Chapter 4

Perceiving Objects

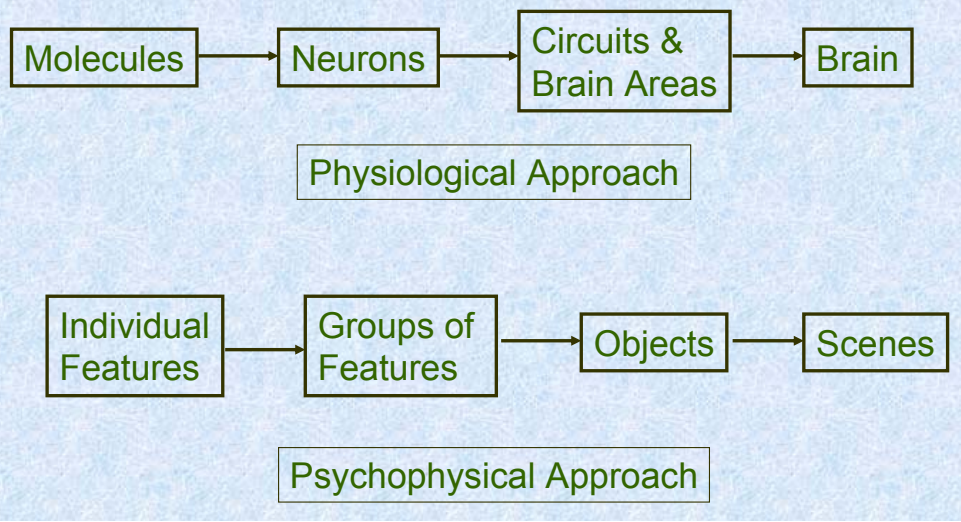

Figure 4.1: Perception being studied from a physiological and psychophysical point of view.

In this course, till now we have seen how the perception takes place from behind the scenes. We have studied it from a physiological point of view. We have studied the physiological aspects at different scales. We have started with molecules, then studied about different types of neurons, how they form circuits in the brain and finally about the brain itself. In this chapter we will study perception from a psychophysical point of view by studying the relationship between the stimulus and perception. In fact, perception can be studied psychophysically also at different levels. We can see how individual features affect perception, then how groups of feature have its effect, and finally objects and scenes affects our perception. These different levels of studying perception both psychophysically and physiologically is illustrated in Figure 4.1. Now we will study some theories of perceptual organization, perceptual segregation, perceptual construction and the role of intelligence in perceiving objects.

4.1 Perceptual Organization

As we have seen from the study of the eye, LGN and brain, it is obvious that we have a sophisticated mechanism of organizing information and then using it for survival. This kind of organization has been studied in a psychophysical manner and models have been proposed to describe our perceptual organization.

4.1.1 Gestalt's Model of Perceptual Organization

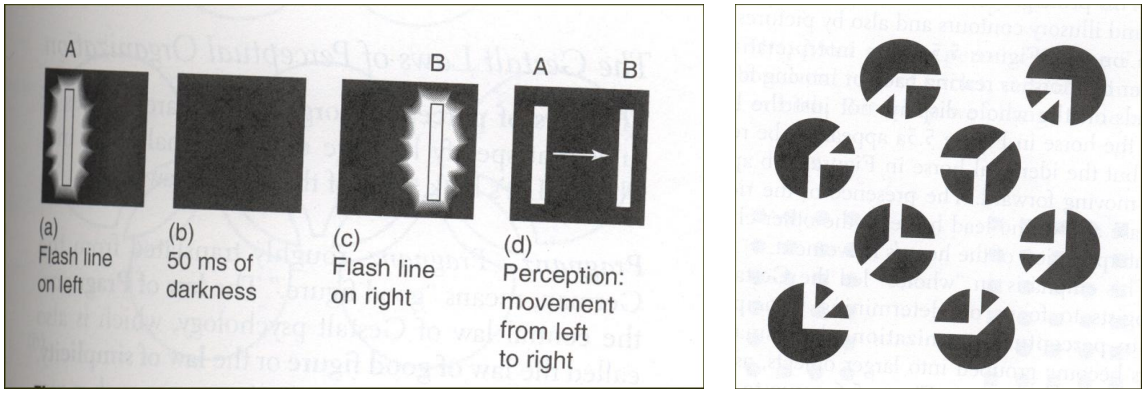

Figure 4.2: Left: Apparent Movement. Right: Vanishing of Illusory Contours.

The first approach and psychophysical model of perceptual organization is called *structuralism*. The concept was first proposed by Wilhelm Wundt in 1879, but was formalized and propagated by Gestalt. Hence it is often called Gestaltism. The first doctrine of structuralism is that perception in created by combining elements called *sensations*. But this single doctrine could not explain many perceptual phenomenon.

- *Apparent Movement:* If two stimuli at slightly different positions are flashed with appropriate timing, it is perceived as if the object has moved from one place to another. This is called *apparent movement*. This is the fundamental philosophy behind movies and television. However, note that the perception of these two images cannot be created by the summation of the perception of each of them. This is illustrated in Figure 4.2.

- *Vanishing of Illusory Contours:* Another such phenomenon is shown in the right image of Figure 4.2. If you see this as a cube hanging in front of black circles, you probably perceive faint illusory contours that represent the cube. However, they are not actually present in the image. You can find this out by placing your hand on one of the black circles and the illusory contours related to that circle vanishes.

These brought in the second doctrine of Gestaltism which says that whole is more than the sum of its parts. This concept is additionally supported by pictures as in Figure 4.3. In the both the images the exact same horse is shown. Yet with different context, in one it seems to be rearing. In the other, it seems to be following the other horse.

Examples like this led Gestalt to describe the second doctrine of structuralism, *the whole is different that the sum of its components*. This basic idea led to development of theories of *perceptual organization*, that is, how small elements are grouped into larger objects. This is defined by a set of *laws of perceptual organization* that specify how we organize elements to wholes.

1. *Law of Simplicity:* This says that every pattern is seen is such a way that the resulting structure is as simple as possible. The familiar olympic symbol in Figure 4.4 is hence perceived as five circles and not the more complicated assembly to nine shapes.

2. *Law of Similarity:* This tells that similar things appear to be grouped together. This is why in Figure 4.5, the squares and circles are grouped together by their shape (left), the circles are grouped together by their lightness (middle), and the dancers are grouped together by their orientation (right).

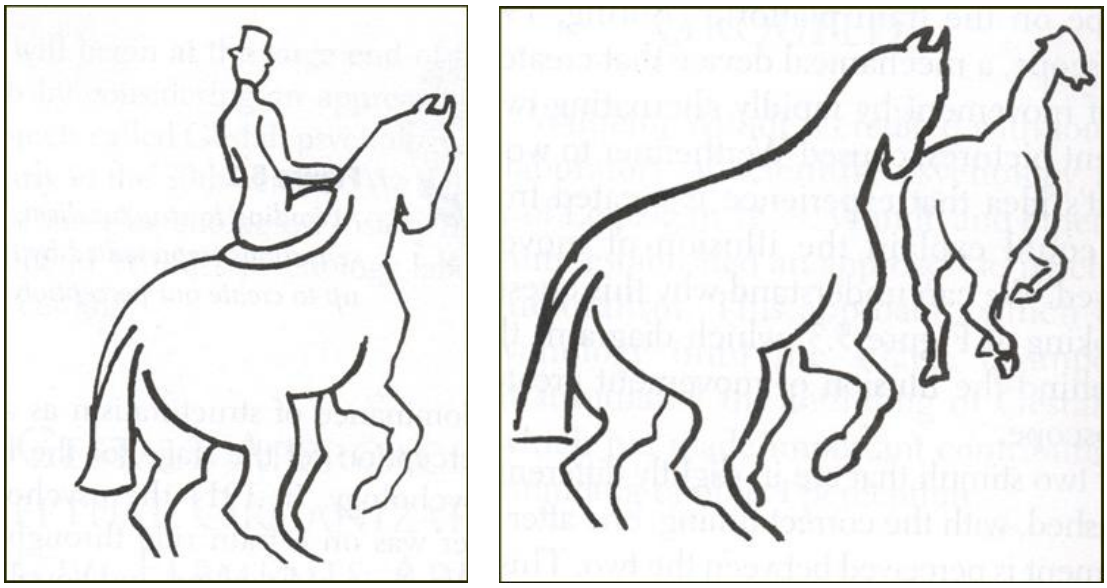

Figure 4.3: Left: A rearing horse. Right: A following horse.

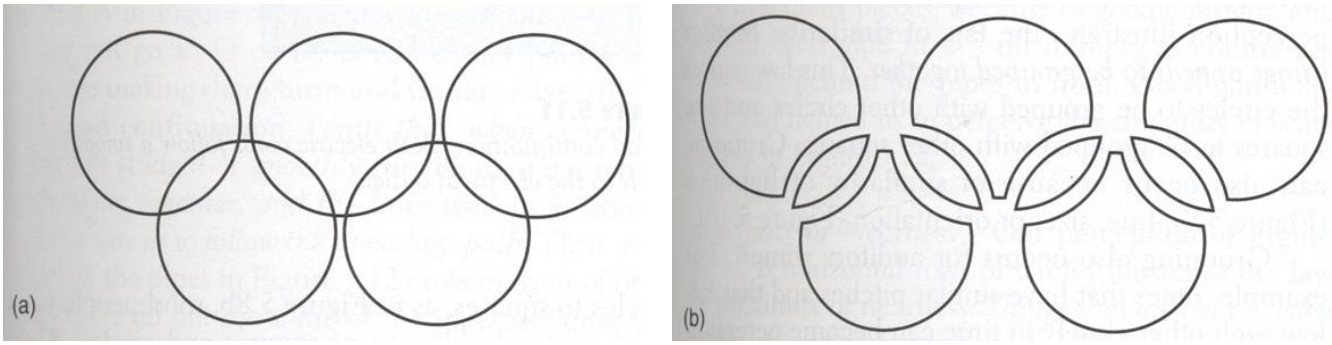

Figure 4.4: (a) is usually perceived as five circles and not as the nine shapes in (b).

3. *Law of good continuity:* This tells that points when connected in straight or smoothly curving lines appear to belong together. In Figure 4.6, notice how we see the electric chord from A to go to B and not to D. Similarly the chord in C seems to continue to D and not to B. This is because both the chords follow a smooth path. Similarly, even when the pipes in Figure 4.6 cross in the front, we do not perceive them as broken pieces of pipes but as each pipe continuing behind the other.

4. *Law of proximity:* This tells that objects that are close together appear to be grouped. In Figure 4.7 note the the perception of the horizontal rows does not change even when the alternating circles are replaced by sqaures. Note the similarity of this example with the example in Figure 4.5 (left). In this case, law of proximity overrides law of similarity.

5. *Law of common fate:* This states that thing moving in the same direction appear to be grouped together as shown in the case of the dancers in Figure 4.7. This fact is extensively used for choreography.

6. *Law of Familiarity:* This says that things are more likely to form groups if they appear familiar or meaningful. In the image shown in Figure 4.8, many people find it difficult to find faces at first. Then they succeed. The change in perception from ”trees, streams and horses in the forest” to ”faces” is a change in perceptual

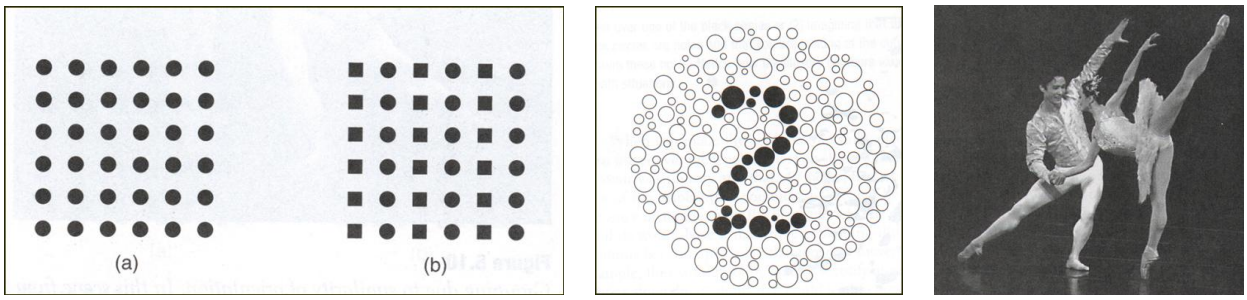

Figure 4.5: Left: (a) is perceived as both horizontal rows or vertical columns, nut (b) is perceived as vertical columns. Middle: Grouping by lightness. The lighter objects form a group and the darker objects form another group. Right: Grouped by orientation. In this picture from Swan Lake the perceptual unity of the dancers are enhanced by the similar orientations of their arms and bodies.

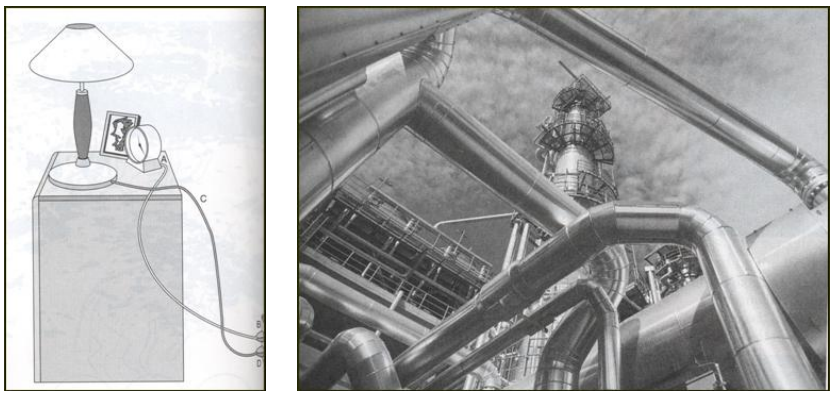

Figure 4.6: Left: Both chords follow a smooth path to the power point. Right: Pipes in a Refinery.

organization. In fact, once you start seeing the faces, you will see that it is difficult to go back and forth between the two perceptions due to the change in perceptual organization that is needed to do so.

However note that Gestalt's laws are more like good heuristics and not a definitive algorithm. Hence, it may not work for all cases. For example, in Figure 4.9 we perceive the branch being occluded by a tree. But, the branches having a good continuity from the tree can be perceived as belonging to the tree. This does not violate any of the other laws and is a valid perception by Gestalt's law of continuity. However, in this case the heuristic does not work.

The next question is where these heuristics come from. It seems that learning plays an important role. After all, we have been practicing "perceptual problem seeking" from childhood and have immense experience of the common regularities and irregularities found in our everyday environment. These regularities are the basis of Gestalt's heuristics and hence they work most often. More recently, one more theory of perceptual organization has been proposed by Stephen Palmer and Irvin Rock. More quantitative analysis has been carried out for the Palmer-Irvin model to give it more solid basis.

4.1.2 Palmer-Irvin Model of Perceptual Organization

Palmer and Irvin proposed three new grouping principles: principle of common region, principle of element connectedness and the principle of synchrony.

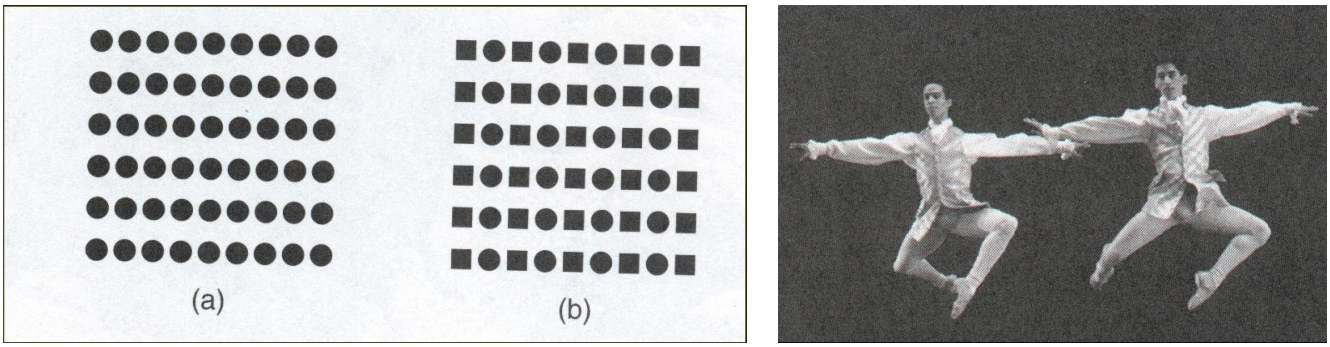

Figure 4.7: Left: Example of law of nearness. (a) is perceived as horizontal rows of circles and (b) is still perceived as horizontal rows even though alternating circles have been changed to squares. Right: Example of law of common fate. These two dancers are grouped not only by orientation but also by the common fate that they are moving in the same direction.

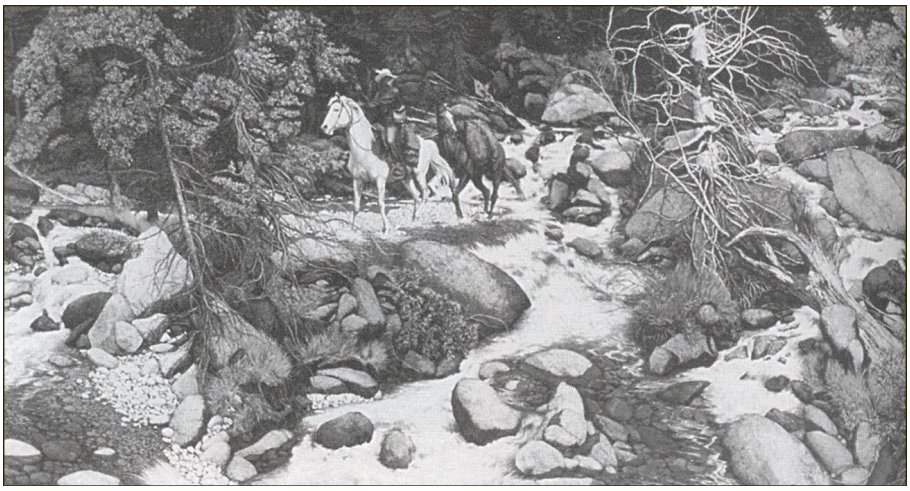

Figure 4.8: The Forest has Eyes by Bev Doolittle (1985). Can you find 12 faces in this picture?

- *Principle of Common Region:* This says that elements that are within same region of space are grouped together. In Figure 4.10, even though the dots inside the ellipses are farther apart than the dots that are next to each other in the neighboring ellipses, we see the dots inside the ellipses as belonging together. This happens because each ellipse is perceived as a separate region of space. Note that this case contradicts the law of proximity by Gestalt in the presence of explicit separation of space that was not studied by Gestalt.

- *Principle of Connectedness:* This says that things that are physically connected will be perceived as an unit. For example, in Figure 4.10, we perceive a series of dumbbells, even though the dots separated by spaces are closer together than the dots connected by lines.

- *Principle of Synchrony:* This states that visual events that take place at the same time (synchronously) will be perceived as being grouped together. For example, the lights in Figure 4.10 that blink together are seen to belong together. Note the similarity of this principle with Gestalt's law of common fate. Both are dynamic, but synchrony can happen without movement. It is more related to a synchronous change rather than direction of movement.

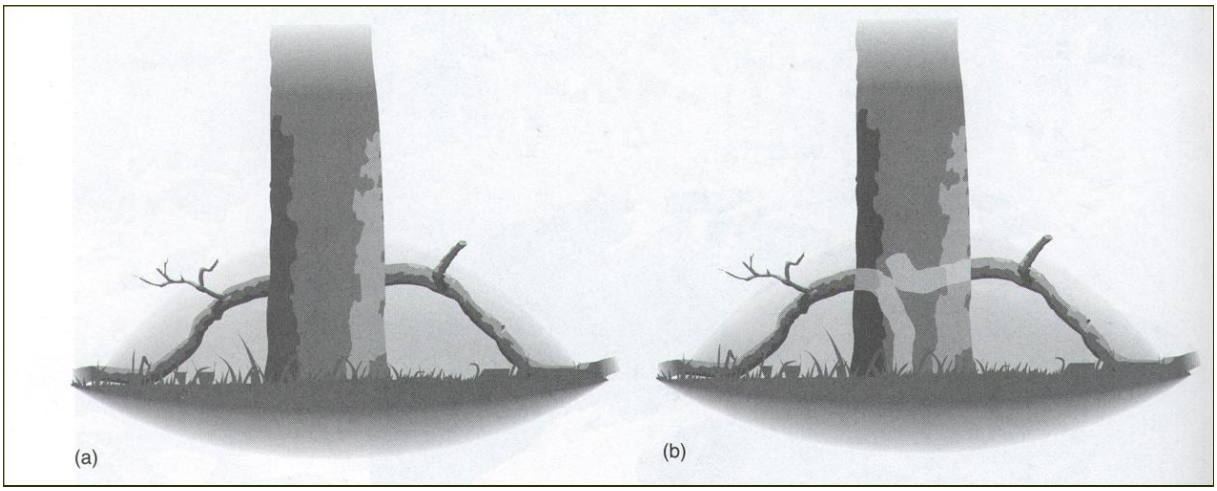

Figure 4.9: (a) Is this a single branch behind a tree? Most of us would perceive this in that fashion. (b) But by laws of good continuity we should have perceived the branches as the part of the same tree.

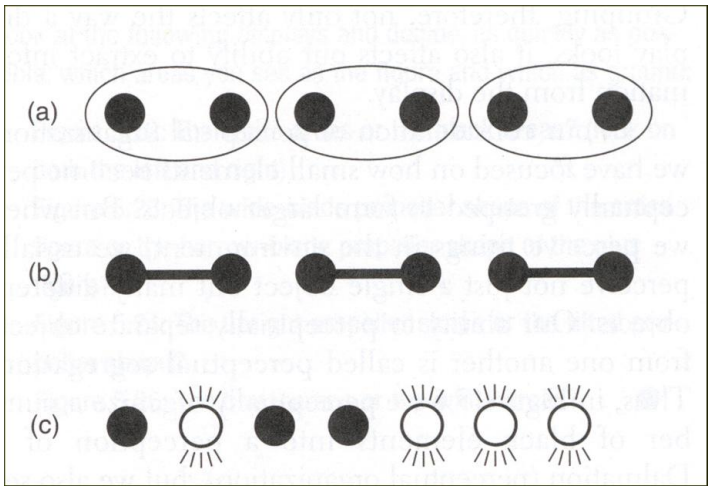

Figure 4.10: Grouping by common region (a), connectedness (b) and synchrony (c).

4.1.3 Does Perceptual Organization Help?

Now that we have studied all these heuristics of perceptual organization, the question to ask is, "Does this really help us?" To answer this question, quantitative measurement of grouping effects on our actions were studied by a technique called *repetation discrimination task*. For this users were presented with moving series of circles and squares and were asked to respond when they see similar shape one after another. These shapes were not place equally distant from each other, but was grouped as shown in Figure 4.11. A pair of shapes were placed close to each other to form a group and a series of such pairs were presented. It was found the reaction time to identify adjacent similar shapes changed with which groups they belong to. If the similar adjacent shapes belonged to the same group, the reaction time was about 700 ms. If they belonged to different groups, the reaction time was about 1150 ms. This proved that perceptual organization does play a role in our actions.

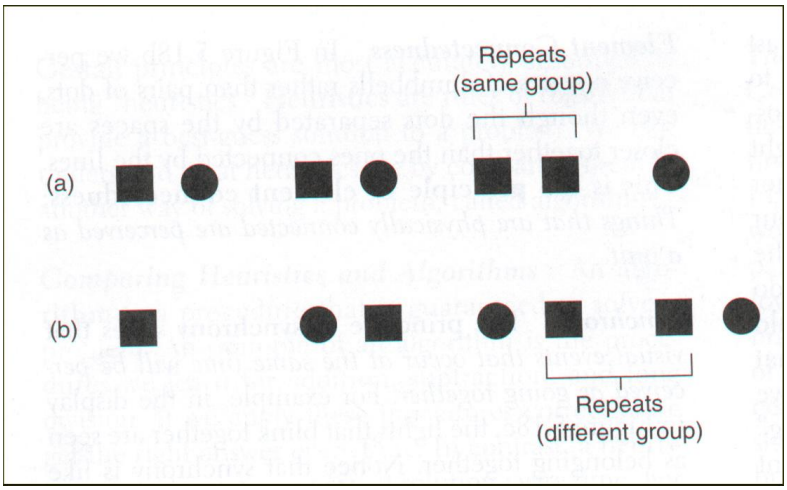

Figure 4.11: Repeatation Discrimination Task

4.2 Perceptual Segregation

So far we have discussed how do we achieve perceptual organization. But while perceiving a stimulus we also separate out objects from one another. The most important separation that we perform extremely efficiently everyday is separating the foreground from the background. This is what we call *perceptual segregation*. Both Gestalt and many recent psychologists have been concerned about this. This is often called the *figure-ground* segregation problem. This indicates that when we see a object, it is usually seen as a *figure* that stands out from its background, which is called the *ground*.

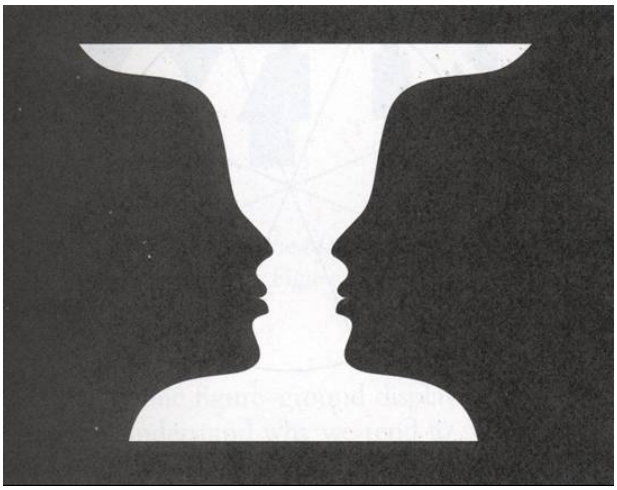

Figure 4.12: A version of Rubin's reversible face vase figure.

4.2.1 Gestalt Approach

Again Gestalt and the psychologists of their league were the first to study perceptual segregation. One way of studying this was by considering patterns like in Figure 4.12 introduced by Danish psychologist Edgar Rubin in 1915. This is an example of reversible figure ground images, since it can be alternatively perceived as a pair of

black faces looking at each other on white background, or a white vase on a black background. From this, some basic properties of figure and ground was proposed as follows.

- The figure is more memorable or more familiar to an object than the ground.
- The figure is seen as being in front the ground.
- The ground is seen as unformed material that seems to extend behind the figure.
- The contour separating the figure and ground seems to belong to the figure.

If you observe Figure 4.12 carefully, you will find that each of these is true for each of the different perception of the figure. In fact, this is the reason that simultaneous perception of this figure both as a vase and the twin faces is difficult, if not impossible. When you perceive the black faces as figure, the white is perceived as unformed material extending behind the faces and hence cannot be perceived anymore as a vase. The fact that contours belongs to the faces in that case makes it difficult to perceive the vase.

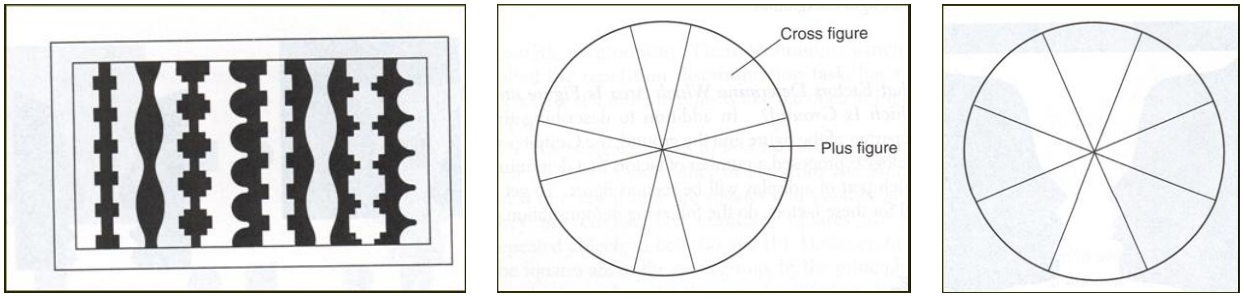

Figure 4.13: Left: Symmetry and figure and ground. Middle: Area and figure and ground. Right: Orientation and figure and ground.

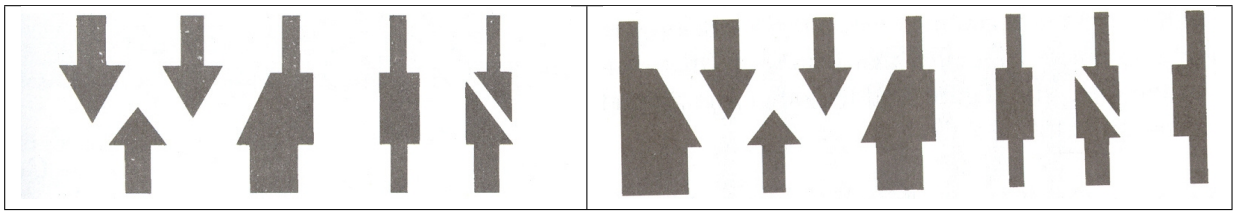

Figure 4.14: This shows the role of familiarity in deciding figure and ground.

In fact, there is a number of factors that contribute to do this perceptual segregation. To get a feel of what there are check out a quickly as possible which areas appear to be figure and ground:

1. In the left image of Figure 4.13, the white or black areas?
2. In the middle image of Figure 4.13, the wide blade propeller shape of the cross or the narrower blade shape?
3. In the right image of Figure 4.13, the upright or the tilted shape?
4. In left image of Figure 4.14, the white or black areas?

Though there is no “correct” perception of this displays, but experiments have shown that certain properties of stimulus tend to influence which areas are seen as figure and which are seen as ground.

1. *Symmetry:* Symmetrical areas tend to be seen as figure. This is the reason, in the left image of Figure 4.13, black areas appear to be in foreground on the left and the white areas appear to be in foreground on the right.

2. *Area:* Stimuli with comparatively smaller area tend to be seen as the figure. This is the reason, the narrow propeller blades in the middle image of Figure 4.13 tend to be seen as foreground.

3. *Orientation:* Vertical or horizontal orientation are more likely to be seen as figure. This is why the upright shape appears to be the foreground in right image of Figure 4.13.

4. *Familiarity:* Finally, meaningfulness or familiarity as mention in perceptual organization also, plays an important role in perceiving an object as figure. Note that in left image of Figure 4.14, the familiar black arrows make the black areas appear to be in the foreground. However, note that this perception changes when the extra black areas are added to the left and right of this same pattern in the right image of Figure 4.14. Now the word 'WIN' in white is in foreground.

4.2.2 Modern Ideas

Modern researchers have extended Gestalt's study by looking more closely as to *how* the underlying figure-ground segregation happen. The two important questions that they have tried to study are,

- What is the role of contours in this process?
- Exactly when does this figure-ground segregation occur?

Role of contours

The role of contours is probably very important. Study the image of Figure 4.12 and ask yourself what is probablity of the scanerio where two objects like the faces and the vase will have exactly similar contour and they will coincide with each other. Though it is not entirely impossible for this to occur, it is highly like. We have seen before that we tend to perceive things that are familiar or likely to occur. This high unlikeliness of the event makes it so difficult for us to perceive the two meaningful things at the same time.

When does this segregation happen?

In fact, researchers are not yet sure when this happens. Some experiments show that this figure-ground segregation happens after perception, some other experiments show that it happens before perception and yet some other show that these two may be parallel process. The idea that this timing is difficult to define is evident from our physiological knowledge since we saw that the physiological happenings are not linear. There are lot of feedbacks involved between the higher and lower center of visual processing and lot of cross talk between the different areas of the brain.

4.3 How Objects are Constructed?

The models we discussed so far were dealing with the higher levels of the psychophysical process of perception described in Figure 4.1. There has been several models which have developed the theory starting at the very beginning of the diagram. These models believe that the objects are broken down into small elements which are then analyzed in a systematic computationally organized fashion to create the perception of a object. We will study three such models here and then combine the finding from these to create a more comprehensive computational model for visual perception.

4.3.1 Marr's Computational Approach

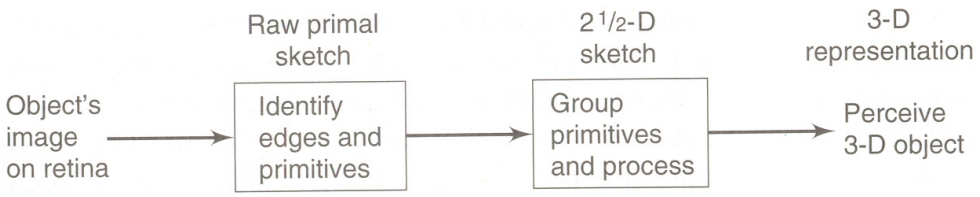

Figure 4.15: Flow diagram describing Marr's computational model.

Marr's model is called the *computational approach* since he assumes that vision happens in such a way as if a computer is programmed to do it. Figure 4.15 shows the flow chart of his model. According to Marr, perception starts with the image of the object on the retina. This image is analyzed to determine light and dark areas, especially edges. This determines a collection of basic features called *raw primal sketch*, that includes closed areas like circles, ellipses and also segments of lines and corners that defines the important features of an object. The challenge for the visual system lies in identifying these features. Marr believed that the visual system achieves this raw primal sketch in using two informations.

- By analysis of the changes in image intensities of the retinal image.
- By taking into account the *natural constraints*. This is what we called regularities in the environment while discussing Gestalt's law of perceptual organization. For example, edges in the object can be created by both illumination and change in shape. To determine, which edge represents what is a big challenge for the visual system. Marr believed that illumination edges are usually sharp whereas those due to change in curvature or shape of the object is often sharp.

This raw primal sketch thus generated is not something that we see. This primal sketch is then further processed using different rules of perceptual organization and segregation to create a 2.5D sketch, which is then transformed to the 3D objects that we see.

Note that Marr's model was just the first step towards development of computational model of vision. Marr died early due to leukemia and his model had many holes. However, it acted as one of the major sources of inspirations for later vision scientists to go for more sophisticated models. For example, one important theory that was developed around the same time being inspired by Marr's theory is called the *feature integration theory*.

4.3.2 Feature Integration Theory

Figure 4.16 shows the flow chart of the feature integration theory (FIT). According to this theory, there are two stages of perception. In the *preattentive stage*, several basic features are detected. This is followed by a *focussed attention* stage when these features are combined together to perceive the object. In the next stage this perceived object is compared with the database of different objects in the memory for recognition.

Basic Feature Determination

FIT identifies the *pop out* property as the most important factor for identifying basic features. This popping out can happen in many ways. Here we are going to give two examples. The first is of *pop out boundaries*. When two sets of elements are displayed in adjacent areas to create a textured field, the boundary between can pops out. Whether this will happen depends on the features of the elements forming the pattern and the values of the features. For example, in Figure 4.17, the first image has same feature but different orientation which creates a

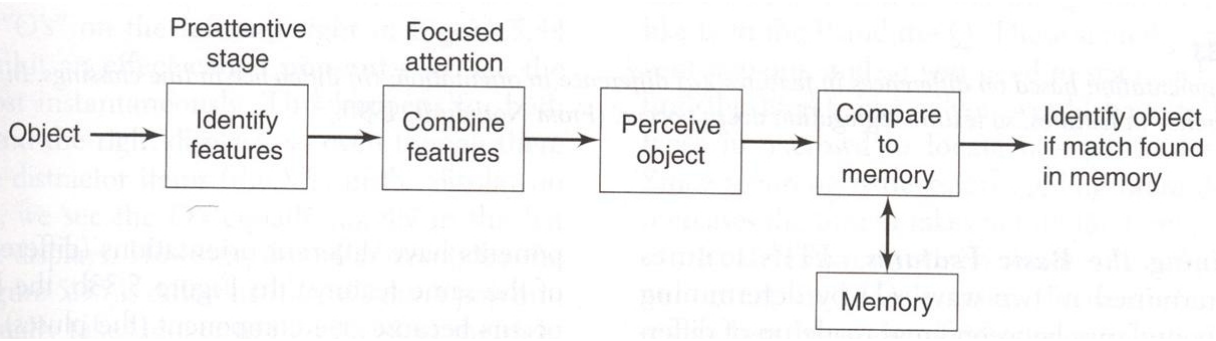

Figure 4.16: Flow diagram describing the feature integration theory.

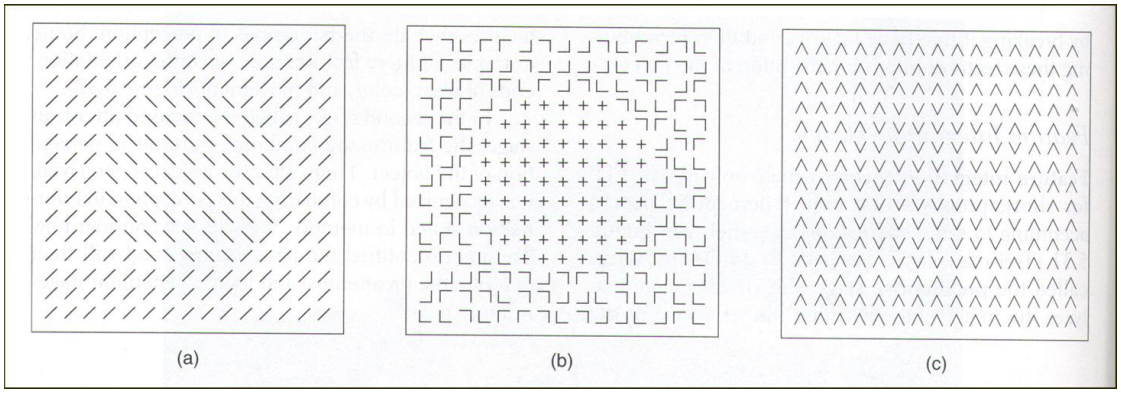

Figure 4.17: Pop out boundaries.

pop-out boundary, the second image have different features and create a pop-out boundary, and the third image do not have a pop-out boundary even though it has different values of same feature. The second example of how popping out can help in determining features was explained by the following experiment.

In Figure 4.18 the test patterns of the experiment is shown. The subject is presented with a set of distracters and a target. The subject is asked to search the target and the time he takes to accomplish this noted. The distracters are then increased and the effect of this increase on the search time is recorded. When the target was 'O' and the distracter was 'V', it was found that the search time was constant irrespective of the number of distracters present. But when the distracters were 'P'and 'Q' and the target was 'R', the visual search time increased linearly with the increase in the number of distracters. This curves are shown in Figure 4.18.

The reason to this has been attributed to how much the target pops out. 'O' has completely different curvature than 'V' and hence pops out from a sea of distracters. As a result, the subject does not have to search for 'O', it just pops out at him. On the other hand, parts of 'R' have exactly similar features as 'P' and 'Q' and hence it does not pop out. So, the subject has to search through every distracter to locate the target and hence the search time increases linearly.

Similar experiments were performed to find the important properties that make a feature pop out. Some of them were identified as curvature, tilt, line ends, movement, color, brightness and direction of illumination.

The next important thing that FIT shows is that in this preattentive stage the basic features are detected but they are all independent, with no combination made to associate them together with one object. To prove this the following experiment was performed. A red cross, blue S and a green T was flashed in the perceptive field for $\frac{1}{5}$ sec and then a random dot mask was flashed to remove any afterimage that can be parsed by the brain. When asked, the subjects could identify the colors they saw and the shapes they saw but ended up wrong associating the

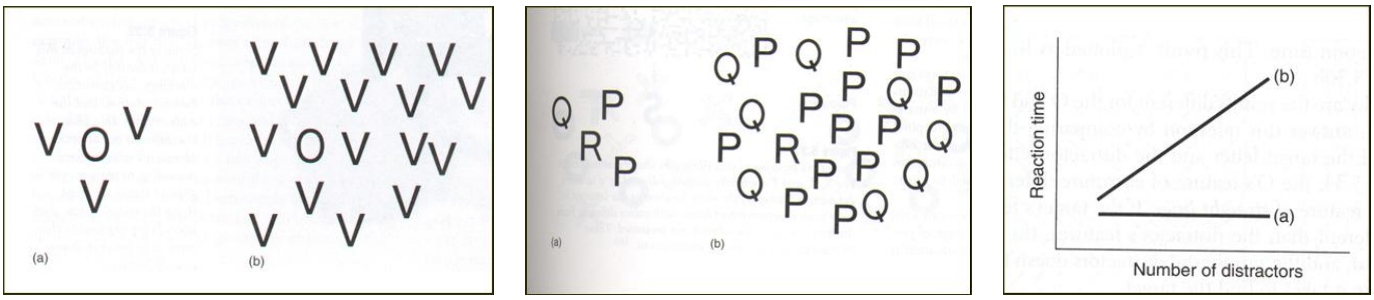

Figure 4.18: Left: The target 'O' is found equally fast irrespective of the number of distracters 'V'. Middle: The targer 'R' takes longer time to search as the number of distracters ('P' and 'Q') increases . Right: The plot of the search time with different levels of "pop outedness".

colors with shapes, i.e. which shapes had which colors. This showed that this association happens later. The fact that these features exist independently is unusual, but remember that we learnt that all these features are detected in different areas in the brain and are tied together by studying the synchronization in the firing pattern which supports this idea.

Combining the Features

FIT says that focussed attention is needed to combine the features. This was demonstrated by the following experiment. A green X was briefly flashed on a display containing many red Os and blue Xs. The subjects task was to identify if the green X was displayed and if so, where. It was found that the subjects were able to identify if the green X was displayed but failed to find its location. Next, instead of a green X a blue O was flashed. This time the subjects were able to tell the location also.

The reason for this was explained as follows. Since the green X is different in color from the red Os and blue Xs, it automatically stands out and the subject does not need to pay attention. So the location of X is not detected. However, when blue O is the target, it has same color as Xs and same shape as the Os. This forces the subject to pay attention and then he or she can tell the location of the blue O. Note that we have seen enough evidence before that attention plays an important role in perception and hence this finding is hardly surprising.

4.3.3 Recognition by Components

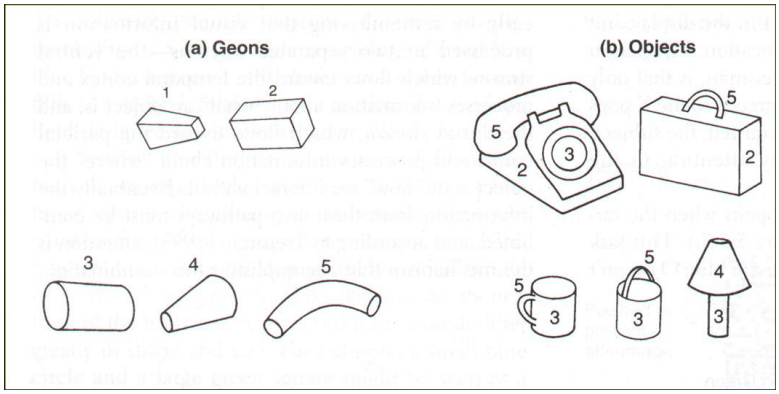

Figure 4.19: Geons used as primitives for Recognition by Components Model.

The third model that has been in vogue is the recognition-by-components (RBC) model. This model assumes volumetric primitives - three dimensional objects corresponding to object's parts - celled *geons*. Figure 4.19 shows some geons and how they combine to form objects. This theory proposes that a scene or object is analyzed into these volumetric 3D features called geons. Biedermann, the proponent of this theory, identified 36 different types of geons. He defined geons as volumes that would satisfy the following properties.

- *View invariance:* They can be recognized and identified from almost any views. This are attributed by the invariant geometric properties of the geons. For example, three parallel edges of a rectangular solid.
- *Discriminability:* The geons can be discriminated from each other from almost any view.

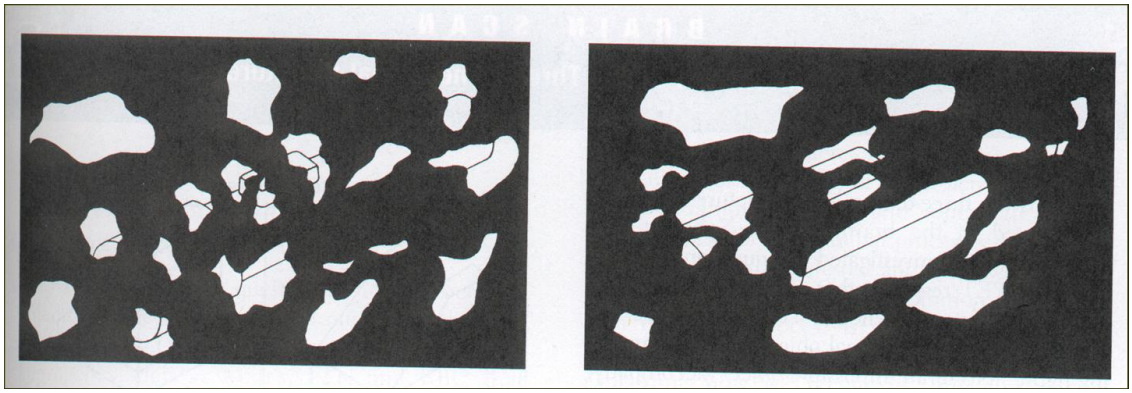

Figure 4.20: The object cannot be perceived when the geons are obscured (left), but can be perceived as a flashlight when the geons are visible (right).

- *Resistance to Visual Noise:* The geons can be perceived even under noise conditions. This is illustrated in Figure 4.20. Note that when the geons are obscured it is difficult to perceive the objects from its parts, however when the geons are visible, it can be perceived as a flashlight.

The basic principle of RBC is if there is enough information for us to identify the geons, we can identify the object, else not. This is called the *theory of componential recovery*. The strength of this theory lies in the fact that we can identify objects based on the knowledge of a few basic shapes. However, it cannot explain how we identify the difference in details. For example, two birds may have different types of beaks and yet can be described by the same geons. But we identify them differently based on their beaks. Thus, it differentiates between how we see objects that differ grossly, but not the ones that differ only in details.

4.4 A Comprehensive Model

All these models above are usually very qualitative. It is difficult to approach vision quantitatively using these. So, what vision scientists did was to combine all the elements and come up with a model that is more quantitative and algorithmic in nature. This model describes visual perception as having four stages: the image based stage, the surface based stage, the object based stage and category based stage.

4.4.1 Image Based Stage

The retinal image is the input to the image based step. The optical image that strikes the retina is completely continuous, but its registration in the receptors of the eye is discrete. So, this image is represented by a 2D array

picture elements, or *pixels*, the two dimensions denoted by x and y. Due to the non-uniform distribution of the receptors in the eye, the discretization is complex. However, in the formal and computational theories of vision, it is almost always simplified as a uniform discretization where every pixel is of same dimensions. The intensity at any pixel (x, y) is given by $I(x, y)$. Of course, this image is a set of numbers and looks completely meaningless. An example of such a retinal image is shown in Figure 4.21.

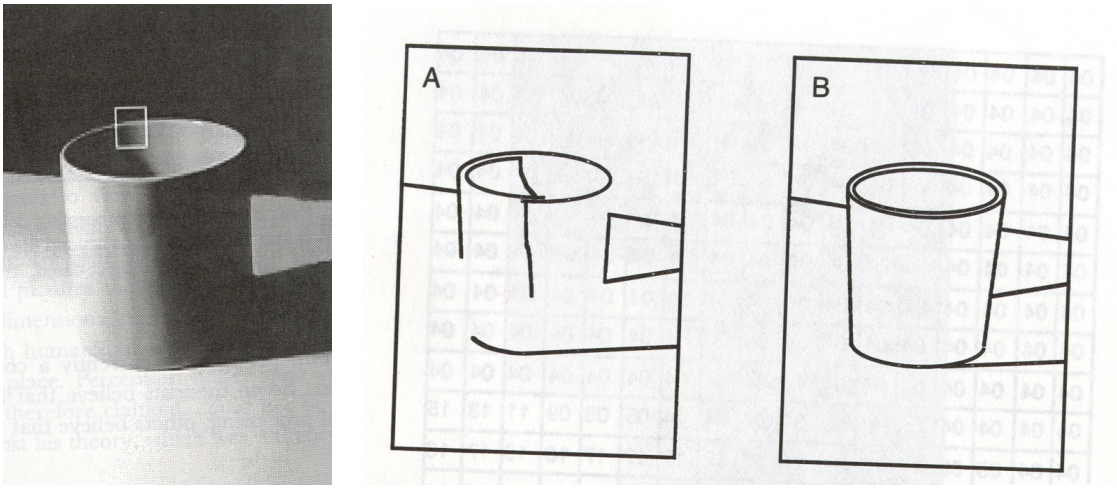

Figure 4.21: Left: The retinal image of a cup. Right:(a) The edges in the image as detected by an edge detection algorithm. (b) The edges found by the computational method is not the same as that of the clean line drawing which is more akin to what we perceive.

In this step, several image based operations are performed, like detecting local edges and corner, combining them with global edges and corners, detecting other features like blobs or squares, finding the correspondence between the left and right eye images. Though these may seem simple operations for us, they are computationally very complex. Figure 4.21 shows an example. It shows the cup image on which a computational edge detection algorithm is performed. Note that the edges detected are not same as the clean line drawing of the image which is much closer to what we see. There are edges detected which we often eliminate, maybe because they are soft edges or are illumination edge. At the same time, very important edges that we perceive right away in a scene are omitted.

Vision scientists call the images that are generated by the image based operations as *primal sketches*. However, they suggest that there are two of them. The first that is formed by local and elementary detection of lines and features, to some extent like (a) of Figure 4.21. This is called the *raw primal sketch*. The second one includes the global grouping and organization of these features and is called the *full primal sketch*. This is to some extent like (b) of Figure 4.21. Note that this part of the model is similar to the concept of Marr's primal sketched, but refined and formalized more to adapt to an algorithmic definition.

The important components of this stage are the following.

1. *Image based Primitives:* The primitive elements are 2D structure of the intensity image like edges and corners rather than external objects of the 3D world.

2. *2D geometry:* The geometry of spatial information is purely two dimensional.

3. *Retinal Coordinate Frame:* The coordinate frame within which these 2D features are analyzed is specified relative to the retina, in the sense that the principle axes are aligned with the eye.

4.4.2 Surface Based Stage

This stage is concerned about recovering the intrinsic property of visible surfaces in the external world that may have formed the primitives detected in the previous stage. The fundamental difference of this representation from the image based stage is the actual representation of the layout of the visible surfaces in the 3D world, as opposed to features detected in the 2D image. This is often called the *intrinsic image*. Note that similarity of this stage with Marr's 2.5D sketches. However, Marr could not formalize these 2.5D sketches which we are going to do here.

Note that constructing the surface represenatation is the first step towards recovering the complete 3D information from the 2D world. Hence, it does not contain information about all the surfaces, but just the *visible one*. But note that they cannot be computed from the retinal image without some additional constraints. But these additional constraints required are very few and almost always true. Since, this representation contains only the visible surface, this can be thought of as a rubber sheet with a texture on it being wrapped over the surfaces in the external world light reflected from which can reach our eye.

To represent these surfaces, small locally flat 2D patches are used. The theory is that any complicated surface can be approximated by infinitely large number of planar elements. Each of these elements have three information: the distance from the viewer, the slant, and the color either in terms of texture or color gradients. This representation is formed from different sources like stereopsis, motion parallax, shading and shadows all of which we will discuss in details in the subsequent chapters.

So, the important components of this stage are the following.

1. *Surface Primitives:* These are local patches of 2D surface at located with some slant from the view direction, at some distance from the viewer and has some texture or color.
2. *3D geometry:* Though the surfaces are themselves 2D, their representation is embedded in 3D space.
3. *View Centered Coordinate Frame:* The coordinate frame within which these surfaces are described is defined in terms of the viewer's distance and orientation with respect to the object. This is called *view centered reference frame*.

4.4.3 Object Based Stage

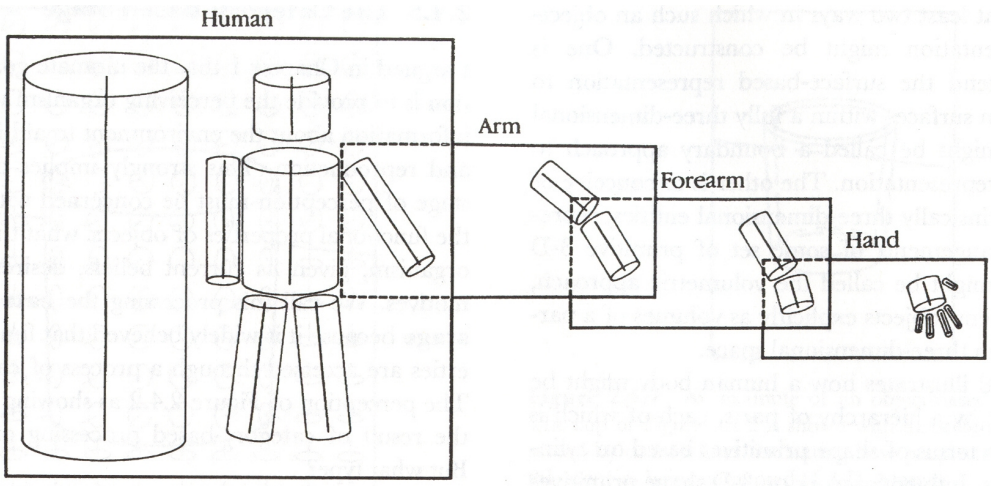

Figure 4.22: Using hierarchical volume representation for a person's body.

Visual perception clearly does not end with surface based representation. In that case we would get surprised if a hidden surface gets exposed which hardly happens. This says that there is an object based stage of our visual

perception when we generate complete 3D representation of objects. However, the volumetric representations that we use can be different. One can be continuing to use 2D patches embedded in 3D to represent the external surfaces of the objects. The second is use volume primitives, that is small elements of volumes to create larger volume. This is basically extending the idea of pixels to 3D. These volumes can in turn be hierarchically arranged as shown in Figure 4.22. Note the similarity of volume based description with the recognition-by-components model.

The components of this stage as thus the following.

1. *Volumetric Primitives:* Objects are defined using volumetric primitives which can either be 2D patches embedded in 3D or 3D volumes.

2. *3D geometry:* The primitives are either embedded in 3D or are in 3D.

3. *Object Based Reference Frames:* The coordinate system within which the relationships are defined can be with respect to each objects one 3D space. This is possible since we have the whole 3D information now, as opposed to 2.5D information in the previous stage. Thus, we can identify a global coordinate system that is not necessarily related to the viewer.

4.4.4 Category Based Stage

This final stage, though very important, is very difficult to quantify. We mentioned that our perception helps us to perceive and understand the functions of different objects around us with respect to us also to the other objects surrounding us. This is defines as a two stage process.

1. The first part is categorization where we use all the information achieved from the previous step to categorize the object to be of a particular class.

2. The second part that comes with this identification, is the retrieval of a large body of information regarding how it can be accessed and used, expectations and behaviors.

This step basically talks about the knowledge in the perceptual process and deals with the cognitive science behind it. There are several views amongst scientists about how this happens and studied mostly in the discipline of cognitive science.

4.5 Role of Intelligence and Experience

The first three stages of the above model is computationally definitive but extremely hard to achieve. Computer vision is striving for years to achieve these. One important reason for this is what we call *intelligence*. In almost all situations in everyday life, we rely on our intelligence to resolve many confusions. Intelligence is very difficult to model. It is exactly what computers lack that make the job so difficult for computers. Here are some examples.

1. *Stimulus is ambiguous.* Objects viewed from one viewpoint give us ambiguous information as shown in Figure 4.23. This is mainly due to the inverse projection problem that we have discussed so much. Loss of 3D depth creates these kinds of ambiguities.

2. *Objects may not be separated.* As shown in left image of Figure 4.24, the two objects are not separated explicitly. We use intelligence to decipher exactly where one object starts and another ends.

3. *Objects may be hidden.* Note that the middle image of the tiger shown in Figure 4.24 is almost completely hidden. But the little information that is present helps us to identify the tiger.

Figure 4.23: When viewed from the right vantage point, the scene looks like a circle of rocks (left) and reveals its true structure from another view point (right).

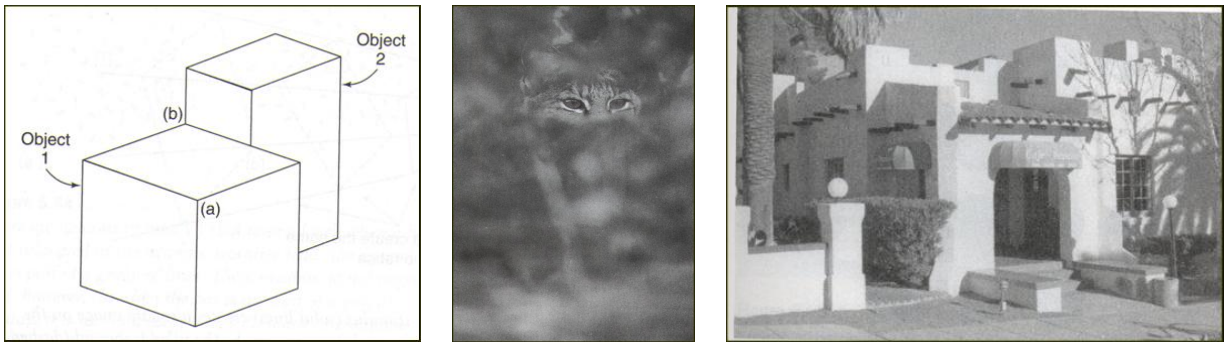

Figure 4.24: Left: The objects are not separated explicitly. Middle: Almost all of the object is hidden or occluded. Right: Some changes in the lightness is due to object and rest are due to illumination. Yet for all of the above we have absolutely no problem figuring out the 3D objects that make the scene.

4. *Lighting changes ambiguously.* Also note that in the right image of Figure 4.24, many of the lighting changes are due to shape and rest are due to illumination. However, they cannot be distinguished easily without intelligence.

All these examples illustrate our dependence on intelligence and experience for successful visual perception. In order to succeed in computer vision, we need to find some way of achieving intelligence and experience. There are work going on the artificial intelligence and neural networks domain the achieve intelligence and experience respectively.

However, here are a few heuristics on how we use intelligence and experience. Just food for some thoughts in those directions.

- *Occlusion Heuristic:* Check Figure 4.25. Note that occlusion helps us in perceptual organization. When the pieces get occluded by the inkblot, they are perceived as continuation of a bigger object in the background which helps in the perception of that object for us. Gestalt's law of good continuation comes to play here too.

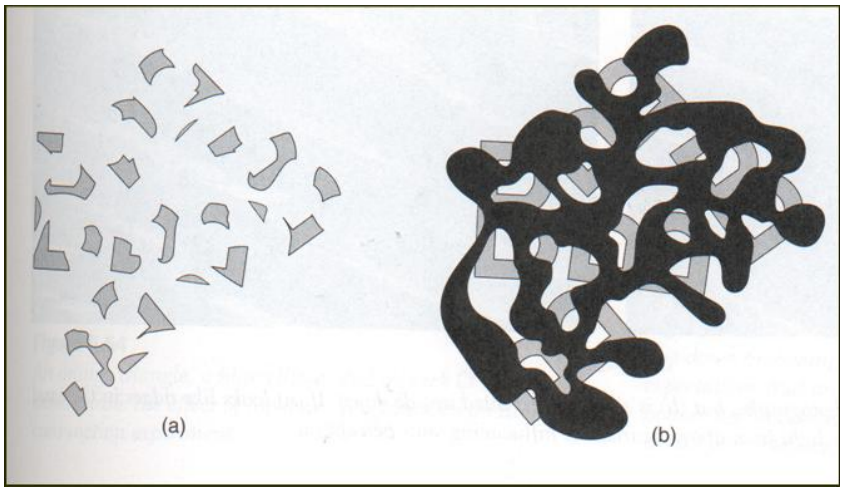

Figure 4.25: The gray areas in (a) are the same in (b). But they become perceptually organized in (b) when the occluding blobs fills in the missing parts.

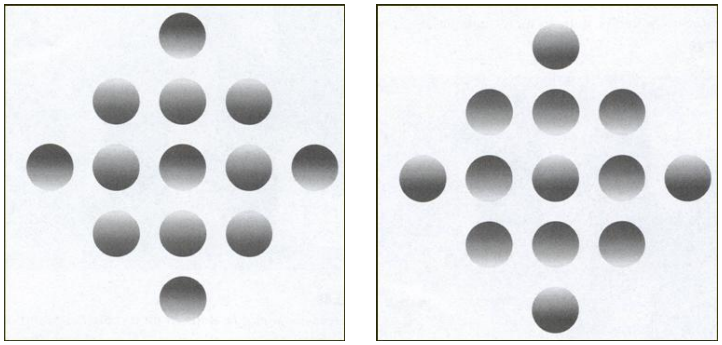

Figure 4.26: The left image is just an upside down version of the right image. Note how the square pattern changes from spheres to indentions in the images.

- *Light from above heuristic:* Your perception of the spheres and indentions in Figure 4.26 is being affected by the light-from-above heuristic. Since we are used to sunlight, our visual scene assume that the light is coming from above. This causes these perceptual effects.