Vision

- Vision is our most powerful sense. It provides us with an enormous amount of information about our environment and enables us to interact intelligently with the environment, all without direct physical contact. It is therefore not surprising that an enormous amount of effort has occurred to give machines a sense of vision (almost since the beginning of digital computer technology!)
- Vision is also our most complicated sense. While we can reconstruct views with high resolution on photographic paper, the next step of understanding how the brain processes the information from our eyes is still in its infancy.

CCDs

- The most common vision sensors are camera that use CCD (charged couple device) chips.
- A CCD chip is essentially an array of capacitors (around 5-25 microns) that are light sensitive.
- The array is comprised of 20,000 to several million pixels depending on the resolution.
- On exposure, a photon hits a pixel and releases energy which is stored until the image is read off the array.
- Usually give 640x480 (or larger) images at 30 Hz (or FPS)

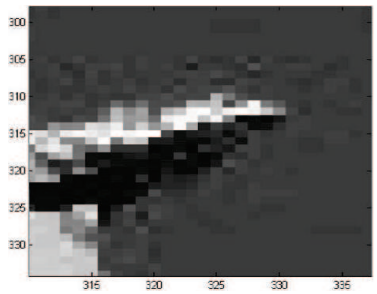
Grayscale images

These are referred to as grayscale or gray level images

- Corresponds to achromatic or monochromatic light
- Light “devoid” of color
- Also results from equal levels of R-G-B in a color image
- Typically 8-bit unsigned chars with a dynamic range of [0,255]

$$0 \leq I(x, y) \leq 255$$

Image Representation



61	29	29	57	199	192
222	200	197	135	167	222
203	203	203	137	137	165
208	208	201	124	142	111
208	203	200	190	127	92
204	201	200	218	173	139

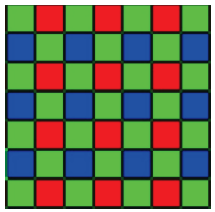
It's just a bunch of NUMBERS!

RGB Colorspace

- Motivated by human visual system
  - 3 color receptor cells (cones) in the retina with different spectral response curves
- Used in color monitors and most video cameras
- Typically 3 8-bit unsigned chars for each color (R: [0,255], G:[0,255], B:[0,255])

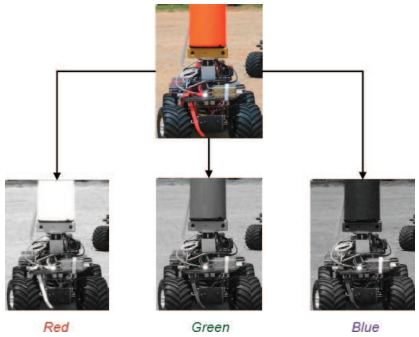
Color CCDs

- In normal, inexpensive color cameras, CCD array are divided into 2x2 regions of green, red, and blue receptors.
  - Human vision is more responsive to green than red or blue
- Half the pixels in the CCD are allocated to green, a quarter to red and a quarter to blue
- Color is generated for the whole CCD by interpolating neighbor values
- The image we get has already undergone a “lossy compression”



Color CCDs

- In more expensive cameras, there are 3 CCD chips, one that measures wavelengths of blue light, one for read, and one for green. green color
  - Three images are taken and combined into a single color image

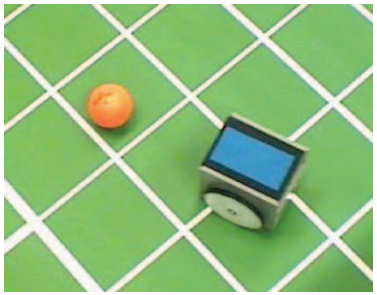


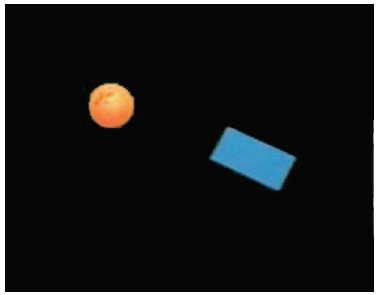
Vision-based Object Detection

- Color Segmentation
- Edge Detection
- Line/Plane Fitting
- Road Detection
- Object Detection
  - Specially trained detectors for specific objects
  - Machine Learning Techniques
- Segmented color blobs or objects can be tracked across multiple frames of video.

Color Segmentation

- Computationally inexpensive (relative to other features)
- “Unnatural” colors are easy to track: e.g., bright pink
- Combines easily with other features for robust tracking

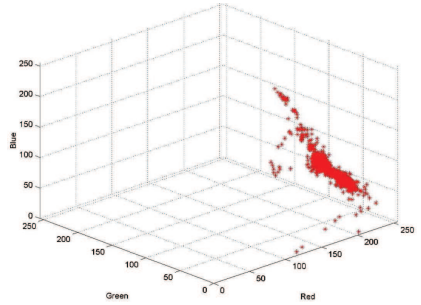




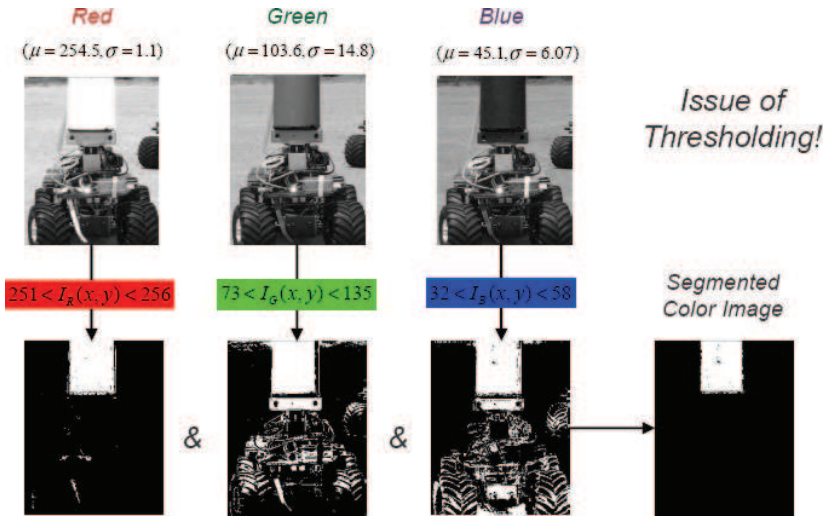
Color Segmentation

How do we segment a single color?

- We need to model is mathematically a priori
- In other words, the robot needs models of colors it is looking for in its memory

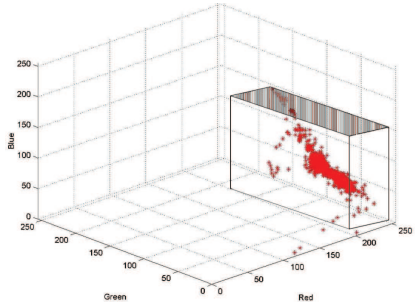


Simple RGB Color Segmentation

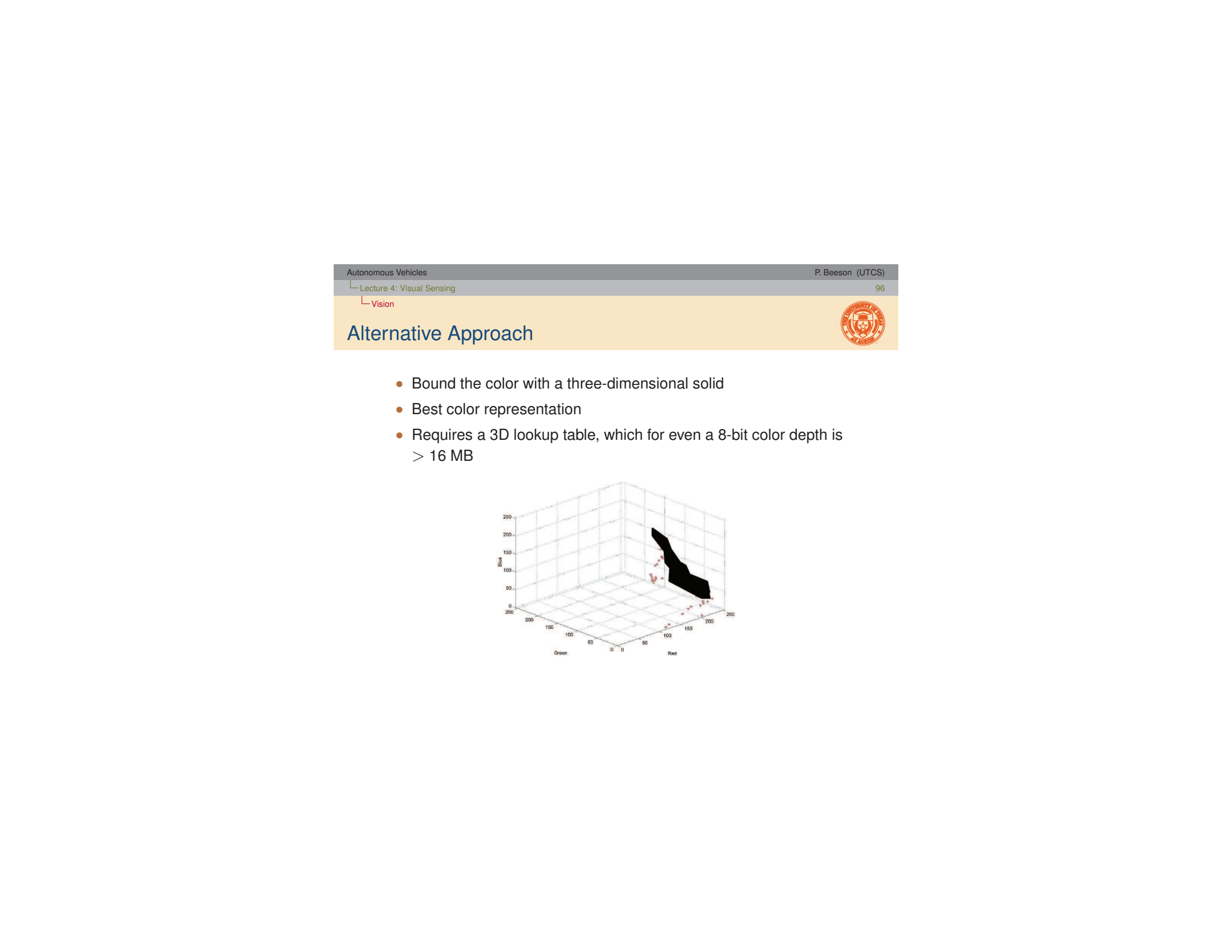


Segmentation Issues

- The approach surrounds the color with a box
- This captures the color, but also many other colors that are not of interest
- Remember, each POINT represents a unique color

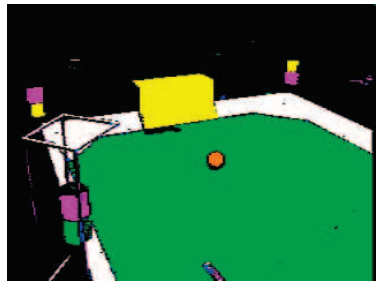


Alternative Approach



- Bound the color with a three-dimensional solid
- Best color representation
- Requires a 3D lookup table, which for even a 8-bit color depth is $>$ 16 MB

Good Results in Static Illumination

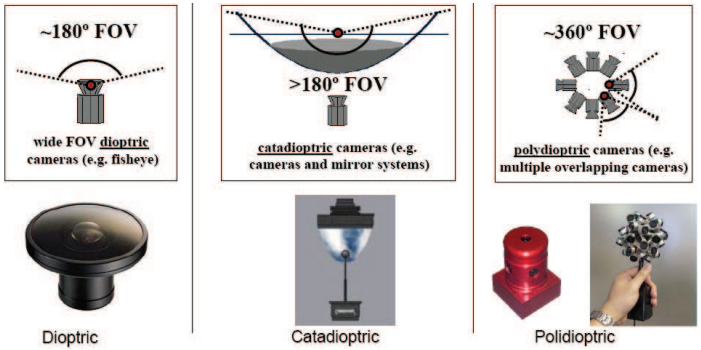


Standard Vision Issues

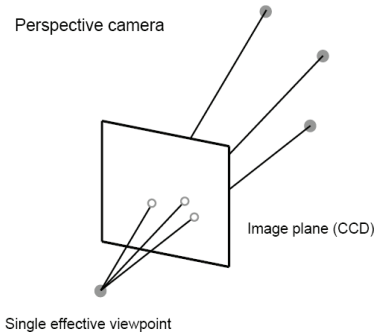
- Benefits:
  - No explicit maximum range
  - Passive sensor
  - Relatively inexpensive
- Drawbacks:
  - Noisy across multiple frames
  - Detects blue light more poorly than red and green
  - Sensitive to illumination changes / dynamic range
    - Color constancy
    - Dark illumination yields little information
    - Too bright of illumination saturates pixels: causing white images or blooming (bleeding of energy into neighboring pixels).
  - 3D information is lost in 2D projection

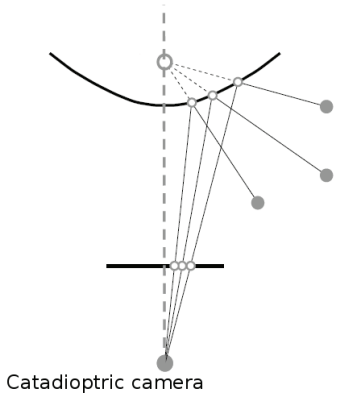
Omnidirectional Cameras

- Omnidirectional Cameras can come in 3 Types



Omnidirectional Cameras



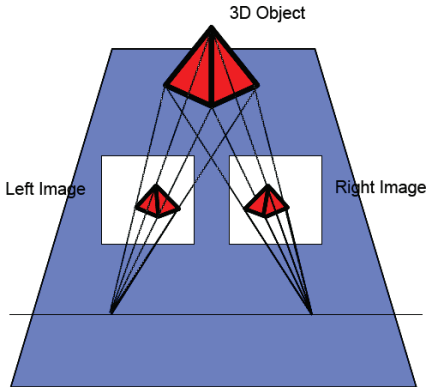


Omnidirectional Cameras

Omnidirectional Cameras

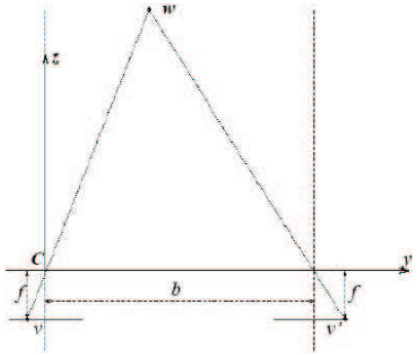
Stereo Vision

Reclaim some 3D info from two images taken at different locations



Stereo Vision

The simplified case is an ideal case. It assumes that both cameras are identical and are aligned on a horizontal axis.



- $\frac{-f}{z} = \frac{-v}{y}$
- $\frac{-f}{z} = \frac{v'}{b-y}$
- b = baseline, distance between the optical centers of the two cameras
- f = focal length
- $v - v'$ = disparity

Stereo Vision

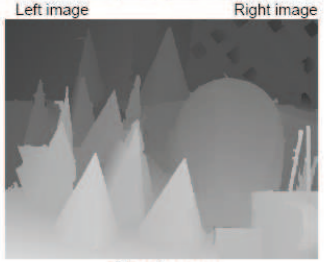
- $z = \frac{bf}{v - v'}$
- Distance is inversely proportional to disparity $(v - v')$
  - closer objects can be measured more accurately
  - Disparity is proportional to b.
    - For a given disparity error, the accuracy of the depth estimate increases with increasing baseline b.
    - However, as b is increased, some objects may appear in one camera, but not in the other.
- Two identical cameras do not exist in nature!
- Aligning both cameras on a horizontal axis is very hard.

Correspondence Problem

- Correspondence Problem: Finding two matching points in the two images which are projection of the same 3D real point
- Hard problem
  - Point (or nearby neighborhood) must be distinctive: no flat untextured walls
  - Sometimes there are multiple matches
  - Looking for m points (or small regions) in an N pixel image is exponential, $O(m^N)$.

Common Solution

Left image Right image



Disparity map

- Lots of calibration needed to get image coordinates to real world coordinates
- Typically distance info is only for objects within a few meters of the cameras.