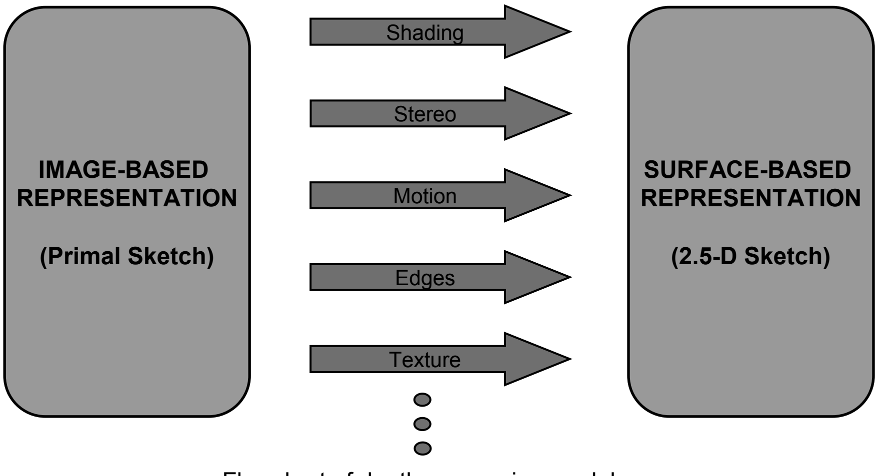
- The most influential proposal to date has been David Marr's conception of the 2.5-D sketch.
- The 2.5-D sketch is somewhere between the 2-D properties of an image-based representation and the 3-D properties of a true object-based representation.



Marr's 2.5-D sketch
At each position in the visual field, a vector is shown (called the surface normal) that is perpendicular to the local patch of surface at that point. The 2.5-D sketch also includes information about the distance.

Marr's 2.5-D Sketch

- There are many independent (or nearly independent) processing modules computing depth information from separate sources
- Marr’s 2.5-D sketch summarizes the many converging outputs of different processes that recover information about the depth and orientation of local surface patches



Flowchart of depth-processing modules.
A 2.5-D surface-based representation is thought to be computed from a 2-D image-based representation by a set of parallel and quasi-independent processes that extract information about surface orientation and depth from a variety of sources, such as shading, stereo, and motion.

Marr's 2.5-D Sketch

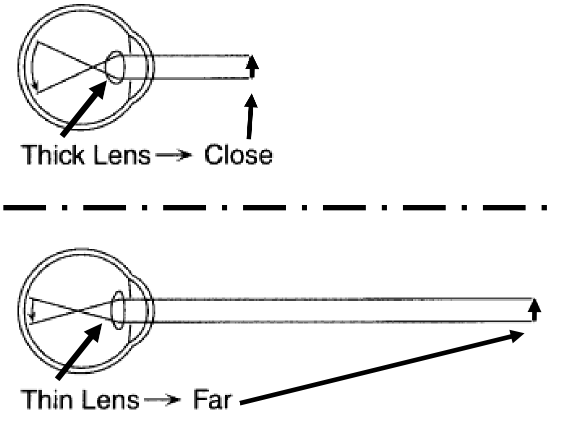
- In the computer vision community, these processes are frequently referred to as the "shape-from-X modules," where X is the source of depth information, as in "shape-from-shading" or "shape-from-motion."
- It is also called "depth-from-X" or "orientation-from-X" modules, or more vaguely, “structure-from-X”.

Ocular Information

- Ocular information about the distance to a fixated surface, arises from factors that depend on the state of the eyes themselves and their various components.
- The focus of the lens (accommodation) and the angle of between the two eyes' lines of sight (convergence) are particularly important for depth perception.

Accommodation

- Accommodation is the process through which the muscles in the eye control the optical focus of the lens by temporarily changing its shape. See example.
- The lens of the human eye becomes thin to focus light from faraway objects on the retina and becomes thick to focus light from nearby ones.
- If the visual system has information about the tension of the muscles that control the lens's shape, then it has information about the distance to the focused object.



Depth information from lens accommodation.
The lens of a human eye changes shape to focus the light from objects at different distances: thin for objects far away and thick for ones nearby.

Back

Accommodation

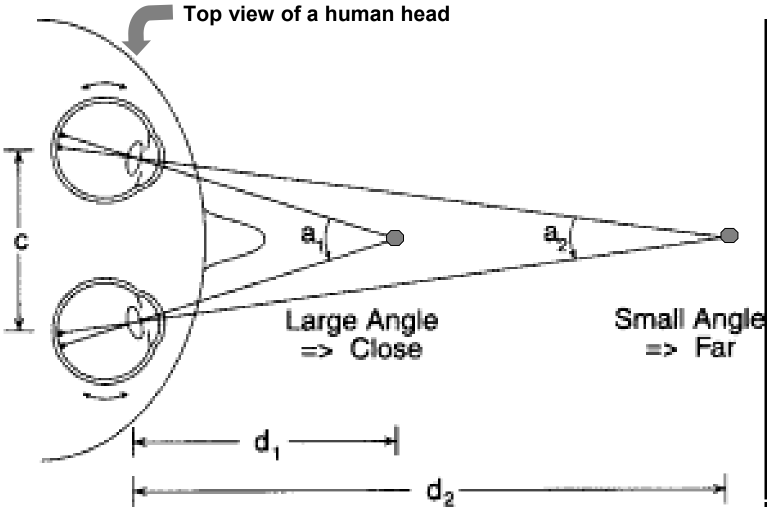
- Experimental results indicate that people use accommodation at close distances.
- Beyond 6-8 feet, however, accommodation provides little or no depth information.
- At this distance the muscles that control the shape of the lens are already in their most relaxed state, so the lens cannot get any thinner.

Accommodation

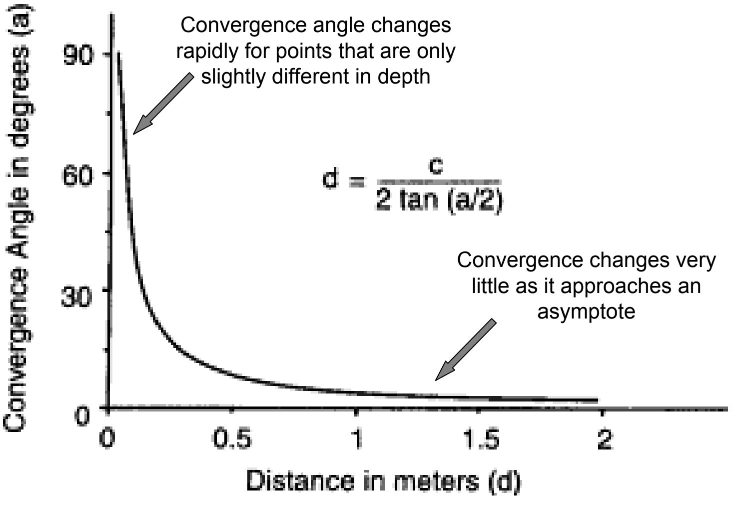
- Accommodation provides information about absolute depth.
- It can specify the actual distance to the fixated object, provided the visual system is properly calibrated.
- Most optical depth cues merely provide information about relative distance, indicating which of two things is closer or the ratio of the distances of two objects.

Convergence

- Another source of information about depth comes from eye convergence: the extent to which the two eyes are turned inward (toward each other) to fixate an object.
- The eyes fixate a given point in external space when both of them are aimed directly at the point so that light coming from it falls on the centers of both foveae simultaneously.



Depth information from eye convergence.
The angle of convergence between the two eyes (a) varies with the distance to the object they fixate: smaller angles for objects far away (a_2) and larger angles for objects nearby (a_1).



Convergence as a function of distance.
The angle of convergence changes rapidly with distances up to a meter or two but very little after that.

Convergence

- The important fact is that the angle formed by the two lines of sight varies systematically with the distance between the observer and the fixated point.
- Fixate a close object ⟶ large convergence angle.
- Fixate a far object ⟶ small convergence angle.
- Only useful up to a few meters. Beyond that, convergence angle changes very little.

Convergence

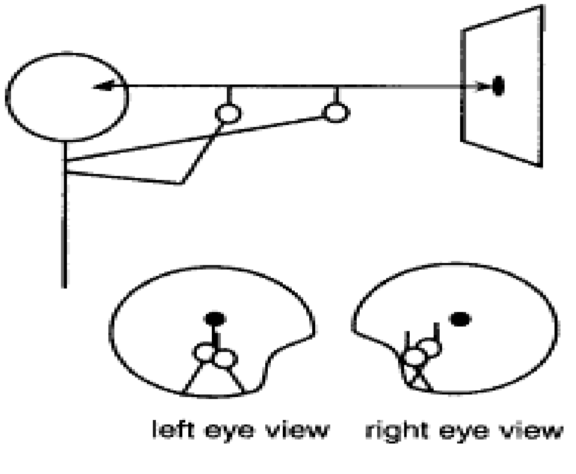
- Because convergence depends on the observer using both eyes, it is a binocular source of depth information, unlike accommodation.
- Convergence provides information about the absolute distance to the fixated object, like accommodation.

Stereoscopic Information

- The most compelling experience of depth comes from **stereopsis**: the process of perceiving the relative distance to objects based on their lateral displacement in the two retinal images.
- Stereopsis is possible because we have two laterally separated eyes whose visual fields overlap in the central region of vision.
- The positions of the eyes differ by a few inches, the two retinal images of most objects in the overlapping portion are slightly different.

Stereoscopic Information

- The same point in the environment projects to locations on the left and right retinae that are displaced in a way that depends on how much closer or farther the point is from the fixation point.
- This relative lateral displacement is called **binocular disparity**. See example.
- The direction of disparity provides information about which points are closer and which are farther than the fixated point.
- The magnitude of this disparity provides information about how much closer or farther they are.



Demonstration of binocular disparity.
The fact that the two eyes register somewhat different views of the world can be demonstrated by performing this finger experiments.

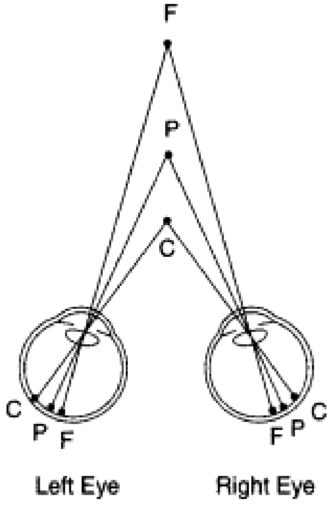
Back

Instructions

- Hold up your left index finger at full arm's length and your right index finger at half arm's length.
- Close your right eye, and align your two fingers using your left eye so that they both coincide with some distant point in the environment.
- Keeping your fingers in the same place and continuing to focus on the distant object, close your left eye and open your right.
- What happens to the images of your two fingers?
- They are no longer aligned either with the distant fixated point or with each other but are displaced markedly to the left.
- These differences occur because the few inches of separation between your eyes provide two slightly different views of the world.

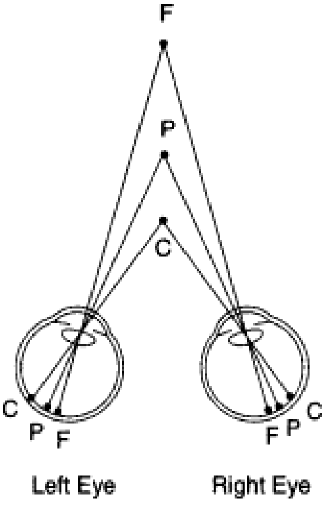
Back

Binocular Disparity



- Consider the diagram of two eyes fixated on a point, P.
- Point P falls on the foveae of both eyes and thus stimulates corresponding points.
- Now consider point C while the eyes are still fixated on point P.
- They do not fall on corresponding retinal points.

Binocular Disparity

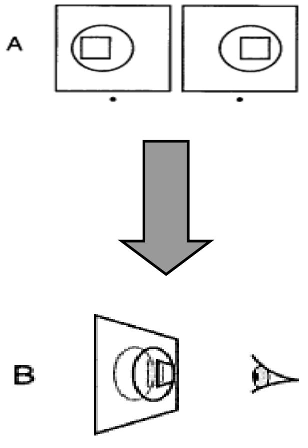


- This outward direction is called crossed disparity (remember "c" is for "close").
- This inward (or nasal) direction is called uncrossed disparity.
- It indicates that the point that gave rise to it is farther away than the fixated point.

Stereograms

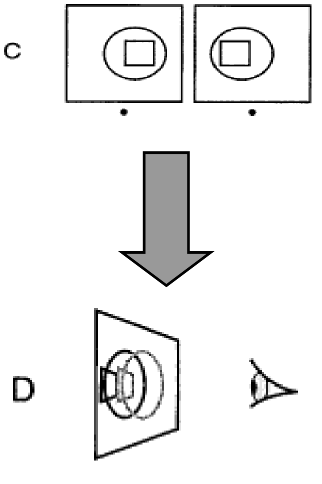
- The most powerful demonstration that binocular disparity can produce depth illusions.
- Pairs of images that differ in the relative lateral displacement of elements.
- When viewed stereoscopically, they produce compelling illusions of depth from a completely flat page
  - Invented by Charles Wheatstone in 1838.
- can also be perceived without any such apparatus.

Stereograms



- **A** shows one such stereo pair. Notice that, relative to the square borders, the circles are displaced somewhat toward the outside, and the squares even more so.
- When these two images are stereoscopically fused by crossing the eyes, this lateral disparity produces a percept in depth like the display shown in **B**.

Stereograms

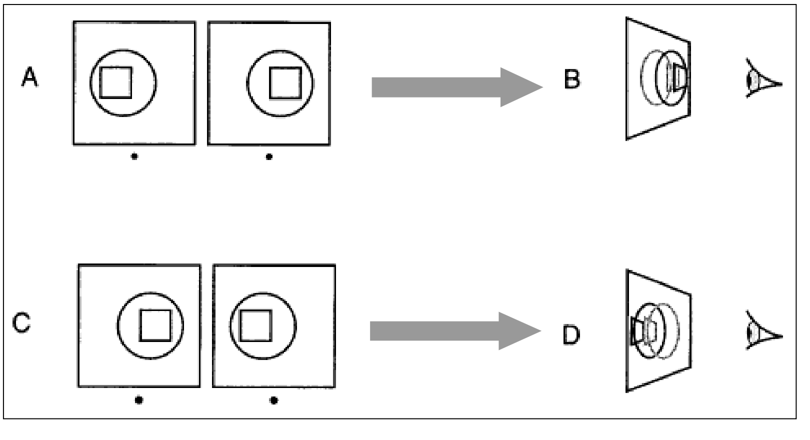


- **C** shows a pair with the reverse disparity, which produces the same percept, except reversed in depth, as depicted in **D**.
- This result is precisely what we would predict given our previous discussion of crossed versus uncrossed disparity.

Stereograms

- To experience these stereograms in depth, you must get your two eyes to register different images that your brain can fuse into a single image.
- There are two ways to do this:
  1. Crossed convergence method
  2. Uncrossed convergence method

Stereograms



Crossed convergence method

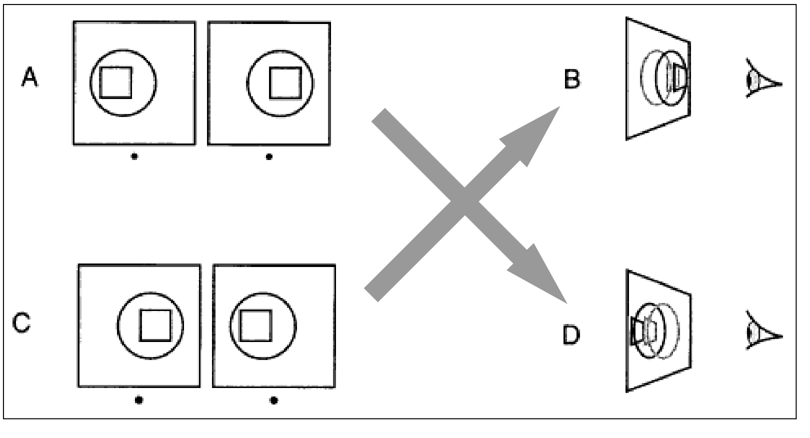
Back

Stereoscopic fusion of a stereo pair by the crossed convergence method

1. Cross your eyes by just the right amount to bring the left image in your right eye into registration with the right image in your left eye.
2. You will generally experience four images once you misalign them by crossing your eyes.
3. The goal is to adjust this misalignment so that there are exactly three images,
4. The middle one of which is the alignment of two images that together produce the perception of depth.

Back

Stereograms



Uncrossed convergence method

Back

Stereoscopic fusion of a stereo pair by the uncrossed convergence method

1. Bring the book very close to your face, with your nose touching the page.
2. Relax your eyes and slowly move the book away from your face.
3. Again you will see four images and four dots most of the time, but at some distance from the page, you will see exactly three images and three dots.
4. Stop at this point and attend to the middle image, which will eventually appear in depth

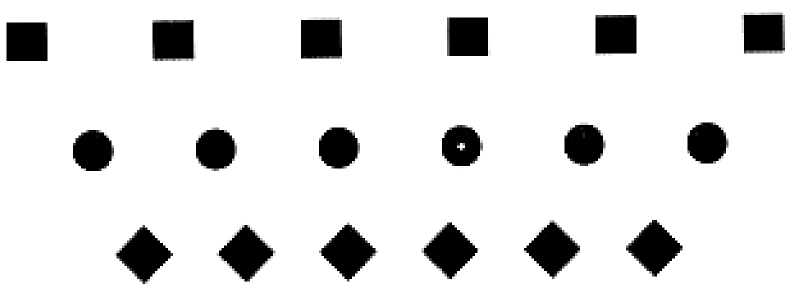
Back

Stereograms

- In both methods, your eyes are misconverged for the distance to the page.
- In the crossed method, your eyes are overconverged,
- in the uncrossed method, they are underconverged.
- These deviations from normal viewing convergence are what allow such stereograms to produce the illusion of depth.

Autostereograms

- Another kind of stereogram has become popular in the past few years.
- They were initially called autostereograms, but they are now more widely known as magiceye stereograms.
- See example.



A simplified autostereogram.
The rows of figures should begin to appear in different depth planes. The squares will appear closer than the circles, and the diamonds will appear farther away than the circles.

Back

1. View it using the crossed convergence method.

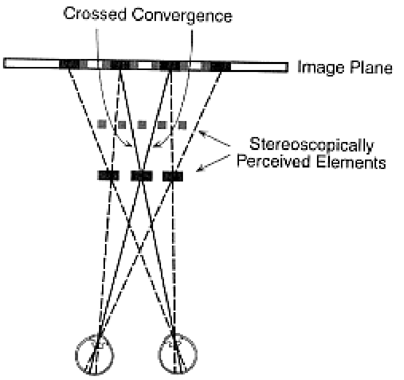
2. When you manage to see two adjacent circles with white dots inside them, the different rows of shapes will appear to be located at different distances away from you.

3. squares closest, diamonds farthest and circles in the middle.

4. If you view this auto-stereogram with uncrossed convergence, the depth of the shapes will reverse.

Back

Autostereograms

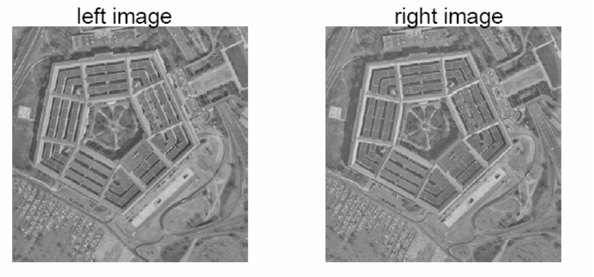


Illusions of depth occur in viewing autostereograms when the eyes are misconverged such that different elements in the repetitive pattern are fused as same object.

The objects in the same row are identical in shape to enable this fusion error to occur.

The Correspondence Problem

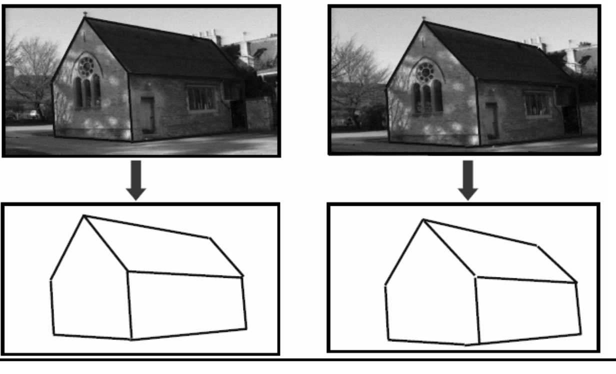
- So far, the only problem stereoscopic vision poses for the visual system is to measure the direction and amount of disparity between corresponding image features in the two retinal images.
- But we have assumed that the visual system has already discovered which features in the left image correspond to which features in the right one.
- This is called the correspondence problem, a very difficult inverse problem.



Stereo pair of Pentagon

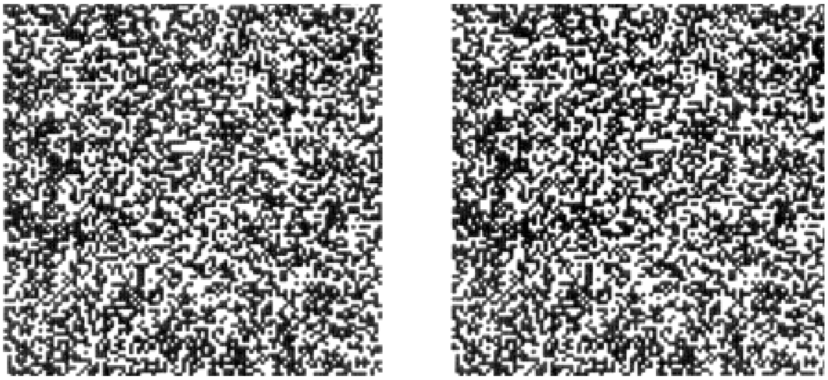
The Correspondence Problem

- For many years, theorists assumed that this problem was solved by some sort of shape analysis that occurred before stereopsis.
  - Shape was assumed to be analyzed first, separately for the left and right images, so that the results could be used to solve the correspondence problem.



Random Dot Stereograms

- The alternative possibility is that stereopsis might actually come first and occur without the benefit of monocular shape information.
- Bela Julesz realized that he could test the shape-first theory of the correspondence problem by constructing what he called random dot stereograms.
- It is a pair of images consisting of thousands of randomly placed dots whose lateral displacements produce convincing perception of depth when viewed stereoscopically.



A random dot stereogram.
These two images are derived from a single array of randomly placed squares by laterally. displacing a region of them as described in the text. When they are viewed with crossed disparity (by crossing the eyes) so that the right eye's view of the left image is combined with the left eye's view of the right image, a square will be perceived to float above the page.

Random Dot Stereograms

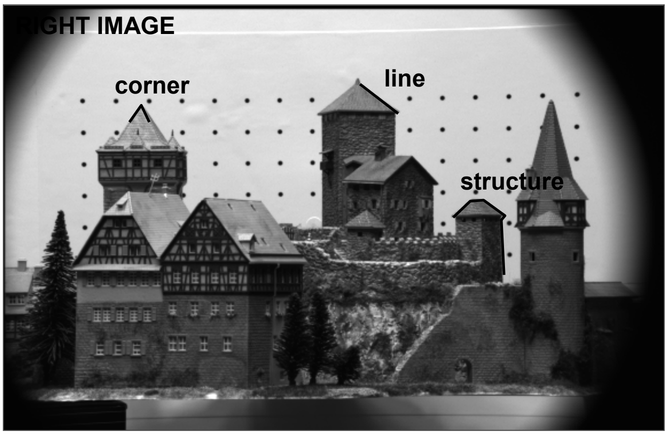
- The dots look random in the sense that no global shape information is present in either image alone when each image of a random-dot stereogram is viewed by itself.
- It should be impossible to perceive depth by fusing random-dot images stereoscopically because the shape-first theory assumes that the correspondence must be based on recognizable monocular shape information.
- Thus, the appropriate conclusion is that the shape-first theory is *incorrect*.

Random Dot Stereograms

- It is important not to overstate the conclusion reached from the perception of random dot stereograms.
- The fact that people can perceive stereoscopic depth in random dot stereograms does not prove that there is no shape analysis prior to stereopsis.
- It shows only that stereoscopic depth can be perceived without monocular shape information.
- The difficulty of achieving depth perception in random dot stereograms compared to stereo pairs of normal photographs suggests that monocular shape information is indeed useful.

Random Dot Stereograms

- Orientation of local lines and edges is one example of further information that would be useful in solving the correspondence problem. Only lines or edges of similar orientation would be potential matches.



Computational Theories

- How does the visual system solve the correspondence problem ? The visual system must somehow figure out which pairs of points go together.
- We will now consider an approach to solve the correspondence problem for the simple case of horizontal binocular displacement.
- If there are 100 points along a given horizontal line in one image, then the number of logically possible pairings is 100! (one hundred factorial).

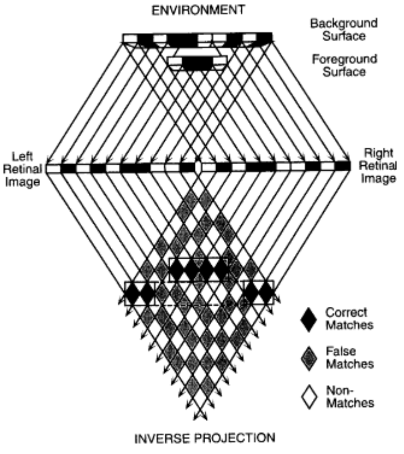
The First Marr-Poggio Algorithm to Solve Correspondence

- A well-known algorithm devised by David Marr and Tomaso Poggio of M.I.T. in 1977.
- It is historically important:

  example of how a dynamic neural network can be constructed to solve a computationally difficult visual task.
- It is also a good example of how heuristic assumptions or constraints can be implemented in such networks.

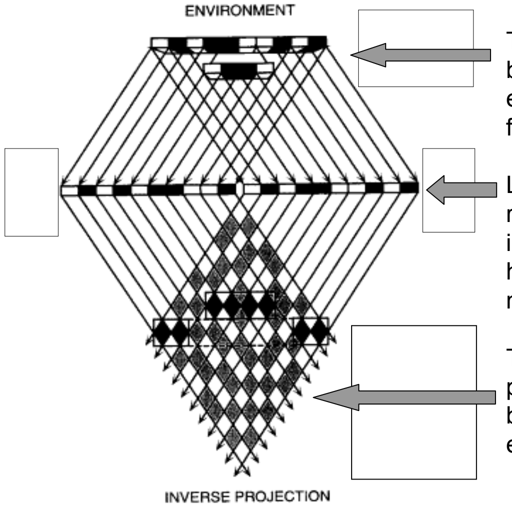
The First Marr-Poggio Algorithm to Solve Correspondence

- The basic idea: inverse projection (or triangulation) from the two retinal images to get 3-D depth.
- See example.
- The fact that there are false matches as well as true ones reflects the fact that this is an underconstrained inverse problem.



Inverse projection of two surfaces in depth.
The upper half of the figure shows a top view of how two surfaces project to (flattened) left and right retinal images. The lower half shows how these images can be used to form an inverse projection in which there are many same-color matches, some of which correspond to the correct depth solution (dark elements) and many of which are false matches (gray elements).

Back

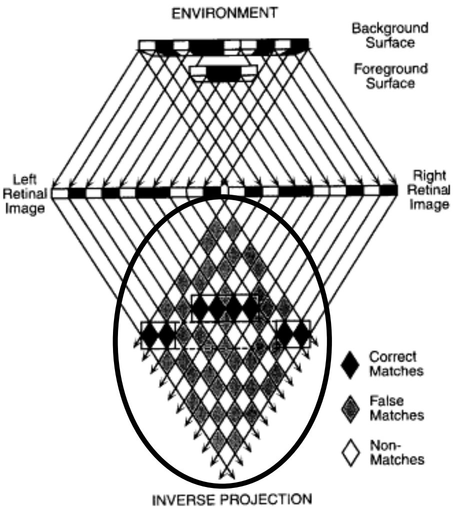


This diagram depicts a top view of two black-and-white striped surfaces in the environment, a small one situated in front of a larger one.

Light reflected from these surfaces registers on the left and right retinal images — which have been flattened here for simplicity — as shown in the middle of the diagram.

To form an inverse projection, each pixel of these images is then projected back into the "mirror-image environment" shown of the diagram.

Back



The shaded cells in this matrix represent positions at which there are color matches between pixels in the two images.

The points match (and are shaded) if both pixels are black or both are white;

They do not match (and are unshaded) if one is black and one is white.

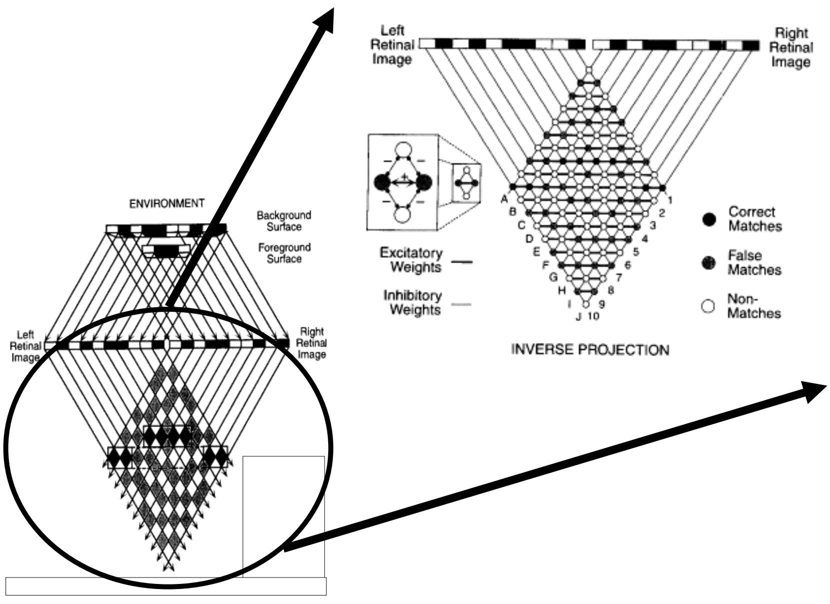
Among these numerous matches are the correct ones that correspond to the visible portions of the actual surfaces in the real world.

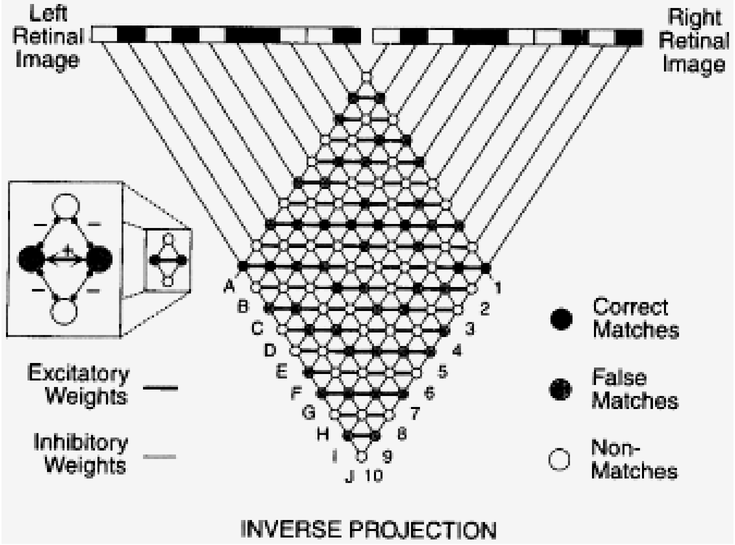
These correct matches are shaded more darkly

Back

The First Marr-Poggio Algorithm to Solve Correspondence

- Marr and Poggio (1977) proposed that the matching could be accomplished by neural network. See diagram.
- The model works by first activating all of the nodes in the "intersection network" that represent like-colored pixels in the left and right images.
- These are just the shaded nodes, indicating that they have been activated in the initial phase of the algorithm.





Only the intersections that come from pixels of the same color are possible matches because pixels that are projected from the same point in the environment must have the same color (both white or both black).

Back

The First Marr-Poggio Algorithm to Solve Correspondence

- Matching by color does not produce a unique solution to the correspondence problem because there are still many color matches for each point in the left and right images.
- What further heuristic constraints might be brought to bear on this formulation of the problem that would produce a unique, correct solution most of the time?

The First Marr-Poggio Algorithm to Solve Correspondence

- Marr and Poggio (1977) employed two such further constraints:

  1. Surface opacity:

  – The opacity constraint states that because most surfaces in the world are opaque, only the nearest one can be seen. Thus, if correspondence A10 is correct in previous figure then correspondences B10, C10, D10, and so forth cannot be correct, and neither can A9, A8, A7, and so forth.

The First Marr-Poggio Algorithm to Solve Correspondence

- Marr and Poggio (1977) employed two such further constraints:

  2. Surface continuity:

  – The continuity constraint states that because surfaces in the world tend to be locally continuous in depth (except at occluding edges), the correct solution will tend to be one in which matches are "close together" in depth, as they would be on locally continuous surfaces.

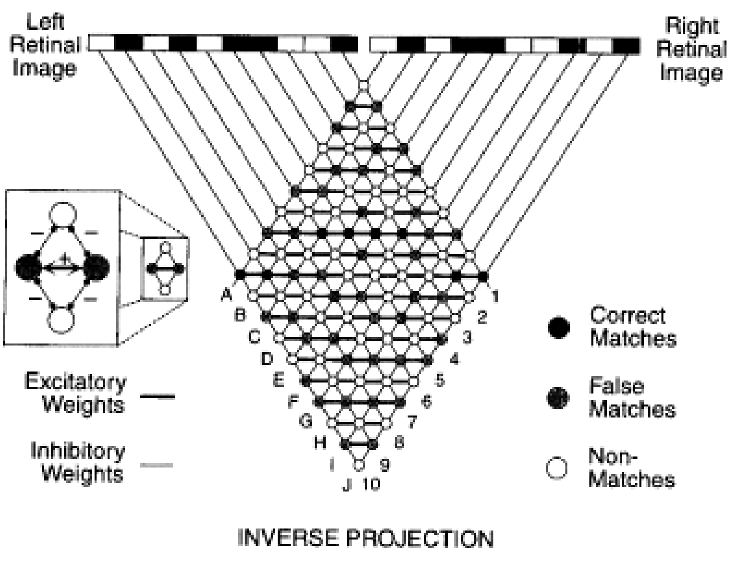
The First Marr-Poggio Algorithm to Solve Correspondence

- Note that both constraints are heuristic assumptions that are usually true but not always.
- If they are true, the solution the algorithm finds will generally be correct.
- If they are not (if one or more of the surfaces are transparent and/or if the surfaces are not locally continuous) the solutions that it finds will tend to be incorrect.

The First Marr-Poggio Algorithm to Solve Correspondence

- Opacity is implemented by having mutual inhibition among all nodes along the same line of sight in the network.
- This part of the architecture is called a **winner-take-all network** because it allows only one node in each diagonal line to remain active after activation has settled into a stable state.
- The continuity constraint is implemented in the network by having mutual excitation between pixels in the same or nearby depth planes.

See diagram.

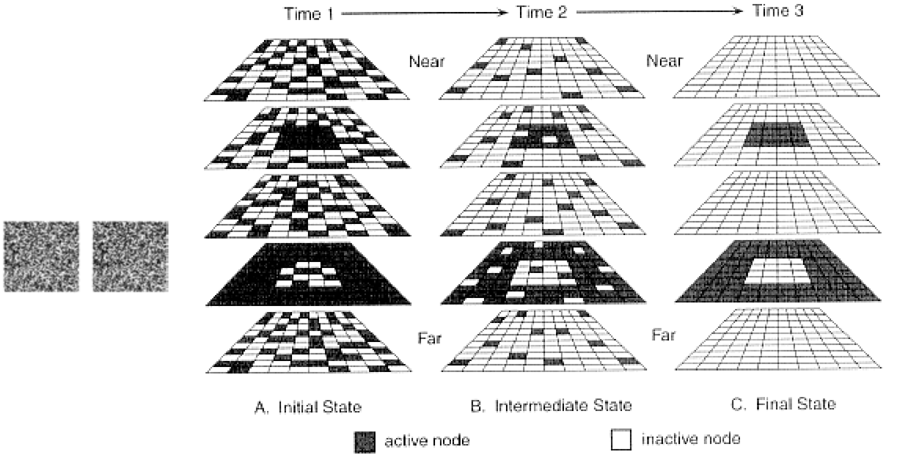


Only the intersections that come from pixels of the same color are possible matches because pixels that are projected from the same point in the environment must have the same color (both white or both black).

Back

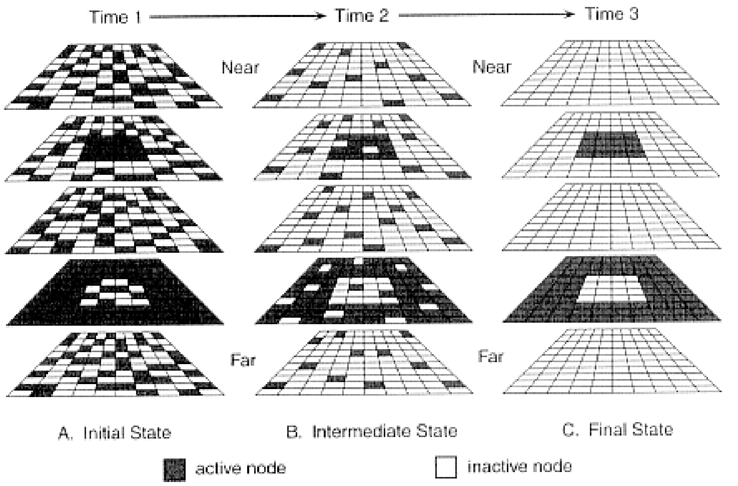
The First Marr-Poggio Algorithm to Solve Correspondence

- Thus possible correspondences in the same depth plane tend to "help each other out" by activating each other through mutual facilitation.
  - cause solutions to tend toward coplanarity
- The network churns away, sending activation and inhibition back and forth through the excitatory and inhibitory connections, until it eventually settles into a stable state. See example.
- This final state of the network usually corresponds to the actual state of affairs in the environment.



The above diagram shows how activation in the intersection network changes over time for the simple case of a square figure in front of a background plane, as in the stereogram on the left.

Back



1. The initial pattern of activation is completely determined by the set of pixel pairs whose colors match.
2. As the nodes begin to interact via the excitatory and inhibitory connections, the opacity and continuity constraints begin to take effect.
3. Eventually, the initial correspondences are reduced to just the set representing a square in front of a background

Back

The First Marr-Poggio Algorithm to Solve Correspondence

- Marr and Poggio's first algorithm is an interesting example of how a process of unconscious inference can be implemented in a neural network without invoking deductions based on either numerical calculations or the sequential application of symbolic logical rules dealing with variables.
- Instead, the assumptions are built into the connection strengths within the network, and they work simply by having a favorable effect on its behavior with respect to achieving veridical perception.

Edge-Based Algorithms

- Marr and Poggio (1979) suggested a second one that differs from their first in a number of important respects:
  1. Edge-based matching.
  2. Multiple scales.

Edge-based matching

- The second Marr-Poggio algorithm finds stereoscopic correspondences by matching edges in the right and left images rather than individual pixels.
- More information is taken into account. Hence more efficient
  - Edges that do not match in orientation and polarity can be eliminated from consideration, thus adding further constraints on the solution.

Multiple scales

- Multiple scale operation like human visual system. First look for corresponding edges at a large spatial scale and only later at smaller scales.
- The early, large-scale process does not work on individual dots but on larger regions of the image.
- Only after possible edge matches are found at this global level is a more detailed matching attempted at the finer-grained level of individual dot contours.

Edge-Based Algorithms

- Edge based algorithm is more plausible biologically because binocular processing begins in area V1 of the cortex,
  - In V1, cells are more sensitive to edges than points
- It also agrees more closely with the results of psychophysical experiments using human subjects (Marr, 1982).
- It is thus a better model of human stereo vision than their first algorithm.

Dynamic Information (Motion)

- This refers to changes in visual structure that occur over time due to object or self motion.
- Retinal displacement depends on:
  1. the observer's motion.
  2. how far away the objects are.

Motion Parallax

- One way in which depth can be recovered from motion information is due to what is known as motion parallax:
  – the differential motion of pairs of nearby image points due to their different depths in 3D.
- See demonstration.

1. Hold your two index fingers in front of your face, one at arm's length and the other halfway to your nose.
2. Close your right eye and align your two fingers with some distant object.
3. Keeping your fingers as still as possible, slowly move your head to the right.
4. Notice that both of your fingers move leftward relative to the distant object, but the closer finger moves farther and faster.
5. Now move your head to the left and notice that your fingers move rightward, the closer one again moving farther and faster.

Back

Motion Parallax

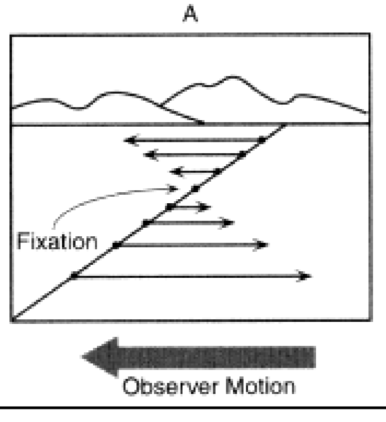
- Motion parallax and binocular disparity are quite similar
- In the case of binocular disparity, the image pair is displaced in space.
- In the case of motion parallax, the image pair is displaced in time.

Motion Parallax

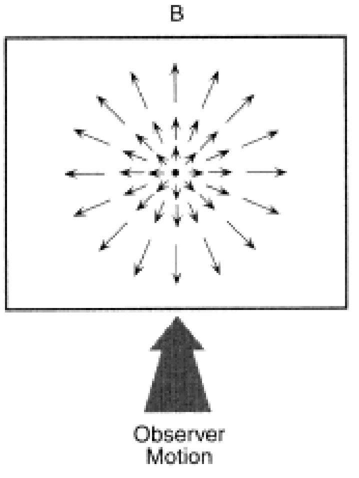
- Both provide only relative information about depth. It does not specify the absolute distance to an object.
- Unlike binocular disparity, however, motion parallax can provide effective depth information at great distances.
  - E.g. if you were to drive past two hills, one of which was 2 miles away and the other 4 miles away, you would be able to determine which one was closer from relative motion parallax.
- Binocular disparity would provide no information in such a case because the distances between the two eyes is too small in comparison with the distance of the mountains.
- With motion parallax, the separation that is achieved from successive views can be much greater, thus affording much greater parallax.

Optic Flow Caused by a Moving Observer

- Observers are usually moving about and actively exploring cluttered environments, engaging in activities that cause complex patterns of optic flow (movement of image features).



The optic flow created by an observer moving leftward (large arrow) while fixating the point in the middle of the line.



The optical expansion pattern that results from an observer moving toward a fixation point straight ahead, as in walking toward a wall.

Optic Flow Caused by a Moving Observer

- These motion gradients are just a few special cases of very simple optic flow patterns that arise from very simple motions with respect to a single environmental surface.
- Optic flow patterns become exceedingly complex as the situation begins to approximate naturally occurring conditions.

Optic Flow Caused by a Moving Observer

- Realistic optic flow patterns are so complex that they could not possibly be catalogued in terms of a few simple types.
- Their structure can be discerned only through sophisticated mathematical analysis. (Longuet-Higgens & Prazdny, 1980; Prazdny, 1980).
- We will discuss these theories in a later chapter.

END OF PART 1

- Ref: handout