- Taxonomy of Vision Problems
  - Reconstruction
  - Visual Control
  - Segmentation
  - Recognition
- Overview the Knowledge of Human Vision
  - Molesular and Cell Mechanisms
  - Early Vision Study

1 Summary of the first part of this course

In the first section of our course, we've learned to solve following problems:

1. Geometry and Photometry of Image Formation

   Static (planar, spherical) and Dynamic Perspective (optic flow, disparity)

2. Contours (Singularities)

   Radiance, Irradiance, **BRDF** (Bidirectional Reflectance Distribution Function) and Global Illumination

Although we worked through all the formula and tricks to retrieve certain qualities of the scene, we still do not know how to obtain/estimate them from image data. The answer will lies in the second part of our lectures.

2 Taxonomy of Vision Problems

In general, these problems belong to below 4 categories:

2.1 Reconstruction

- Estimate parameters of external 3D world

We need to do certain measurement in the image and use this to estimate properties of external world, i.e. use **LS** (Least Square) method to estimate the motion of camera.

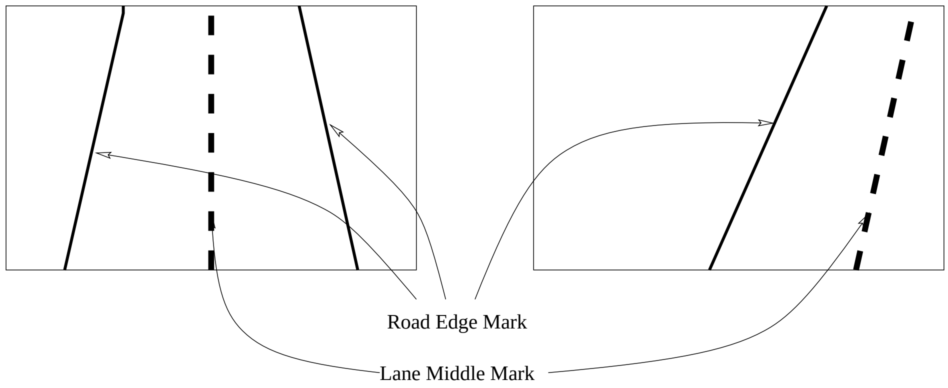

Figure 1: Because we want to maintain our car at the middle of the lane, the left image is desired at all time. Right image shows that the vehicle is off to the left. By tracking the landmark on the road (the dashed line), the vision system sends off-track signal to vehicle control system in order to get back the left image.

2.2 Visual Control

- Visually guided locomotion and manipulation

For example, in Figure 1, people use vision guidance system to stay the vehicle in the middle of the lane and maintain a safe distance from previous car. Visual Control can be done in two ways:

- In the image plane, i.e. keep image projection of lanes centered. See Figure 1.
- In the 3D world.

2.3 Segmentation

- Provide object based image description
- Partition $I(x,y,t)$ into subsets of seperate objects.

In Figure 2, we use texture based segmentation algorithm picking the tiger out of grassland background.

2.4 Recognition

Recognition is often a coupled problem with Segmentation. For instance, in Figure 2 after recognizing the tiger, we express the boundary of it in order to proceed segmentation. In the other way around, we can certainly do edge detection (segmentation) first, match certain enclosed region with our objects database, then recognize it as a tiger.
Here are classifications of recognizing process:

- Classes: face vs. non-face

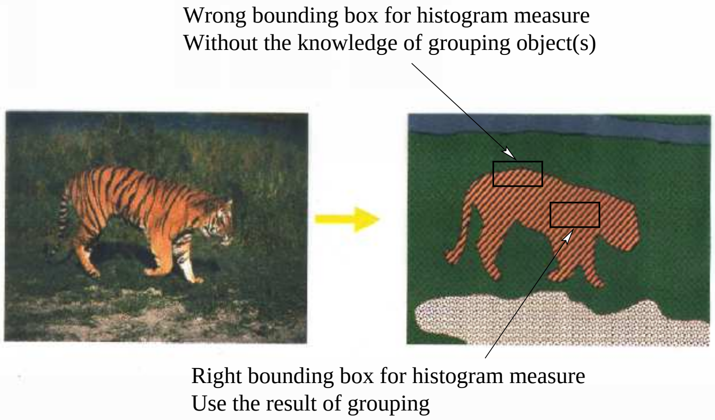

Figure 2: Provide object based image description

- Activities: gesture, expression
- Measurements of image properties, i.e. optic flow, disparity and etc..
- Grouping, i.e. in Figure 2, we need to do grouping object(s) first (recognizing the tiger) then place box inside it to measure the color histogram of its pattern.

Solving any of above problems needs the knowledge of boundaries of images. These boundaries are defined by a number of attributes:

- Brightness/Color
- Texture (texture is often defined as a **Repeated Pattern**, i.e. tiger in above images)
- Motion, i.e. Bird flys in the bush. With single balck and white image, it is hard to tell the bird. But the motion of the bird gives us a cue to pick it out from its "camouflage".
- Stereoscopic depth. Human vision system is born to use the stereoscopic feature. With built-in knowledge of disparity between left and right eyes, we can tell the depth of 0.4 millimeter when the objects are placed in 1 meter ahead of us, referred to the notes of Lecture 9.
- Familiar Configuration by our own knowledge, i.e. table from a sheet of paper, moon from ballon.

As in Figure 3, some of above techniques can be applied to set the boundaries of pillows (**Brightness/Color**), curtain (**Texture**) and etc.. Figure 4 indicates the usage of **Familiar Configuration** to detect the "happy face" among curved, straight line segments. Since having a static single image only, we are not able to take the advantage of notion and stereoscopic depth methods yet.

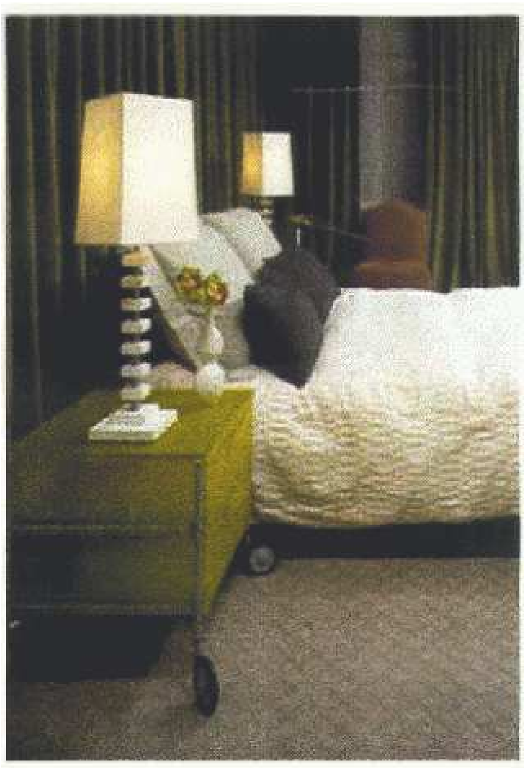

Figure 3: Use Texture and Brightness/Color information to pcik out pillows and curtain, respectively.

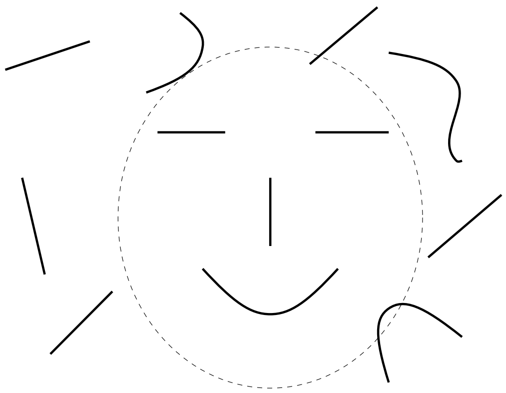

Figure 4: Pciking out "happy face" from scattered surved/straight line segments, using Familiar Configuration.

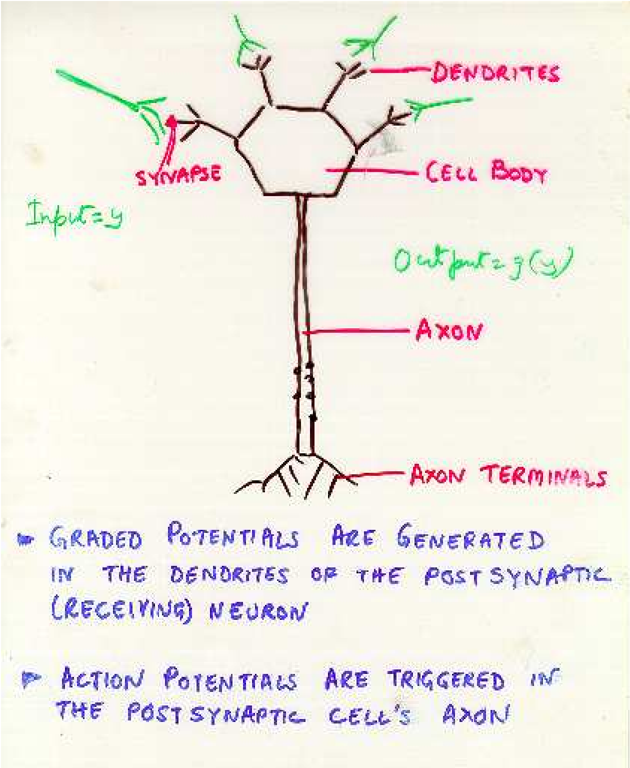

Figure 5: Neuron's structure. Graded potentials are generated in the dendrites or the post synaptic (receiving) neuron. Action potentials are triggered in the post synaptic cell's axon.

3 Overview the Knowledge of Human Vision

- Molecular and Cell Mechanisms

In order to understand the power of our nerve system, we use state-of-art PC as an analogy. If we treat each neuron, the smallest element unit in our nerve system, as a transistor used in computer, the main differences of our nerve system and PC are:

- Human has 10^{10} neurons. Each neuron connects to 10^4 other neurons.
- P-III has 10^{10} transistors. But the connectivity is far less then human.

Let's take a look at the structure of neuron in Figure 5. Transimission in the cell is electrical by the potential from dendrites to axon terminals. Transimission between cells is chemical through synapes. When a neuron is active, it generates potential pulse with amplitude of -55mV to 55mV corresponding resting potential is -70mV in Figure 6. Also, we can view the functionality of neuron as a system box $g(.)$ with $Input = y$ and $Output = g(y)$ in Figure 5.

With enormous number, how are these neurons organized? They are generally divied into 3 systems:

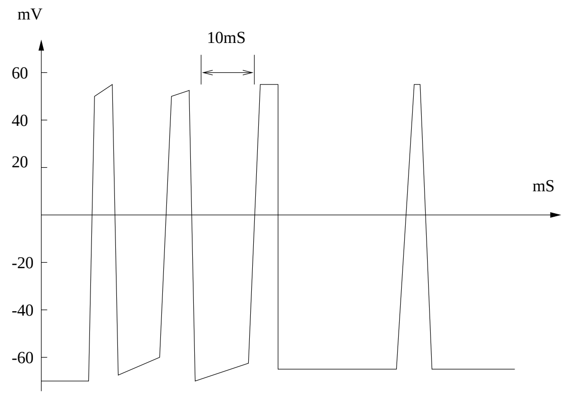

Figure 6: General neuron firing plot. Rest potential is about $-70mV$. Firing potential ranges from $-55mV \sim 55mV$. The smallest firing interval is about $10mS$. This speed is way slower than the response speed of transistors in current CPU. But with complex connectivity among neurons, we still beat computer on complicated tasks.

- **Sensory System**, i.e. hearing, vision and etc.. For human, we use $30 \sim 40\%$ of neurons for vision.
- **Motor System**, i.e. locomotion manupulation, speech and etc..
- **Central System**, i.e. language, planning, learning and etc..

As we discussed above, neuron shoots out signals by positive potential from dendrites to axon. How does it deal with negative signals? There are 2 possible tricks which are easy to think of:

- **Trick 1:** Assume a neuron has a rest firing rate of 50Hz. It adjusts the firing rate to 75Hz in order to send positive signals and reduce the firing frequency to 25Hz to handle negative signals. This is similar to **FM** (Frequency Modulation).
- **Trick 2:** Since the information is transmitted among neurons, it is easy to flip the transmission direction for different signs of signals. For example,
  $A \rightarrow B$ Neuron **A** fires when the signal is positive.
  $A \leftarrow B$ Neuron **B** fires when the signal is negative.

Since **Trick 2** doesn't need constant firing rate at rest, it saves more energy than the first trick and is more likely to be true in the real world.

4 Early Vision Study

Biologist has concluded that different behaviors are controlled by/reflected in different areas of the brain. The vision activity is reflected in the back of our brain shown in below Figure, called **VI**(Visualcortex). Studying the activities of VI has been carried out since the

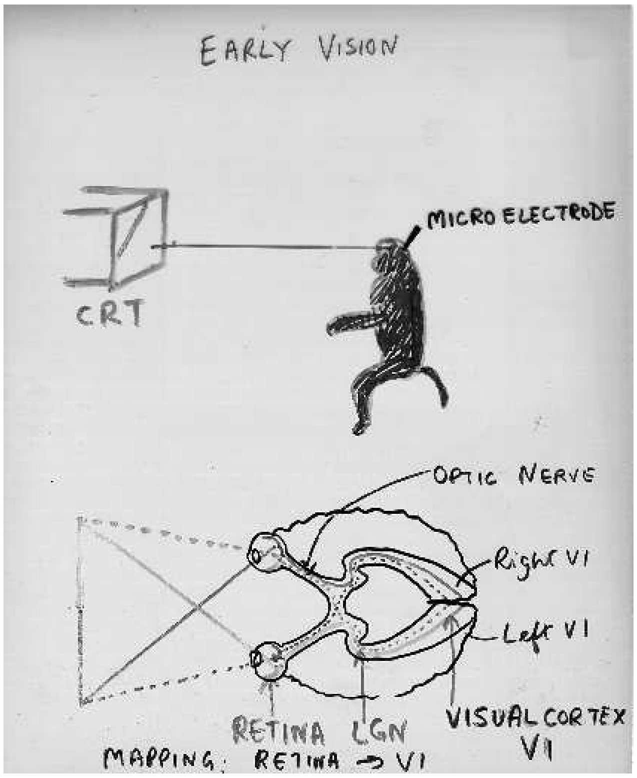

Figure 7: Early Vision study. Monkeys became subjects of our poineers. Monkeys were either trained, rewarded with fruit when look straight, or forced, building cage to fix their heads' direction, to look straight at the CRT. People displayed designed patterns as stimuli on CRT and dissected the monkey's head in order to analyze brain activity. The lower part of this figure shows the vision path inside human/monkey brain. LGN stands for Lateral Genicular Nucleus which acts as an active amplifier and prevents the visual signal generated in retina from dying out through the path. Right VI processes left part of the image. Left VI processes right part of the image.

beginning of this century. In the old days, biologist has no powerful computer. They did experiments on monkeys shown in Figure 7.

By showing below slides, we briefly touch the structure of our retina in Figure 8 and VI structure in Figure 9. After gathering the visual image into VI, our brain carries out two visual information processing streams passing the visual image data into the other parts of our brain for proper responding. This mechanism and streams is discovered by Ungerleider & Mishkin (1982) shown in Figure 10. We can also estimate the computation power of VI by knowing the number of neurons, which can be obtained by measuring the area of iterest. We'll give more details in next lecture.

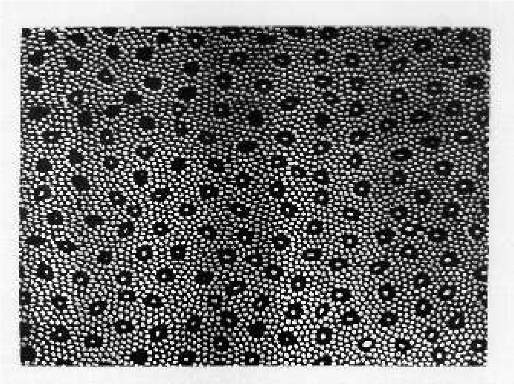

Figure 8: In human retina, we have about 120×10^6 receptors. They are divided into **Rod** which is more sensitive at night and **Cone** that is sensitive to different ranges of wavelength. Cones are responsible for color vision.

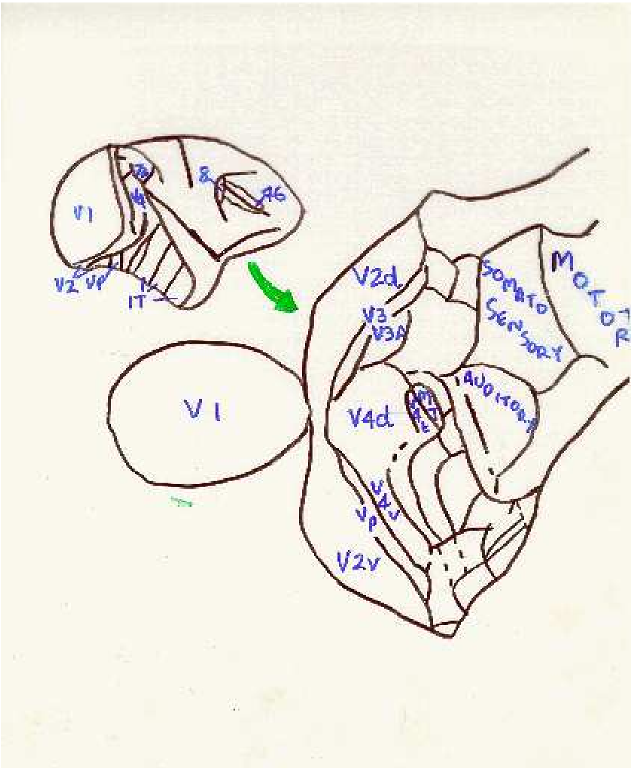

Figure 9: Structure of Visual Cortex (VI). It locates at the back of our brain and divided into multiple regions to sense different visual information.

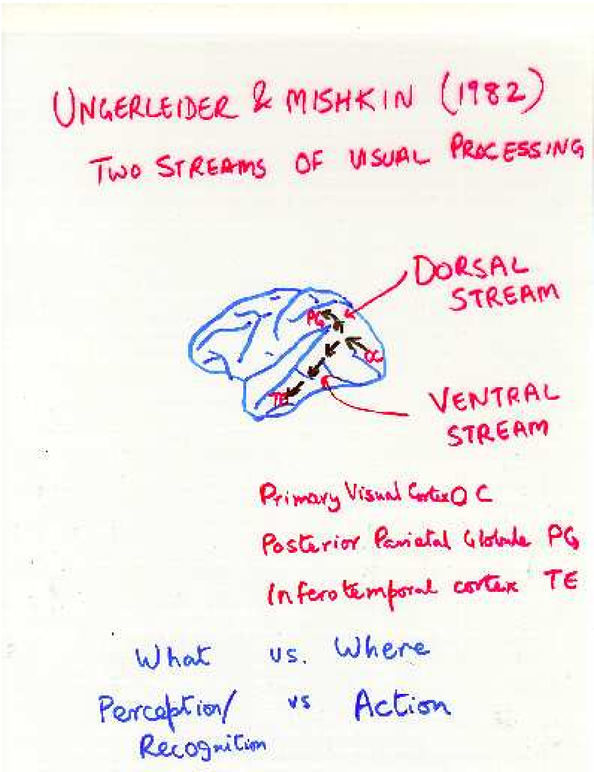

Figure 10: Visual Processing Streams, discovered by Ungerleider & Mishkin (1982)