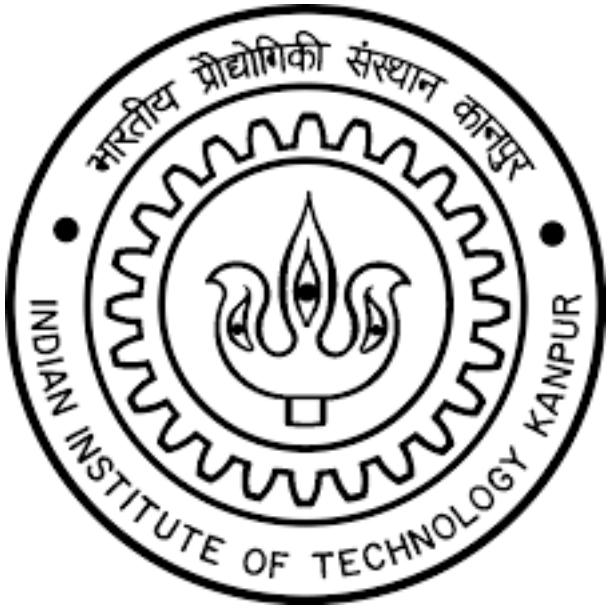


Indian Institute of Technology
Kanpur

References

- Braisby & Gellatly (2005). Cognitive Psychology. *Oxford University Press.*

In Collaboration
with

National Program on Technology Enhanced Learning (NPTEL)

Presents

Course Title:

Basic Cognitive Processes

By: Dr. Ark Verma,
Assistant Professor of Psychology,
Department of Humanities & Social Sciences,
IIT Kanpur

Lecture 19: Approaches to Visual Perception - 2

Marr's Theory of Perception: The 2.5D Sketch

- *"the detection of physical invariants, like image surfaces, is exactly and precisely an information processing problem, in modern terminology. And second, he (Gibson) vastly underrated the sheer difficulty of such detection. … Detecting physical invariants is just as difficult as Gibson feared, but nevertheless we can do it. And the only way to understand how is to treat it as an information processing problem." - (Marr, 1982)*

- While Gibson identified the need for invariants for solving problem of visual perception, he did not specify the possible mechanisms about how the information was "picked up" about these invariants.

- To address, the gap, a theory was needed that attempted to explain exactly how the brain was able to take the information sensed by the eyes and turn it into an accurate, internal representation of the surrounding world. Such a theory was put forward by David Marr (1982).

- Common points between Gibson's & Marr's theories.
  - Like Gibson, Marr suggested that the information from the senses is sufficient to allow perception to occur.
  - Marr adopted an information processing approach in which the processes responsible for analyzing the retinal image were central.
  - Marr's theory is also therefore strongly bottom - up, in that it sees the retinal image as the starting point of perception and explored how this image might be analyzed in order to produce a description of the environment.
    - note that Marr was interested in perception of the kind involved in object recognition & not the one used for action.

- Marr saw the analysis of the retinal image in four distinct states, with each stage taking the output of the previous one and performing a new analyses on it:
    - *Grey level description*: the intensity of light is measured at each point in the retinal image.
    - *Primal sketch*: first, in the raw primal sketch, areas that could potentially correspond to the edges and texture of objects are identified. Then, in the full primal sketch, these areas are used to generate a description of the outline of any objects in view.

- *2.5D sketch*: at this stage a description is formed of how the surfaces in view relate to one another and to the observer.

- *3D object centered description*: at this stage object descriptions are produced that allow the object to be recognized from any angle (i.e. independent of the viewpoint of the observer).

- Marr concerned himself mostly at the computational level and algorithmic levels of analysis and did not say much about the neural hardware that might be involved.

- **The Grey Level Description**:
  - Marr thought that colour information was processed by a distinct module and need not be involved in obtaining descriptions of the shape of objects and the layout of the environment.
    - " Computer scientists call the separate pieces of a process its modules, and the idea that a large computation can be split up and implemented as a collection of parts that are nearly as independent of one another as the overall task allows, is so important that I was moved to elevate it to a principle; the principle of modular design." (Marr - 1982).

- This first stage in Marr's theory acts to produce a description containing the intensity (i.e. the brightness) of light at all points of the retina.
    - A description composed solely of intensity information is referred to as the 'greyscale'.
    - Algorithmically speaking, it is possible to derive the intensity of the light striking each part of the retina, because as light strikes a cell in the retina, the voltage across the cell membrane changes and the size of this change (or depolarization) corresponds to the intensity of light.
    - therefore, a greyscale description is produced by the pattern of depolarization on the retina.

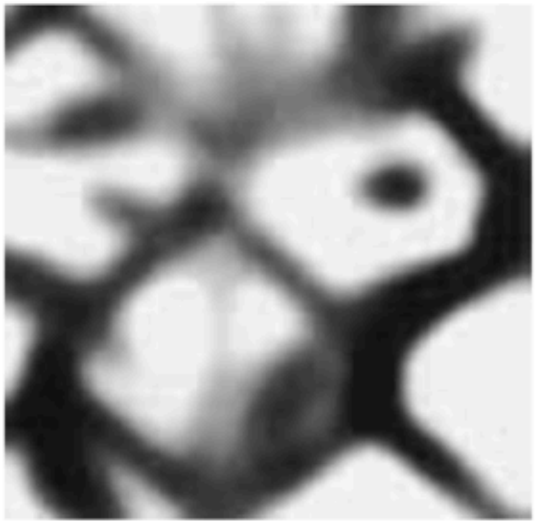
- **The primal sketch**: The generation of the primal sketch, occurs in two phases.
  - The first phase consists of forming a raw primal sketch from the grey level description by identifying patterns of changing intensity.
  - Changes in intensity of the reflected light can be grouped into three categories:
    - Relatively large changes in intensity produced by the edge of an object.
    - Smaller changes in intensity caused by the parts & texture of an object.
    - Still smaller changes in intensity due to random fluctuations in the light reflected.

- Marr & Hildreth (1980) proposed an algorithm that could be used to determine which intensity changes corresponded to the edges of objects, meaning that changes in intensity due to random fluctuations could be discarded. The algorithm made use of a technique called **Gaussian Blurring**, which involves averaging the intensity values in circular regions of the greyscale description.

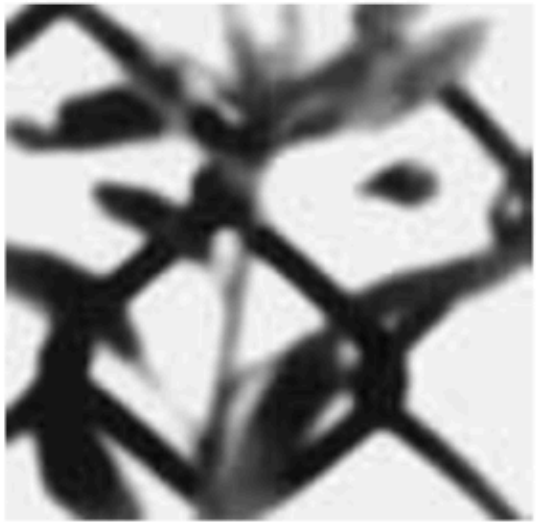
- The values at the center of the circle are weighted more than those at the edges in a way identical to a normal distribution.

- By changing the size of the circle in which intensity values are averaged, it is possible to produce a range of images blurred to different degrees.

(a)



(b)



(c)

Figure 3.17 Examples of Gaussian blurred images

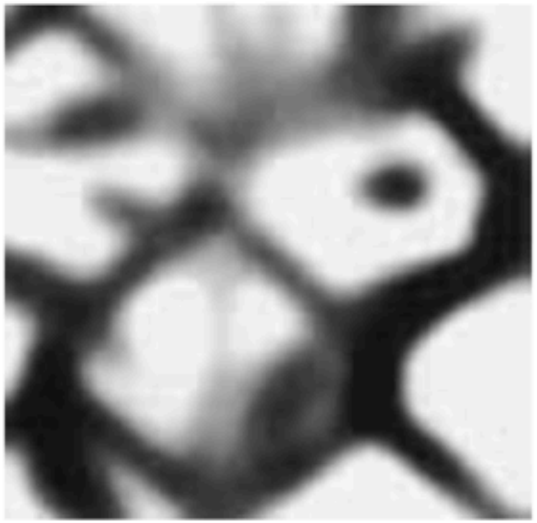
Image: Braisby & Gellatly (2005). Cognitive Psychology. *Oxford University Press.* (Fig. 3.17, p.93)

- Marr & Hildreth's algorithm works by comparing images that have been blurred to different degrees. if an intensity change is visible at two or more adjacent levels of blurring, then it is assumed that it cannot correspond to a random fluctuation and must relate to the edge of an object.

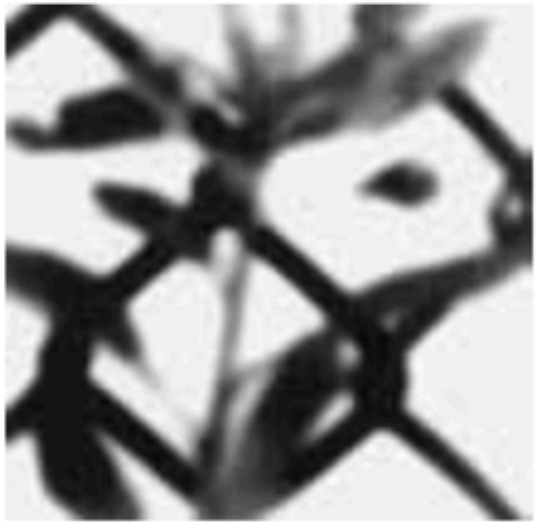
- Although this algorithm was implemented on a computer, there is evidence that retinal processing delivers descriptions that have been blurred to different degrees.

- By analyzing the changes in intensity values in the blurred images, it is possible to form a symbolic representation consisting of four **primitives** corresponding to four types of intensity change:
  - edge - segments: represent a sudden change in intensity
  - bar: a bar represented two parallel edge - segments
  - a termination: represented sudden discontinuity
  - a blob: corresponded to a small enclosed area bounded by changes i intensity

(a)



(b)



(c)

Figure 3.17 Examples of Gaussian blurred images

Image: Braisby & Gellatly (2005). Cognitive Psychology. *Oxford University Press.* (Fig. 3.17, p.93)

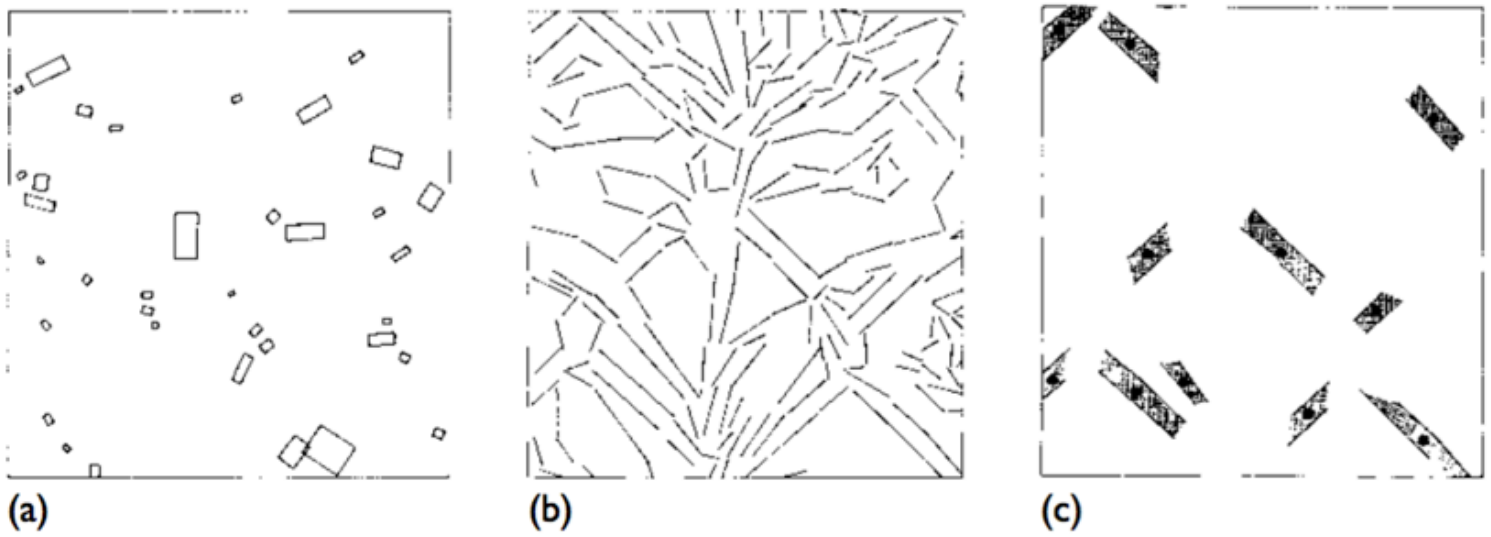


Figure 3.18 Primitives used in the raw primal sketch: (a) blobs, (b) edge-segments and (c) bars

Image: Braisby & Gellatly (2005). Cognitive Psychology. *Oxford University Press.* (Fig. 3.18, p.93)

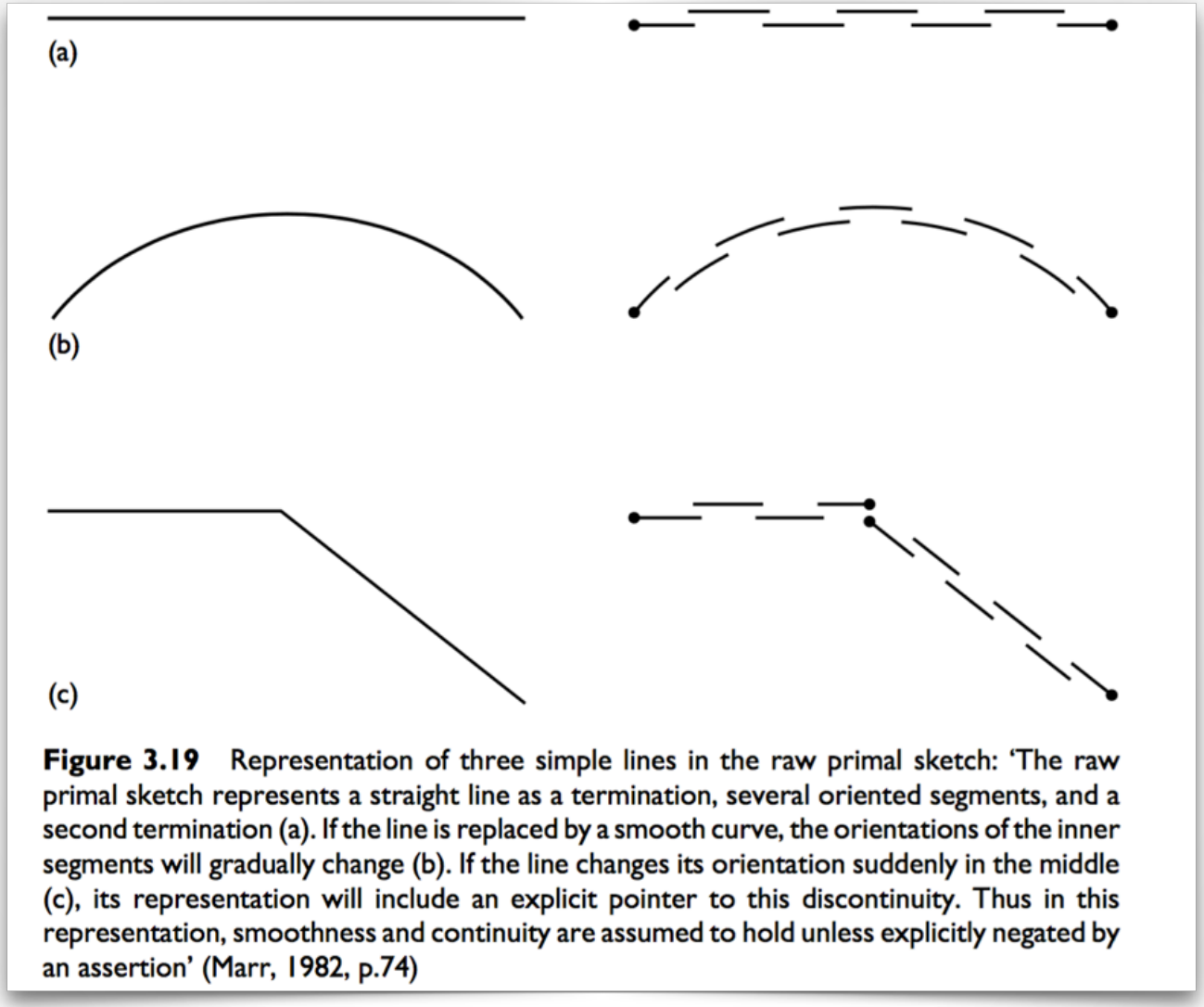


Figure 3.19 Representation of three simple lines in the raw primal sketch: 'The raw primal sketch represents a straight line as a termination, several oriented segments, and a second termination (a). If the line is replaced by a smooth curve, the orientations of the inner segments will gradually change (b). If the line changes its orientation suddenly in the middle (c), its representation will include an explicit pointer to this discontinuity. Thus in this representation, smoothness and continuity are assumed to hold unless explicitly negated by an assertion' (Marr, 1982, p.74)

Image: Braisby & Gellatly (2005). Cognitive Psychology. *Oxford University Press.* (Fig. 3.19, p.94)

- The next step is to transform the raw primal sketch into a description, known as the **full primal sketch**; which contains information about how the image is organized, particularly about the location, shape, texture and internal parts of any objects that are in view.

- Basically, the idea is that **place tokens** are assigned to areas of the raw primal sketch based on the grouping of the edge segments, bars, terminations & blobs.

- If these place - tokens then form a group among themselves, they can be aggregated together to form a new, higher - order place token.

- An example: consider looking at a tiger.
  - the raw primal sketch would contain information about the edges of the tiger's body; but also about the edges & patterns of its stripes and the texture of it's hair.
  - in the full primal sketch, place tokens will be produced by the grouping of individual hairs into each of the stripes. the place tokens for each stripe would then also be grouped into a higher - order place token; meaning that there will be at least two levels of place tokens making up the tiger.

- various mechanisms exist for grouping the raw primal sketch components into place tokens & then for grouping the place tokens together. these include **clustering,** in which tokens that are close to one another are grouped in a way very similar to the Gestalt principle of proximity & **curvilinear aggregation**, in which tokens with related alignments are grouped in a similar fashion to the Gestalt principle of good continuation.

- **The 2.5D sketch:** Marr's modular approach to perception means that while the full primal sketch is being produced, other visual information is being analyzed simultaneously.
    - for e.g. depth relations, distance between a surface & the observer, motion etc.
- Marr proposed that the information from all such modules was combined together to produce the 2.5D sketch. It is called the 2.5D sketch, because the specification of the position & depth of surfaces and objects is done in relation to the observer. thus, the description of the object will be **viewer - centred** and will not contain any information about the object that is not present in the retinal image.
- the viewer centred image is turned into a fully 3D **object - centred** description using some steps (to be discussed later)

- Marr saw the 2.5D sketch as consisting of a series of primitives that contained vectors showing the orientations of each surface.

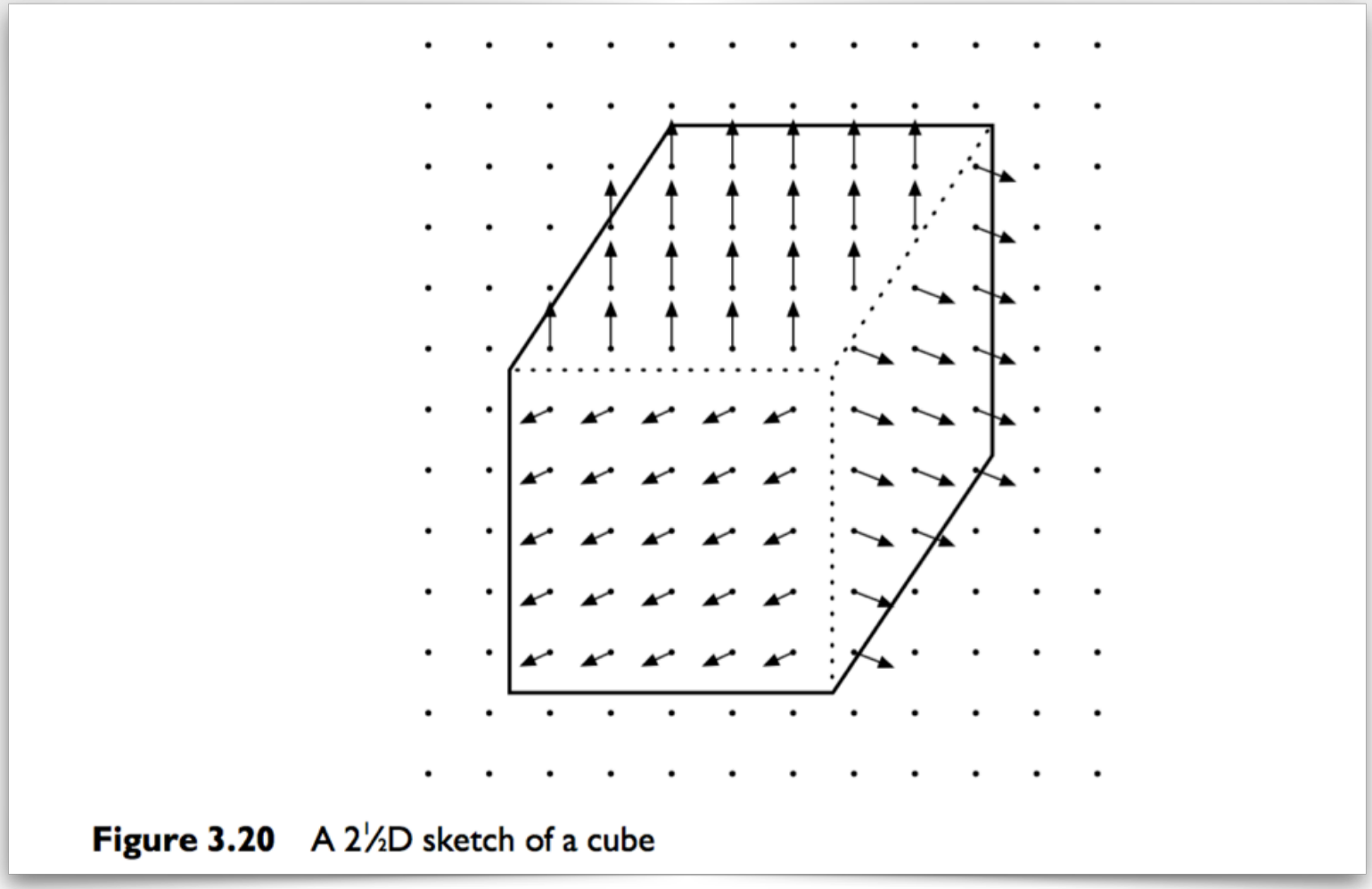


Figure 3.20 A 2½D sketch of a cube

Image: Braisby & Gellatly (2005). Cognitive Psychology. *Oxford University Press.* (Fig. 3.20, p.96)

- **Evaluating Marr's approach**: A lot of research has followed Marr's theory; some of it indeed confirming his proposed mechanisms while some finding out it's shortcomings.

- Marr & Hildreth (1980) proposal of primal sketch being formed by looking for changes in intensity worked well with computer simulations but it could not be guaranteed that the same is followed by human visual system.

- Ens & Resnick (1990) showed that the participants of their study could also use 3D information instead of only 2D information to form a full primal sketch.

- However, Marr's proposal of the integration of depth cues in the 2.5D sketch was supported by experiments done by Young et al., (1993); who reported that the perceptual system does process these cues separately, and will also make selective use of them deepening on how noisy they are.

To Sum Up

- We talked about David Marr's 2.5 D Approach to Perception.
- We saw again that information from the sensory experience can be systematically analysed to construct a good perceptual representation of the world.
- There are however shortcoming & gaps in linking this computational approach to match human performance.