EE4212 Computer Vision

Dr Tan Ping and Dr L.F.Cheong

Course website:
http://courses.nus.edu.sg/course/eleclf/ee4212

Part I – 3D Computer Vision

- A mix of mathematical tools and latest research results in computer vision.
- Aimed both at understanding the theory and applications of computer vision, as well as some knowledge of biological vision.
- Continuous assessment (~40%) and written exam (~60%, allow to bring in A-4 sheet).

Part I – 3D Computer Vision

- Part II of the course focuses on the mathematical / geometrical aspects of 3D Vision.
  - Modeling cameras
  - Mathematical tools
  - Correspondence. Stereo Geometry
  - Optical flow. Shape from motion.

Lecture Schedule

- Introduction to 3D vision and its applications. Marr's paradigm. Shape from X. Depth perception.Various illusions.
- Visual sensors. Geometrical modeling of cameras. Mathematical tools and terminology. Projective geometry.
- Correspondence. Epipolar geometry. Fundamental matrix. Structure from stereo.
- Optical flow. Correspondence. Reconstruction from image streams. Motion parallax methods.

Achieved skills

On completion of this part of the course, you should

- Understand the achievements and the goals of current 3D vision research.
- Understand the jargon in 3D vision literature.
  - Appreciate the difficulties in recovering three-dimensional structure from known view-points.
  - Appreciate the difficulties in finding corresponding features in different view-points.
  - Be able to calculate structure from stereo and motion.
- Appreciate the industrial potential of 3D vision.

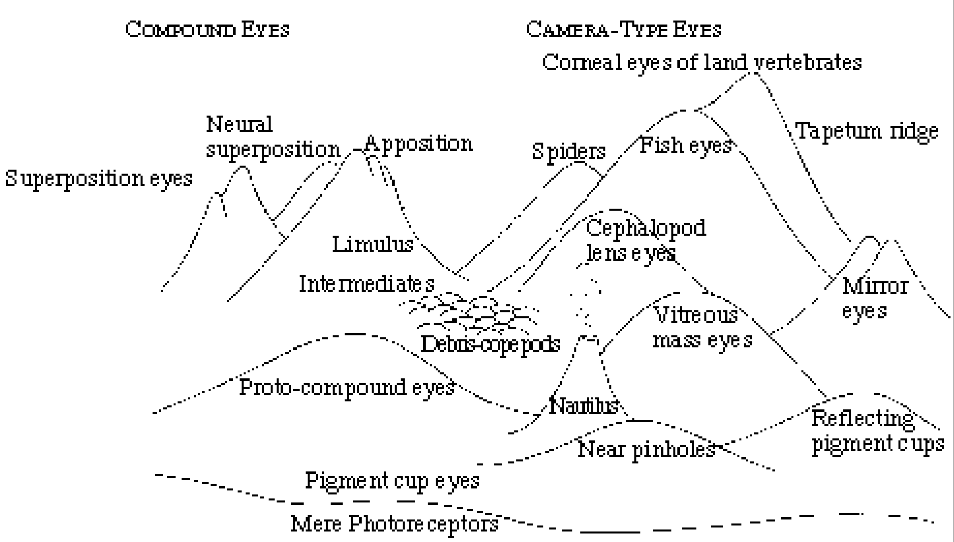
What is 3D Computer Vision?

- How to calculate the 3-D models from their 2-D images
  - 3-D understanding of the scene.
- Most of us take completely for granted our ability to see the world around us.
  - More than half our brain is for processing of visual info.
- How can we determine whether an object is large and distant or small and close? How about their shape
  - Distance and shape perception.
- How can we tell whether we are moving relative to objects in the environment or they are moving relative to us?
  - Motion analysis. Navigation.

Evolutionary View of Vision

- Vision evolved to aid in the survival and successful reproduction of organisms.
  - Desirable objects and situations be sought out and approached.
  - Dangerous objects and situations must be avoided or fled from.
- Motion leads to recognition?
  - Horridge suggests that in the evolution of a visual system, objects are first separated by their motions, and with the evolution of a memory for shapes, form vision progressively evolves.
- Eyes in nature have evolved no fewer than 40 times independently in diverse parts of animal kingdom.
- Eyes "landscape" show 9 basic types of eyes.

Evolutionary View of Vision



Related Disciplines

- Answers to our visual ability come from a variety of different disciplines—biology, psychology, computer science, neuropsychology, linguistics, and cognitive anthropology
- Related disciplines among computer science:
  - Image analysis: (low-level vision)
    - Two-dimensional signal processing, e.g. smoothing, edge detection, deblurring, histogram equalization etc. Extracting 2D information from the retina.
  - Computer Vision: (mid-level vision)
    - Three-dimensional understanding of the scene. Geometry. Structure and motion estimation. Recognition.
  - Artificial Intelligence: (high-level vision)
    - Reasoning on a high level.

Related disciplines among CS

- Photogrammetry:
  - Precise three-dimensional measurements of point configurations using high-resolution images taken with calibrated cameras under controlled circumstances. The speed of computations is not important, since the computations are often made off-line.
- Computer Vision:
  - Less precise estimates of structure and motion using fast, automatic methods that do not require precise calibration. Often speed is important since the vision system often interact with the environment.

Related disciplines among CS

- Computer Graphics:
  - How to generate images of known three-dimensional models.
  - Today advanced models (surfaces, texture, specularity, shading and different lighting conditions) are used for games and special effects in movies.
  - The three-dimensional models are mostly created manually, which is difficult and time-consuming.
- Computer Vision:
  - How to calculate the three-dimensional models from their two-dimensional images.
  - These images could be virtual, but most applications of computer vision use a camera as a sensor for real world objects.

Automated CAD-modeling

- **Idea:** Use video-films, old photographs, digital images.
  - Extract features from image.
  - Find correspondences between features in different images.
  - Calculate relative positions of cameras.
  - Calculate three-dimensional structure of features.
  - Build CAD-model of the scene.
- This is the topic of intense research. Many problems remain to be solved. There is a lot of potential in the applications.
  - Reconstruction of buildings and artifacts from old photographs.
  - Augmented reality. 2d3, NUS, CMU, etc.
  - Virtual reality.
  - Games. Movies. CMU’s virtualized reality.
  - Industrial measurements and inspection.

Introduction to Visual Perception

Perception as a Constructive Act

- **Veridical perception** (from the Latin veridicus meaning to say truthfully): perception that is consistent with the actual state of affairs in the environment.
- **Adaptation and Aftereffects**
  - Visual experience is not a clear window onto reality
  - Visual perception changes over time as it adapts to particular conditions.

Adaptation and Aftereffects

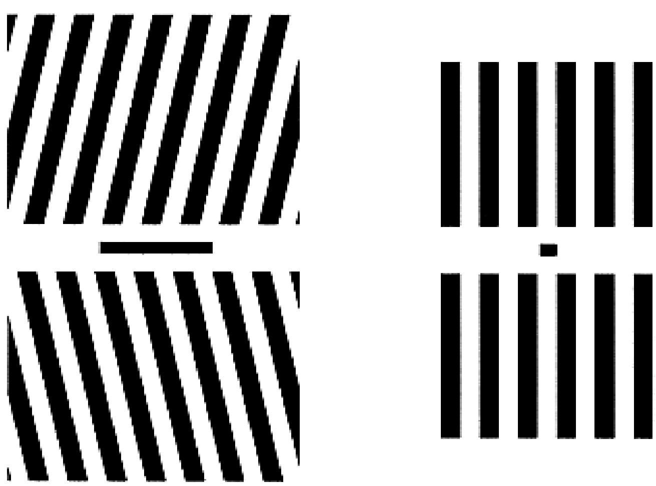
- When you first enter a darkened movie theater on a bright afternoon, you cannot see much except the images on the screen.
- After some time, you can see the whole theater surprisingly well.
- This increase in sensitivity to light is called **dark adaptation**.

Adaptation and Aftereffects

- When someone takes a picture of you with a flash, you first experience a blinding blaze of light.
  - This is a veridical perception.
- It is followed by a prolonged experience of a dark spot where you saw the initial flash.
- This **afterimage** is superimposed on whatever else you look at for the next few minutes.
  - This is not veridical perception because the afterimage stays long after the physical flash is gone.

Reality and Illusion

- Not all aftereffects make you see things that are not there.
  - others cause you to misperceive properties of visible objects. See eg.
  - Another famous example: moon illusion.
- One may think of visual perception as inaccurate and unreliable.
- This is a mistake - Illusions are not terribly obvious in everyday life



An orientation aftereffect.

Run your eyes along the central bar between the gratings on the left for 30-60 seconds. Then look at the square between the two identical gratings on the right. The upper grating should now appear tilted to the left of vertical (counter-clockwise) and the lower grating tilted to the right (clockwise) .

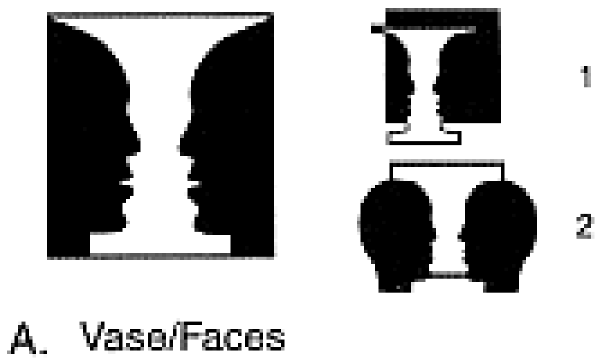
BACK

Reality and Illusion

- To provide us with information about the three-dimensional environment, vision must therefore be an interpretive process that somehow transforms complex, moving, 2-D patterns of light at the back of the eyes into stable perceptions of 3-D objects in 3-D space.
- What we perceive are actually interpretations based on the structure of images rather than direct registrations of physical reality.

Ambiguous Figures

- Thus, what we perceive are actually interpretations based on the structure of images rather than direct registrations of physical reality.
- Further demonstrations of the interpretive nature of vision come from ambiguous figures: Example.
- Two points about these human perception
  1. The 2 interpretations are mutually exclusive - suggests only one model can be fitted to the data at one time.
  2. Once you have seen both interpretations, they are multistable perceptions - suggests 2 models competing with each other



Ambiguous figures
Figure A can be seen either as a white vase against a black background or as a pair of black faces against a white background.

BACK

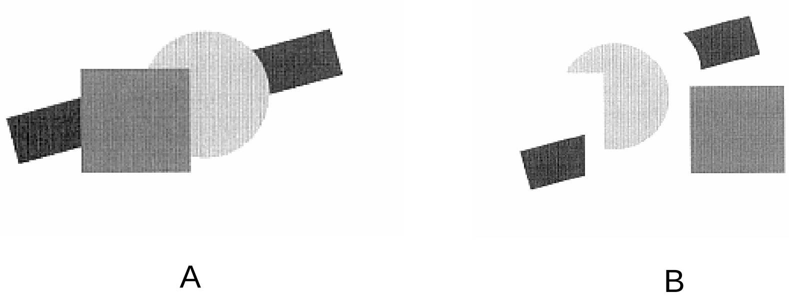
Perception as Modeling the Environment

- Ambiguous figures demonstrate the constructive nature of perception.
- The important idea here is that people's perceptions actually correspond to the models that their visual systems have constructed rather than (or in addition to) the sensory stimulation on which the models are based.
- That is why perceptions can be illusory and ambiguous.

Perception as Modeling the Environment

Visual Completion

- The best evidence that visual perception involves the construction of environmental models comes from the fact that our perceptions include portions of surfaces that we cannot actually see. See example
- This perceptual filling in of parts of objects that are hidden from view is called visual completion.
- It happens automatically and effortlessly.



Visual completion behind partly occluding objects

Figure A is perceived as consisting of a square, a circle, and a rectangle even though the only visible regions are those shown separated in Figure B.

BACK

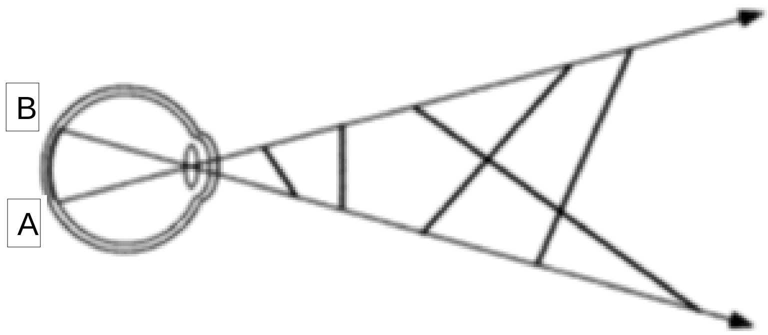
Vision as an "Inverse" Problem

- Light reflected from the 3-D world produces 2-D images at the back of the eye where vision begins.
  - This process of image formation is completely determined by the laws of optics
- But the inverse problem, which is what the early stages of visual perception are all about, is much more difficult:
  - How to get from optical images of scenes back to knowledge of the objects that gave rise to them?

Vision as an "Inverse" Problem

- The inverse mapping from image to environment goes from two dimensions to three dimensions.
- This is not a well-defined function.
- Usually, for every 2-D image on the back of our eyes, there are infinitely many distinct 3-D environments that could have given rise to it.

Illustration



An illustration of inverse projection.

A single line segment on the retina can be the projection of an infinite variety of lines in the environment.

BACK

Vision as an "Inverse" Problem

- Human visual system manages to do create 3-D perception with such remarkable accuracy under most circumstances.
- We assume that 3-D perception results from the visual system making a lot of highly plausible assumptions about the nature of the environment – ecological constraints.
- When assumptions are violated => errors, illusions.

Computer Vision

- Computer vision promoted two important developments that dramatically changed the theoretical branch of vision science:
  - Explicit theories.

    Before computer simulations, theories of visual perception were vague, informal, and incomplete. Computer simulations force the theorist to make everything explicit.
  - Real images.

    Allow theories to be tested on real images of real objects, warts and all.
- Vision turns out to be unbelievably hard to get computers to "see" even the simplest things

Computer Vision

- The major development since the late 70’s is to adopt a mathematical approach to creating working computer vision programs,
- David Marr and his many talented colleagues have effectively articulated it at the Massachusetts Institute of Technology (M.I.T.).
- This research is characterized by mathematical analyses of how the luminance structure in two-dimensional images provides information about the structure of surfaces and objects in three-dimensional space (Marr, 1982).

Three Levels of Information Processing

- David Marr (1982) distinguished three different levels of description involved in understanding complex information processing systems:
  - The computational level
  - The algorithmic level
  - The implementational level
- Important conceptual distinctions among these three levels
  - First is an abstract description of constraints, laws.
  - The next two are algorithms & realization.

Three Levels of Information Processing

- The information processing paradigm is a way of theorizing about the nature of the human mind as a computational process

- The noted philosopher of science Thomas Kuhn (1962) defined a scientific paradigm as a set of working assumptions that a community of scientists shares in conducting research on a given topic.

The Computational Level

- This level of theorizing specifies what computation needs to be performed and on what information it should be based, without specifying how it is accomplished.
- Why segmentation fails?
  - Attempt to segment in accordance to object entity, but goals not appropriate

The Algorithmic Level

- Algorithmic descriptions are more specific than computational descriptions in that they specify how a computation is executed in terms of information processing operations.
- To construct an algorithm for a given computationally defined task, one must:
  1. decide upon a representation for the input and output information.
  2. construct a set of processes that will transform the input representation into the output representation in a well-defined manner.

The Algorithmic Level

- There are, in principle, many different ways in which a given computational-level mapping of input to output might be accomplished.
- Some of the many controversies about the nature of visual representations:
  1. Whether the representation of a given fact is localized in a particular representing element or distributed over many such elements.
  2. Whether a certain fact is represented explicitly or implicitly.
  3. Whether all visual representations can be reduced to a finite set of primitive atoms or constitute an open-ended system.

The Implementational Level

- The implementational level specifies how an algorithm actually is embodied as a physical process within a physical system.
- E.g. thermostat can be implemented as thermister, thermocouple, platinum wire.

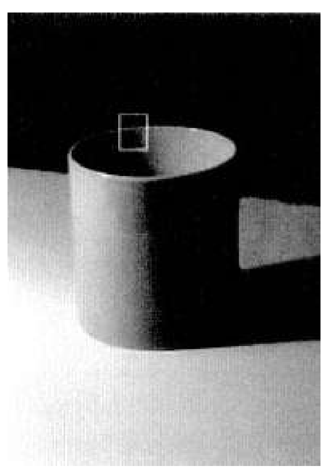
Four Stages of Visual Perception

- Visual perception can be decomposed into four major stages beyond the retinal image itself:

1. The image-based stage of perception.
2. The surface-based stage of perception.
3. The object-based stage of perception.
4. The category-based stage of perception.

The Retinal Image

The input stimulus for vision is a single or a pair of 2-D images projected from the environment to the observer's eyes.



The Image-Based Stage

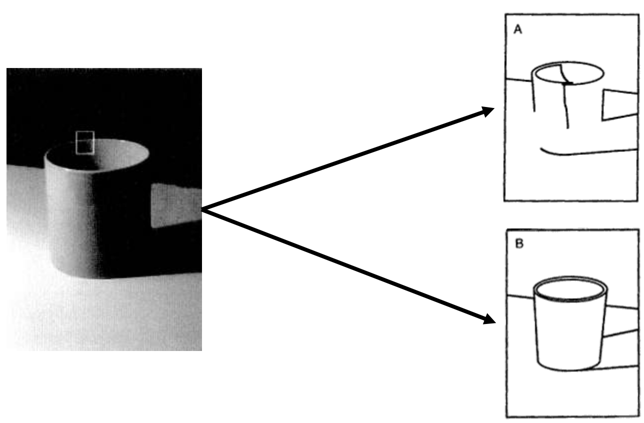
- The image-based stage includes image-processing operations such as:
  1. Detecting local edges and lines,
  2. Linking local edges and lines together more globally,
  3. Matching up corresponding images in the left and right eyes,
  4. Defining two-dimensional regions in the image,
  5. Detecting other image-based features, such as line terminations and "blobs."

The Image-Based Stage

- Marr (1982) called the representations that resulted from such image-based processes primal sketches and suggested that there are two of them.
  – The raw primal sketch: includes just the results of elementary detection processes that locate edges, bars, blobs, and line terminations.
  – The full primal sketch: also includes global grouping and organization among the local image features present in the raw primal sketch.

The Image-Based Stage

- These 2-D features of images characterize their 2-D structure and organization (such as edges and lines).
- But not necessarily information about the 3-D world (such as surface edges or shadow edges).
- See example

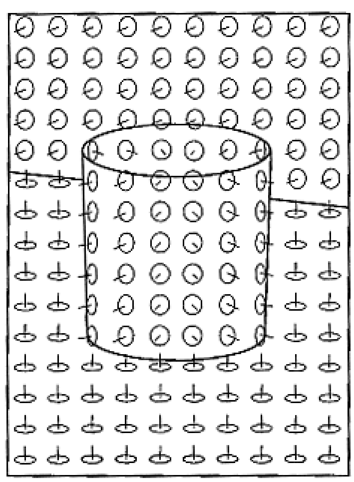


Edges in the cup image.
Local intensity edges can be detected using computerized edge-finding algorithms, as illustrated in part A.
The set of edges thus identified are not the same as those in a clean line drawing of the image (part B),

BACK

The Surface-Based Stage

- The 2nd stage of visual processing: concerned with recovering the intrinsic properties of visible surfaces in the 3-D world.
- E.g. surfaces may be regarded as being composed of many small, locally flat pieces.
- Specify each flat piece completely by just information about the color, slant, tilt, and distance from the viewer.
- See example and flowchart

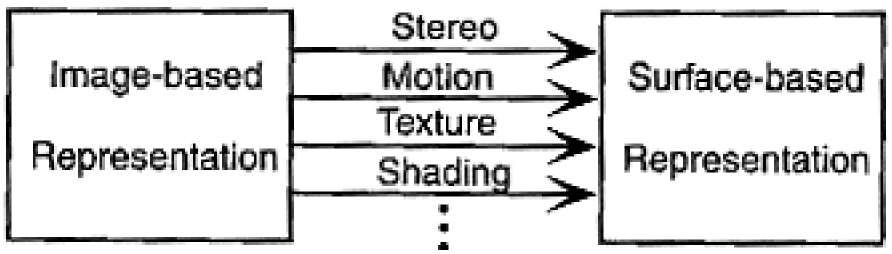


A surface-based representation of the cup scene.

The surfaces visible are represented as a set of local estimates of surface orientation (slant and tilt) and depth with respect to the viewer.

Surface orientation is depicted by a set of imaginary circles on the surface and "needles" pointing perpendicularly out of them at a sampling of image locations.

BACK



A flowchart showing how the surface-based representation might be derived from the image-based representation.

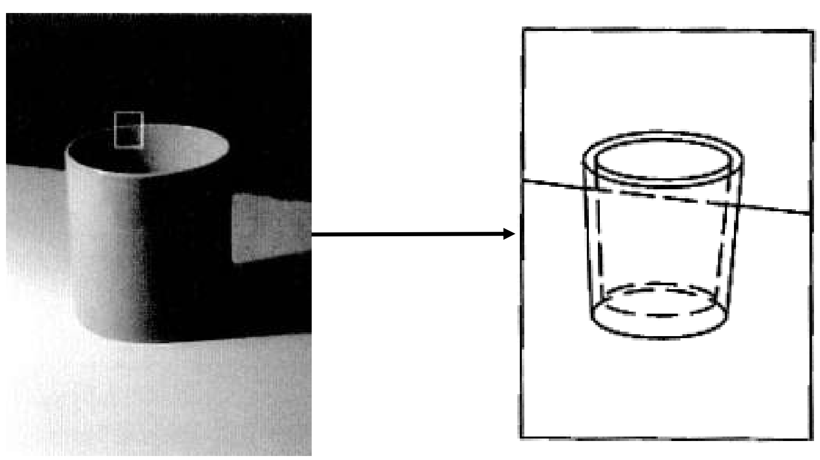
BACK

The Surface-Based Stage

- Marr named his surface-based representation the **2.5-D sketch**
- Because it lies somewhere between the true 2-D structure of image-based representations and the true 3-D structure of object-based representations.

The Object-Based Stage

- Visual perception clearly does not end with a representation of just the surfaces that are visible.
- It is in this object-based stage that the visual representation includes truly 3-D information.
- See example.
- The fact that we complete the hidden surfaces suggests that we have true 3-D representation that includes occluded surfaces in the visual world.



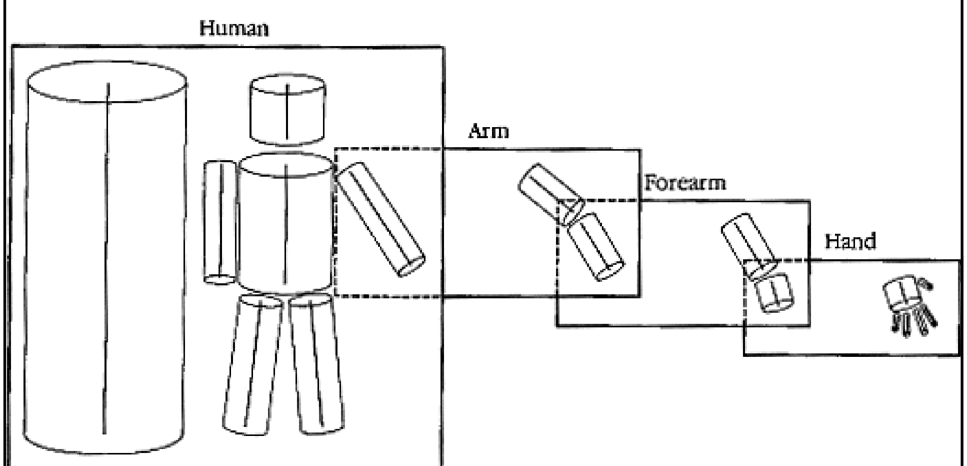
An example of an object-based representation.
The cup shown with its occluded edges represented by dashed lines, indicating how people typically perceive this scene as being composed of 3-D volumes.

BACK

The Object-Based Stage

- One way in which such an object-based representation might be constructed is known as the volumetric approach.
- conceive objects as intrinsically three-dimensional entities, represented as arrangements of some set of primitive 3-D shapes.

Example



Using shape primitives in an object-based representation.

The shape of a person's body as a 3-D volume is roughly represented here as a set of cylinders of the appropriate size, shape, orientation, and connectedness.

Each box shows a perceptual object at a coarse, global level as a single cylinder and at a finer, more local level as a configuration of several cylinders.

BACK

The Category-Based Stage

- The ultimate goal of perception is to provide the perceiving organism with accurate information about the environment to aid in its survival and reproduction.
- This implies that the final stage of perception must be concerned with recovering the functional properties of objects.
- it is widely believed that functional properties are accessed through a process of categorization.

The Category-Based Stage

- The categorization (or pattern recognition) approach to perceiving evolutionarily relevant function proposes that two operations are involved.

The Category-Based Stage

Pattern recognition approach:

1. The visual system classifies an object as being a member of one of a large number of known categories according to its visible properties, such as its shape, size, color, and location.
2. This identification allows access to a large body of stored information about this type of object, including its function and various forms of expectations about its future behavior.

The Category-Based Stage

Functional recognition

- Registering functional properties of objects directly from their visible characteristics without first categorizing them.
- Objects such as chairs and cups have functional properties intimately tied to their visible structure.

The Category-Based Stage

- Objects such as computers and telephones have functions that are so removed from their obvious visual characteristics that they almost certainly need to be categorized first.
- It is possible that people employ both types of processes (direct and indirect) in perceiving function.

The Category-Based Stage

- These four proposed stages of visual processing represent the current best guess about the overall structure of visual perception.
- We have listed them in the particular order in which they must logically be initiated, but that does not necessarily mean that each is completed before the next begins.
- Later processes may feed back to influence earlier ones.
- In the later part of this course, we will only examine mainly stage 2 and stage 3