Lecture 5-1

Complex Sensors:

Cameras, Visual Sensing

The Robotics Primer (Ch. 9)

Course Announcements

- Bring your laptop and robot everyday
- DO NOT unplug the network cables from the desktop computers or the walls
- Tuesday's Quiz is on **Visual Sensing**
- Lab 4 Demo due Thursday, 4/09/09
- Lab 4 Memo and code due by midnight on Friday, 4/10/09
- Lab 4 modified to include Line follower with PID Control
- Upload memo and code to Angel
- Take your robot to the parts room at the end of class on Thursday
- Spring Break, 4/11/09 – 4/19/09

Quote of the Week

Asimov's Robot Laws

1. *A robot may not injure a human being, or, through inaction, allow a human being to come to harm*
2. *A robot should obey a human being, unless this contradicts the first law*
3. *A robot should not harm another robot and protect its own existence unless the contradicts the first or second law.*

From *Handbook of Robotics*, 56th Edition, 2058 A.D., as quoted in I, Robot.

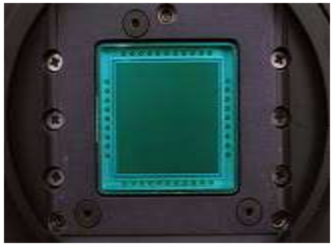
Visual Sensing and Image Processing

Machine Vision

- The research field that deals with vision in machines, including robotics is called *machine vision*.
- The machine vision questions are:
  - What is that?
  - Who is that?
  - Where is that?
- The robotics questions are more action related
  - Should I keep going, turn or stop?
  - Should I grab or leg go?
  - Where do I turn?

Vision-based Sensors: Hardware

- CCD (charge-coupled devices)


2048 x 2048 CCD array


Orangemicro iBOT Firewire


Sony DFW-X700


Cannon IXUS 300

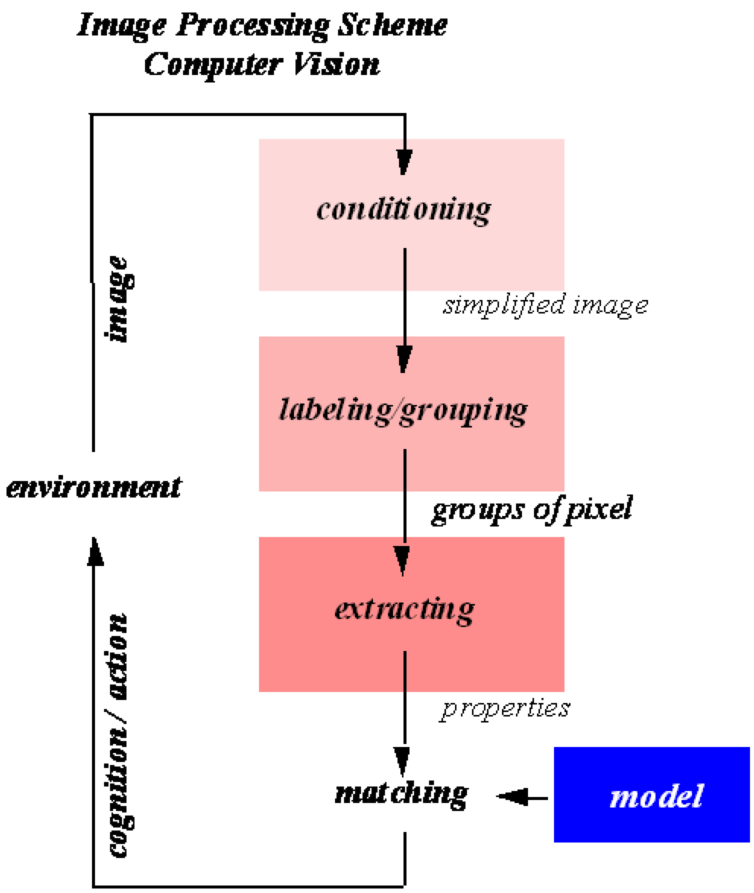
- CMOS (*Complementary Metal Oxide Semiconductor technology*)

Image Processing

- Edge detection is the first step in image processing which is determined by identifying significant changes in brightness
- Smoothing involves using convolution to eliminate isolated peaks in intensity caused by noise
- Segmentation is the process of dividing or organizing an image into parts corresponding to continuous objects

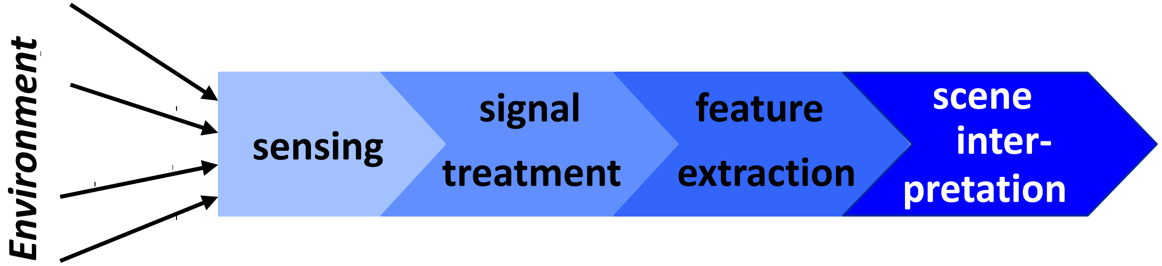
Visual Appearance: Image preprocessing

- Conditioning
  - Suppresses noise
  - Implemented with
    - gray-scale modification (e.g. thresholding)
    - (low pass) filtering
- Labeling
  - Determination of the spatial arrangement of the events, i.e. searching for a structure
- Grouping
  - Identification of the events by collecting together pixels participating in the same kind of event
- Extracting
  - Compute a list of properties for each group
- Matching



Feature Extraction: Scene Interpretation

- A mobile robot must be able to determine its relationship to the environment by sensing and interpreting the measured signals.
  - A wide variety of sensing technologies are available
  - However, the main difficulty lies in interpreting these data, that is, in deciding what the sensor signals tell us about the environment.
  - To extract information from one or more sensor readings to generate a higher level ***percept*** to inform the robot's environment model and action is ***feature extraction***



Feature Extraction: Features

- Features are distinctive elements or geometric primitives of the environment.
- Good features are always perceivable and easily detectable form the environment
- They usually can be extracted from measurements and mathematically described.
  - *low-level features* include *geometric primitives* like lines, circles
  - *high-level features* include edges, doors, tables or trash cans.

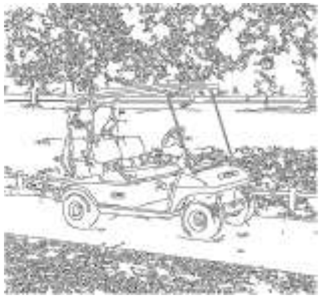
In mobile robotics, features help for localization and map building.

Feature Extraction:
Filtering and Edge Detection

- The single most popular spatially localized feature is *edge detection*
- Edges
  - Locations where the brightness undergoes a sharp change,
  - Differentiate one or two times the image
  - Look for places where the magnitude of the derivative is large.
  - If there is noise then filtering/smoothing is required before edge detection
- Gaussian Smoothing
  - Removes high-frequency noise
  - Convolution of intensity image I with G

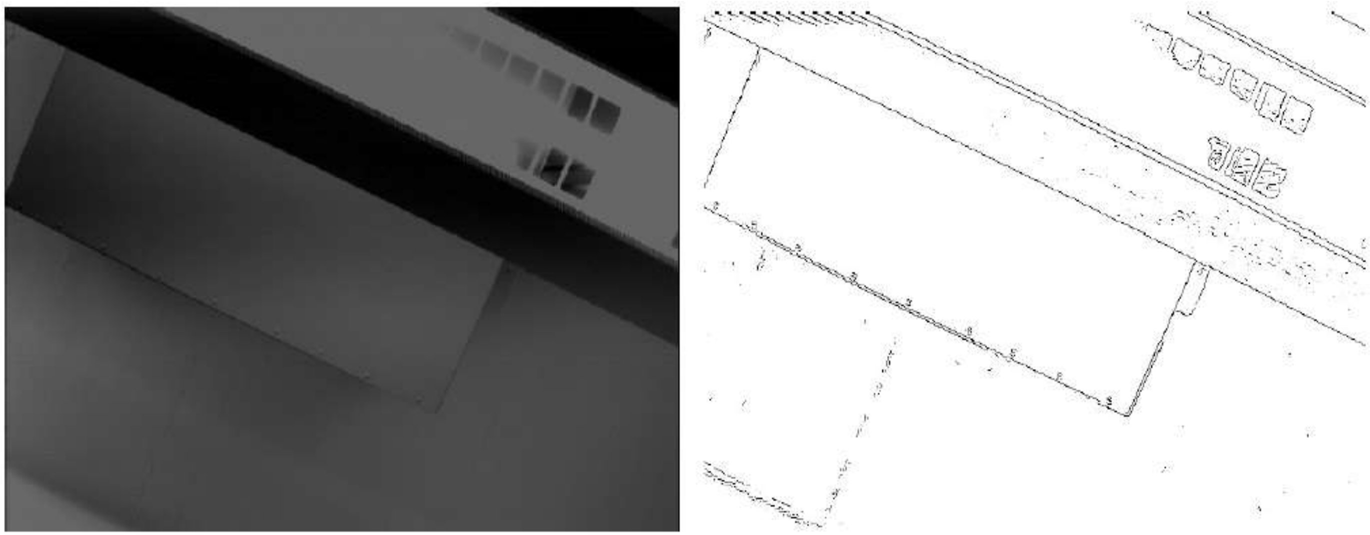
Edge Detection

- Edge = a curve in the image across which there is a change in brightness
- Finding edges
  - Differentiate the image and look for areas where the magnitude of the derivative is large
- Difficulties
  - Not only edges produce changes in brightness: shadows, noise
- Smoothing
  - Filter the image using convolution
  - Use filters of various orientations
- Segmentation: get objects out of the lines



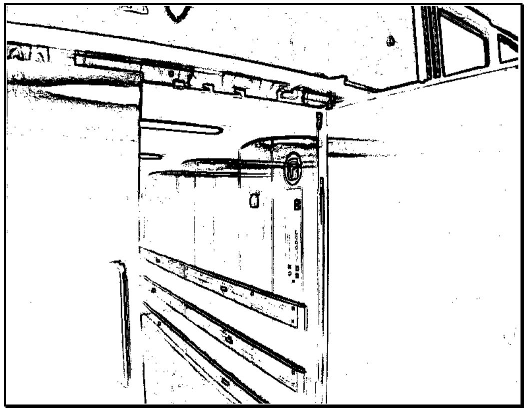
Feature Extraction: Edge Detection

- Ultimate goal of edge detection
  - an idealized line drawing.
- Edge contours in the image correspond to important scene contours.

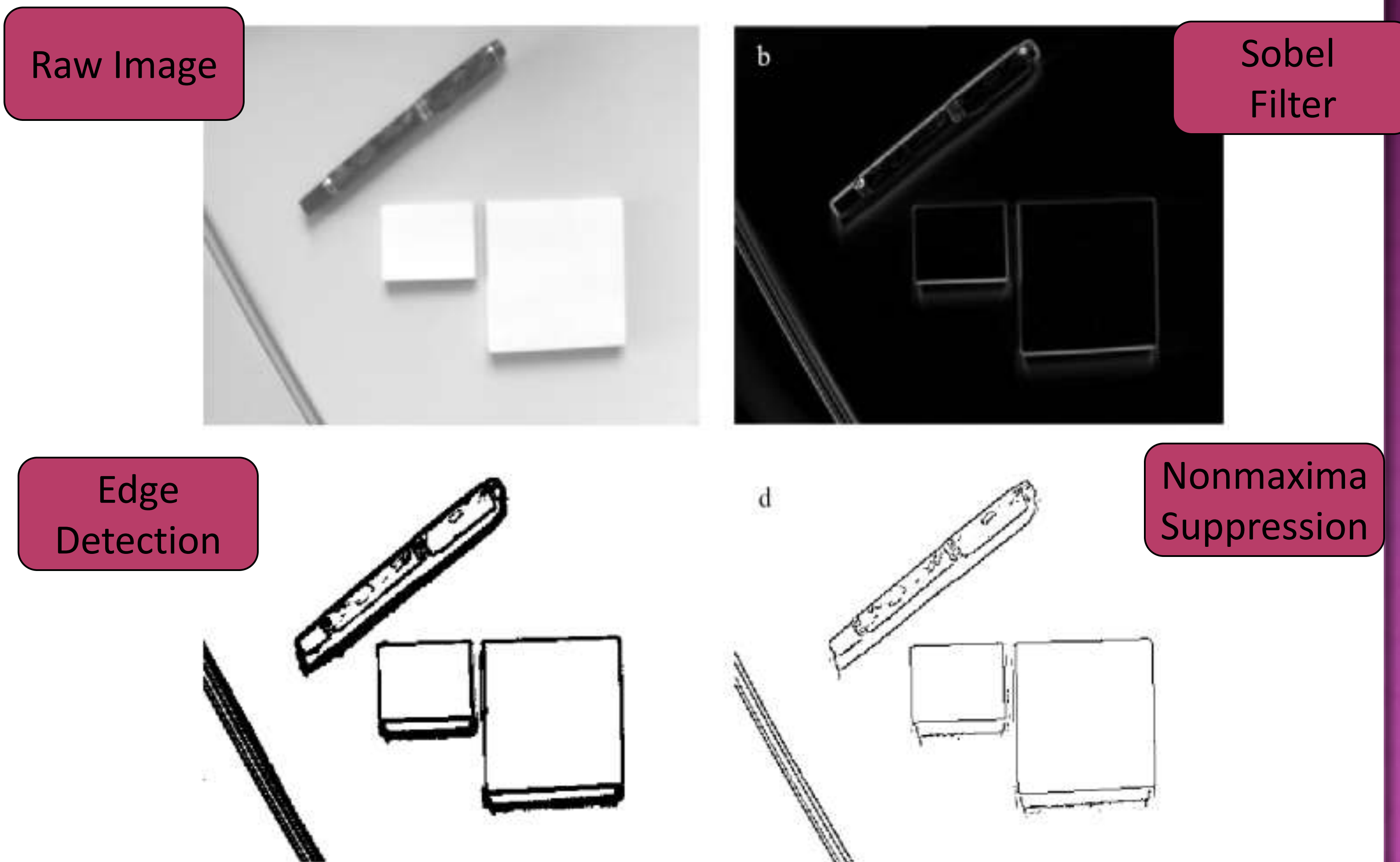


Feature Extraction:
Canny Edge Detector and Nonmaxima Suppression

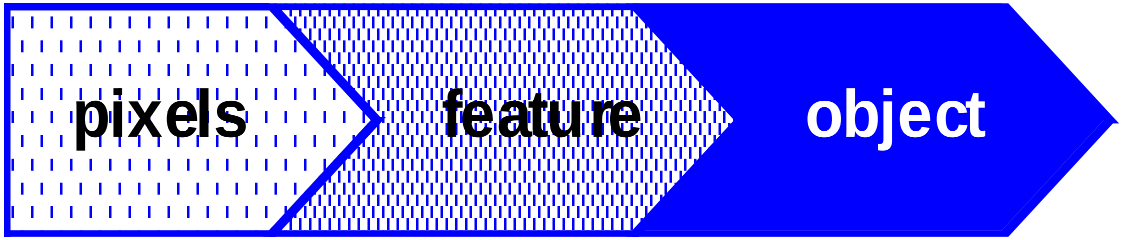
- Canny edge detection is usually a black and white image where the pixels with gradient magnitude above a predefined threshold are black and all the others are white
- *Nonmaxima suppression* sets all pixels to zero that do not represent the local maxima
- *Nonmaxima suppression* generates contours described with only one pixel thinness



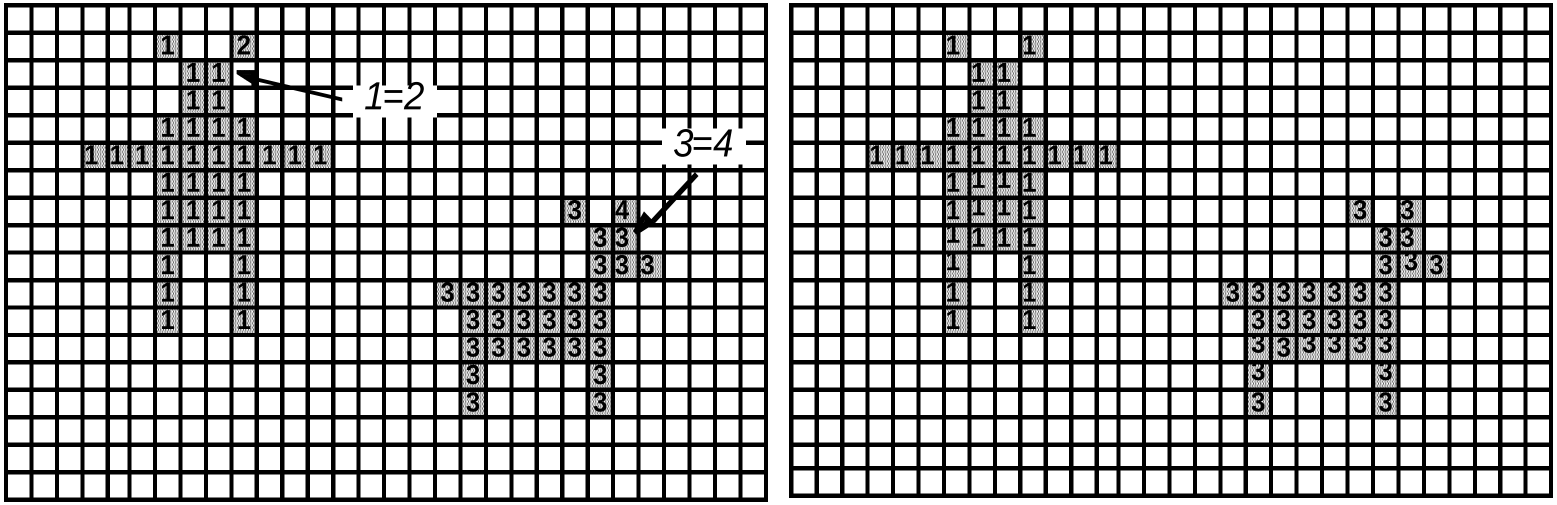
Feature Extraction Example



Grouping, Clustering: Assigning Features to Features



Connected Component Labeling



Feature Extraction:
Floor Plane Extraction

- Vision based identification of a traversable path
- The processing steps
  - As pre-processing, smooth the image using a Gaussian smoothing operator
  - Initialize a histogram array H with *n* intensity values
  - For every pixel (x,y) in the image increment the histogram
  - Mark intensities outside a certain range as white

Model-Based Vision

- When a robot has stored drawings these are called models
- When these models are compared with a camera image with the model this is called model-based vision (i.e. face recognition).
- This is computational intensive and takes plenty of memory

Vision Ranging Sensors

Motion Vision

- Motion vision is a subset of machine vision that uses motion to facilitate visual processing
- A static system can detect moving objects
  - By subtracting two consecutive images from each other which represents the movement between frames
- A moving system can detect static objects
  - By identifying continuous objects which move as one at consecutive time steps as long as the exact movement of the camera is known
  - Robots are typically moving themselves so it is necessary to consider the movement of the robot

Stereo Vision

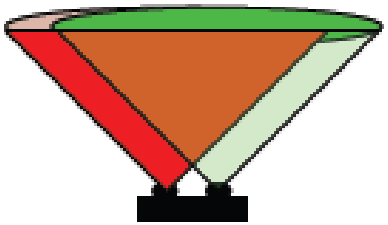
- In nature, creatures have two eyes and using 2 cameras gives a robot binocular vision
- Binocular stereopsis uses combined points of view from 2 cameras to reconstruction three-dimensional objects and to perceive depth.
- Stereo vision is like motion vision except the robot does not have to move to get 2 images
- You must know how the 2 cameras are positioned relative to each other like you must know how the robot moves in motion vision

Vision ranging sensors

- In mobile robotics, it is natural to attempt to implement ranging using vision
- Vision collapses the 3D world into a 2D image
- To recover depth information look at several images of a scene
  - The images must be different
  - They should provide differ viewpoints yielding *stereo or motion algorithms*
  - Alternately, do not change the viewpoint but change the camera geometry (i.e. focus or lens iris) yielding *depth from focus algorithms*

Stereo Ranging Systems

- Similar to laser, robots with stereo cameras can obtain 3D range maps of the environment
- Usually implemented with 2 cameras or one used from multiple locations
- Resolution
  - Camera covers roughly a 45° cone

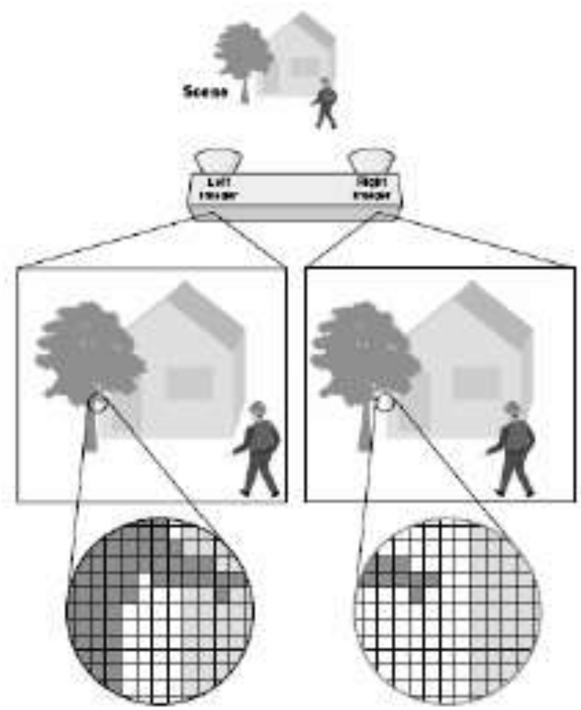


Stereo Ranging Systems: Stereo Vision

- Image depth is inversely proportional to disparity
  - Stereo is most accurate for close objects
- *Disparity* is an integer since it is a difference in x values of pixels
- Accuracy of depth can be increased by increasing baseline distance between cameras
  - However this reduces the overlap of the camera and scene width
  - It is more difficult to match pairs of points since the left and right images have less in common due to larger difference in viewing angle
  - some objects may appear in one camera, but not in the other
- A point visible from both cameras produces *a conjugate pair*
  - Conjugate pairs lie on *epipolar line*

Stereo Ranging Systems: Disparity and Correspondence

- Objects in left camera appear horizontally shifted from objects seen in right camera
- The size of the shift is the *disparity*
- The ideas is to find a *correspondence* (or match) between points in one image with points in other image

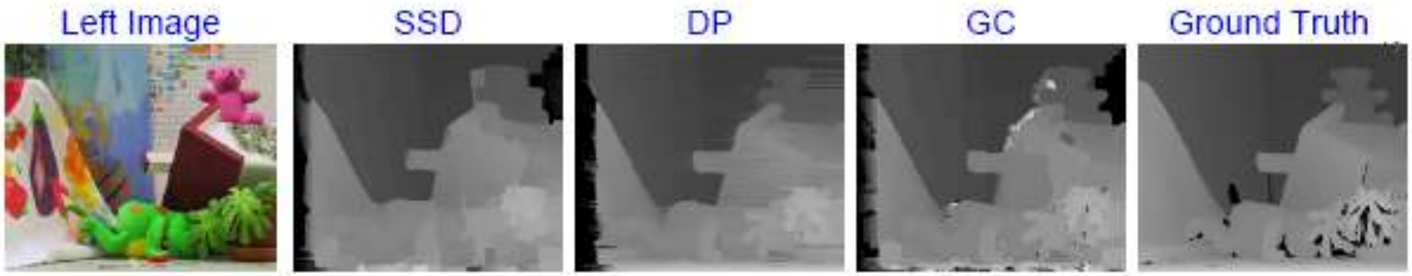


Stereo Ranging Systems: Correspondence

- Desired characteristics
  - Corresponding image regions are similar
  - Each point matches a single point in the other image (unlikely)
- Two main matching methods
  - *Feature-based*
    - Start from image structure (e.g. edges)
  - *Correlation-based*
    - Start from grey levels

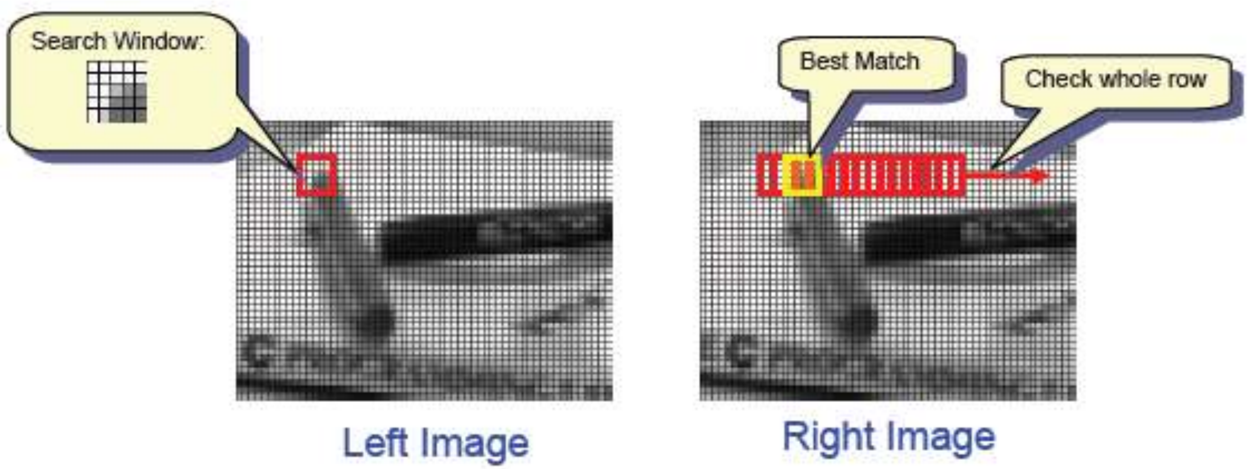
Stereo Ranging Systems: Correlation

- There are several methods:
  - Sum of Squared Difference (SSD)
  - Dynamic Programming (DP)
  - Graph Cut (GC)
  - Ground Truth
  - Belief Propagation (BP)
  - Markov Random Fields (MRF)



Stereo Vision: SSD Correlation

- Take a small area of data in left image and compare it with similar-size area in the right image along the same *epipolar line* (i.e. same height in the image if the cameras are horizontally level)

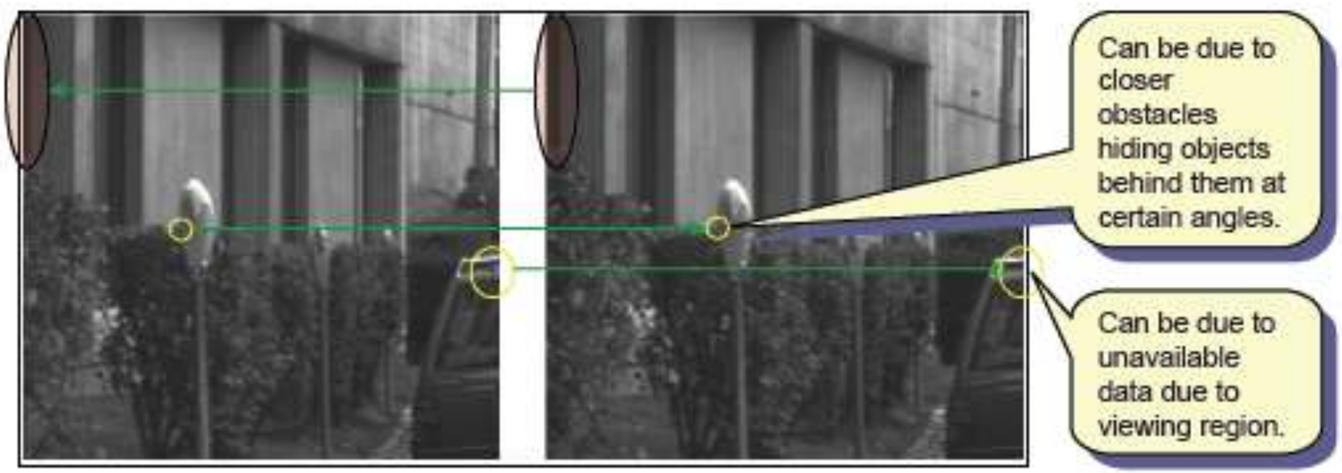


Stereo Vision: Correlation

- To improve matching
  - Apply image filters before and after processing
  - Identify corners and edges to help fill in areas with no data available
  - Use sensor fusion (i.e. data from other sensors) to fill in missing gaps
  - Project structure light onto objects to improve matches

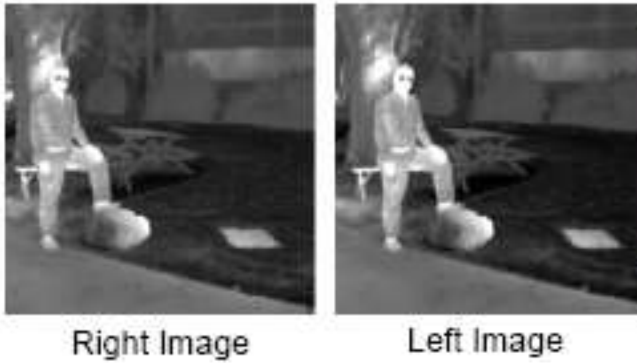
Stereo Ranging Systems: Occlusion

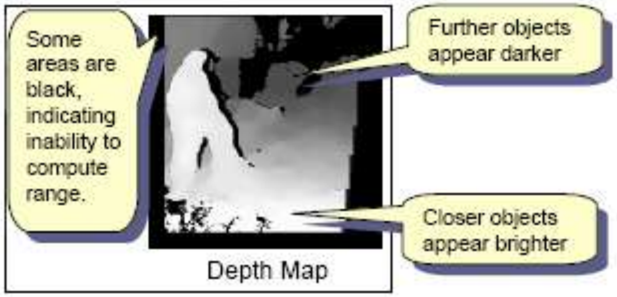
- It is difficult to find corresponding pixels in 2 images
- It is better to find the most likely match
- In some cases, the pixel in one image may not be visible in the other (*occlusion*)



Stereo Ranging Systems: Depth Maps

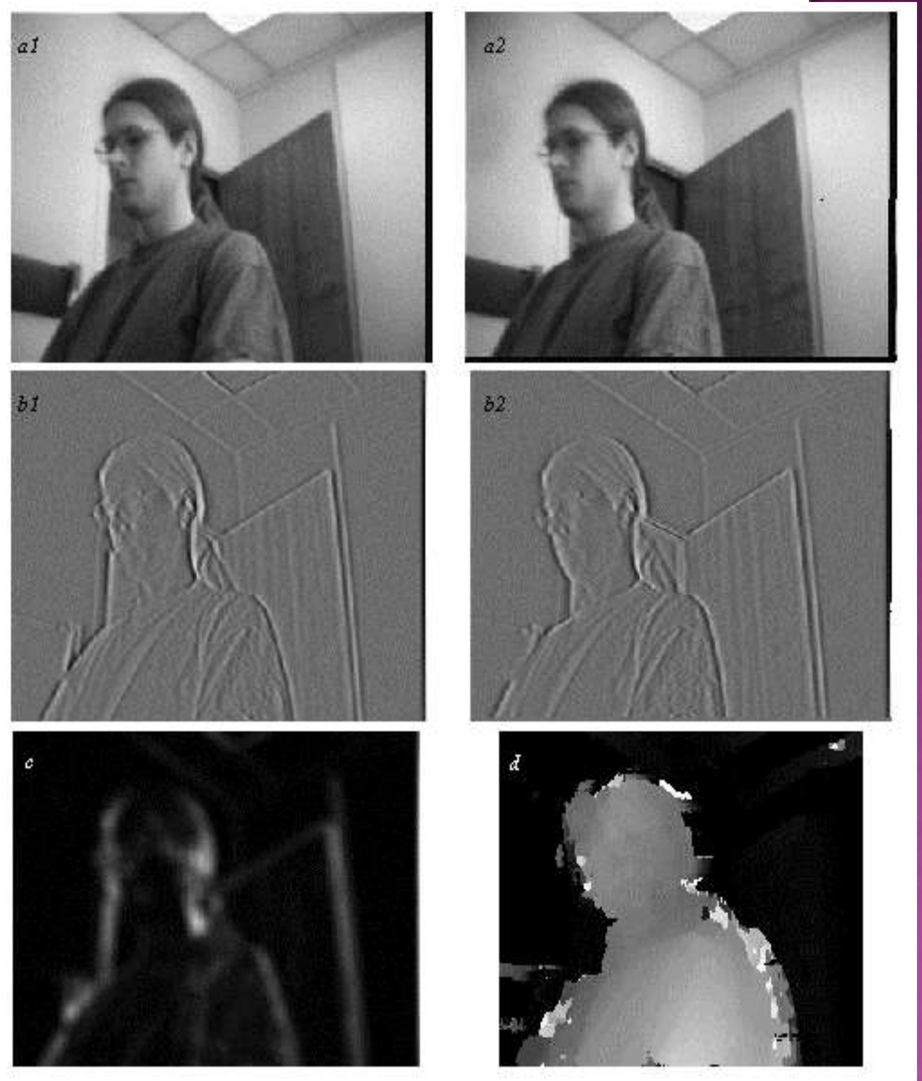
- Calculate the depth or distance of features in an image relative to the sensors (construct a *depth map*)
  - Use images from dual cameras aimed at the same object
  - Locate the same ‘feature’ in both images
  - Use geometric relationships between the 2 cameras and the location of the feature in each image
  - The depth of each feature can be triangulated and a depth map constructed



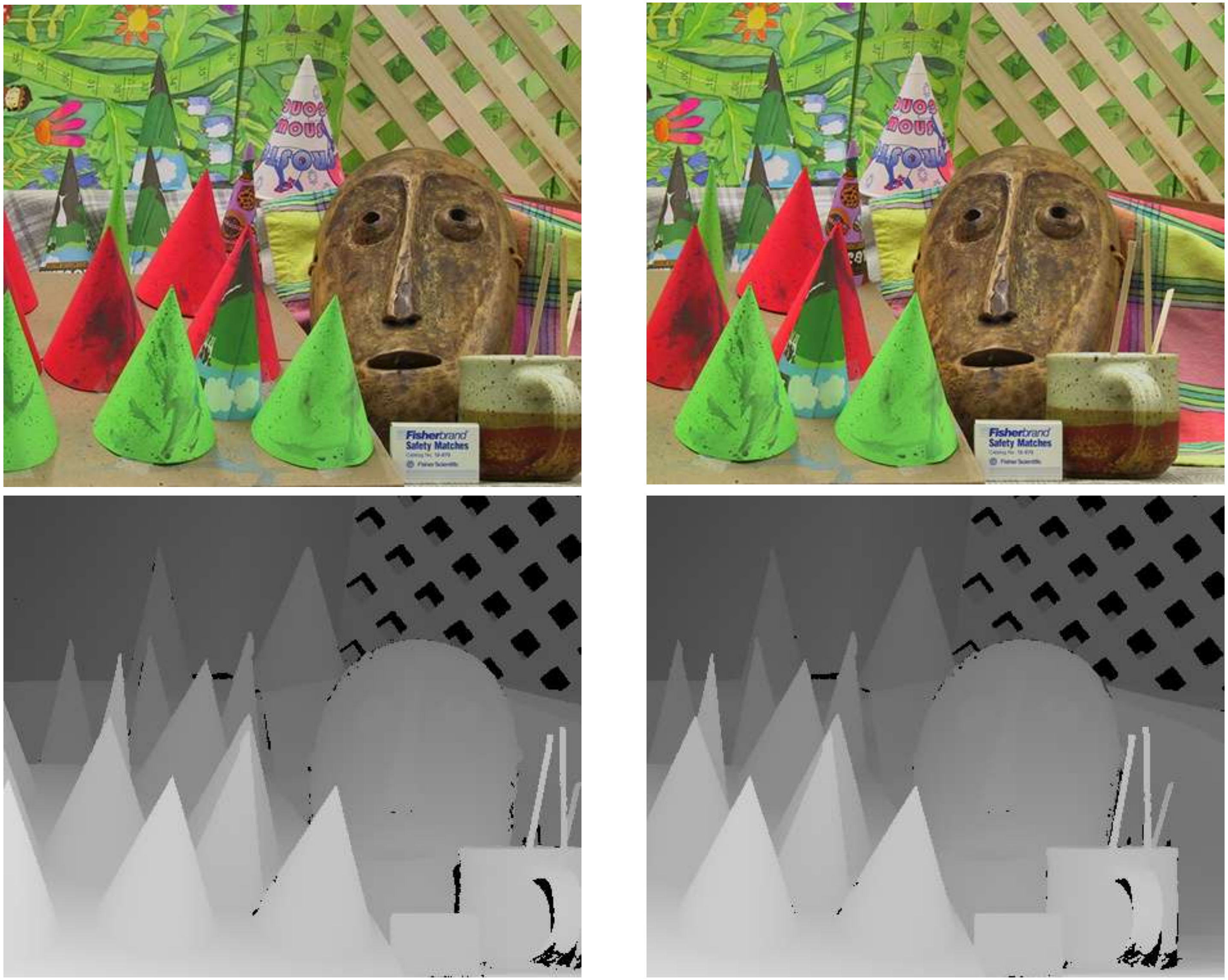


Stereo Vision Example: Depth Information

- Extracting depth information from a stereo image
  - a1 and a2: left and right image
  - b1 and b2: vertical edge filtered left and right image;
  - c: confidence image:
    bright = high confidence (good texture)
  - d: depth image:
    bright = close; dark = far



Stereo Vision:
Depth Maps



Scene Reconstruction

From depth maps, 3D models can be constructed by a triangular mesh

3D model from one angle   3D model from different angle   Completed model

Stereo Ranging Systems: Depth from triangulation

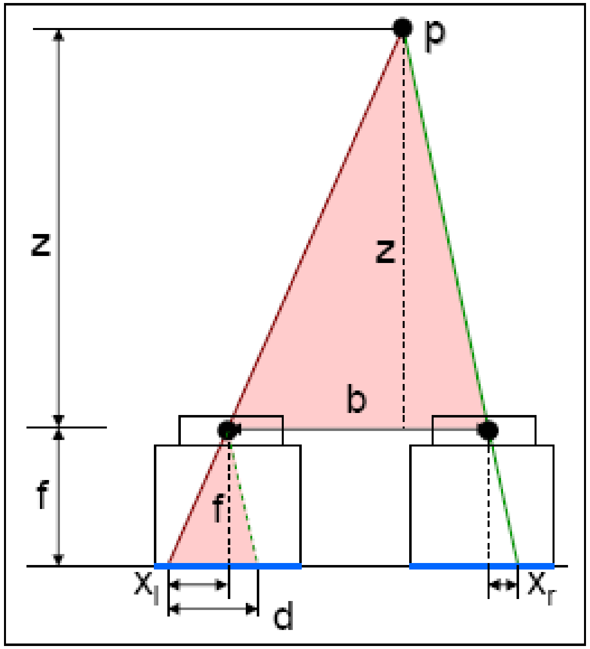
- If cameras point in the same direction and are aligned use geometry to calculate depth
  - b = baseline of cameras
  - z = depth of point p
  - d = disparity = $x_l - x_r$
  - f = focal point of cameras
- The 2 shaded triangles are *similar*, so

$$z = (fb)/d$$

$$y = y_l \, z/f = y_r \, z/f$$

$$x = x_l \, z/f$$

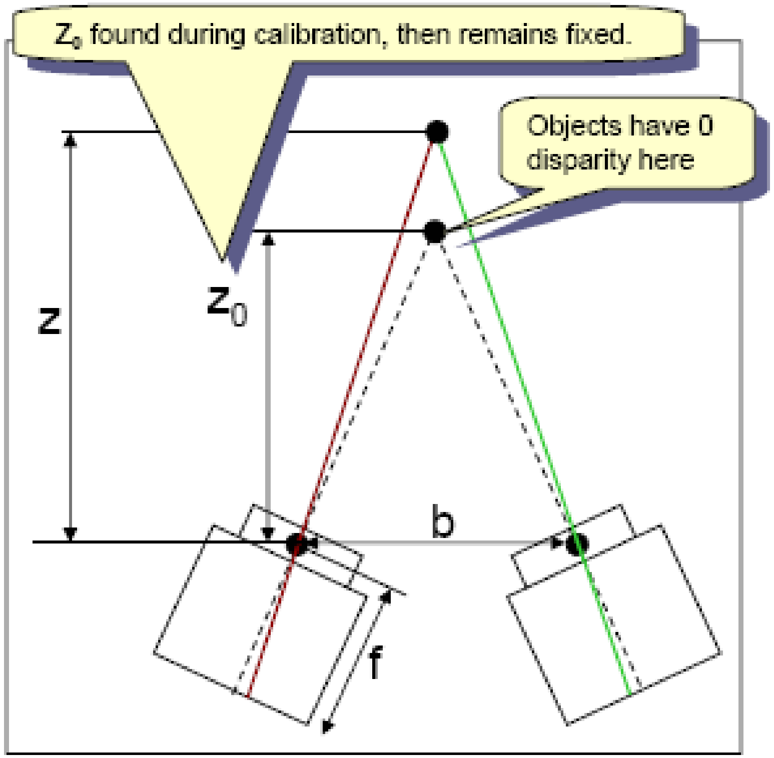
$$x_r = f(x - b)/z$$



Stereo Ranging Systems: Stereo Vision

A more realistic scenario is when the cameras do not lie on the same plane

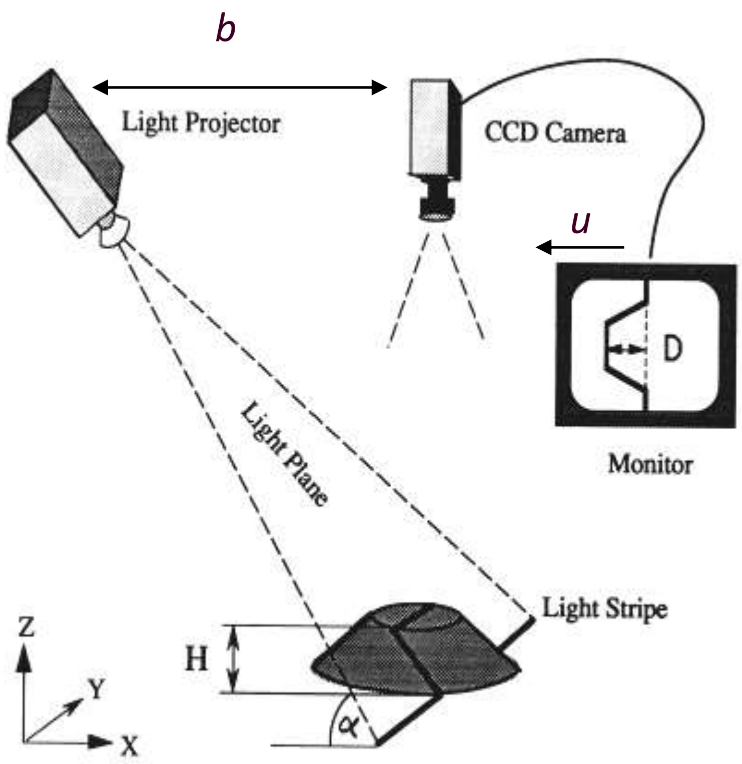
$z = (f \cdot b)/(d + (f \cdot b)/z_o)$

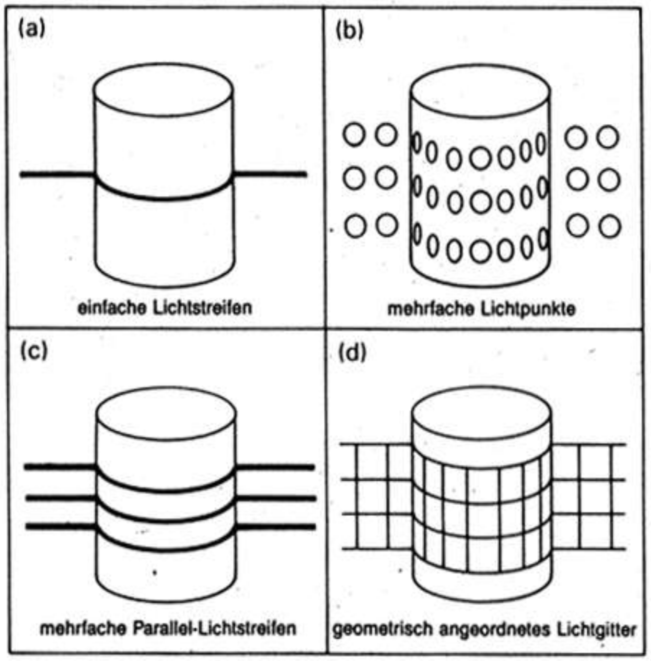


Structured Light (vision, 2 or 3D)

- *Triangulation* can be used to find the distance to a large set of points by replacing a 2D receiver by a CCD or CMOS camera
- The emitter must project a known patter, or *structured light*, onto the environment
  - Light textures
  - Collimated light with a rotating mirror
  - Laser stripe using a prism

Structured Light (vision, 2 or 3D)

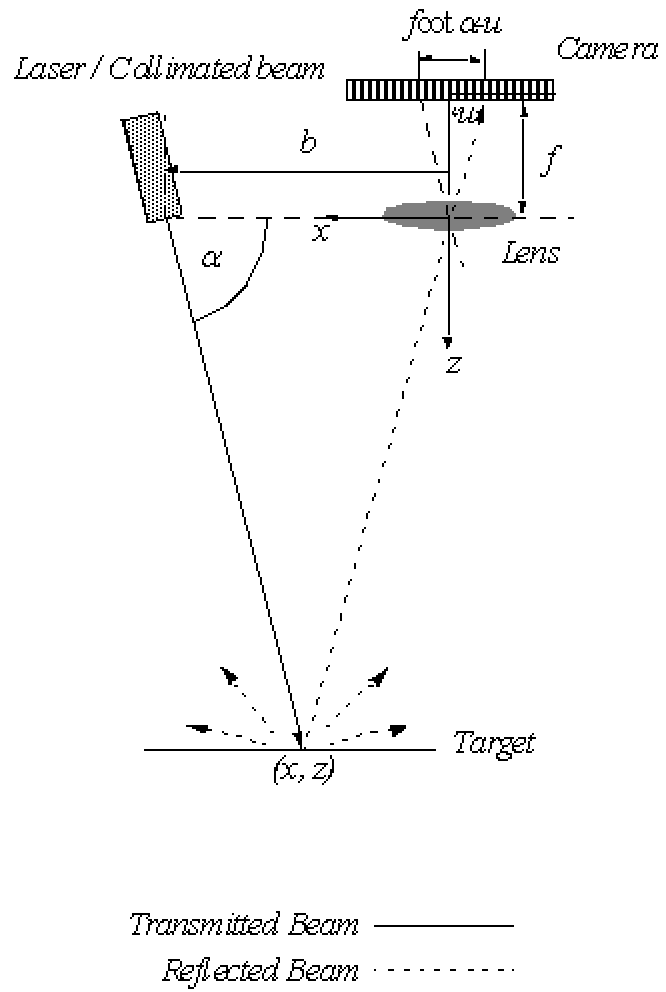




$$H = D \cdot \tan\alpha$$

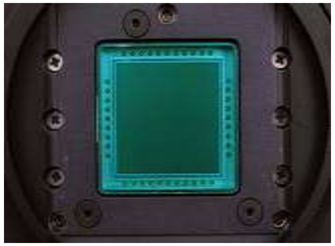
Structured Light (vision, 2 or 3D)

$$x = \frac{b \cdot u}{f\cot\alpha - u} \; ; \quad z = \frac{b \cdot f}{f\cot\alpha - u}$$



Vision-based Sensors: Hardware

- CCD (*light-sensitive, discharging capacitors of 5 to 25 micron*)



2048 x 2048 CCD array

Orangemicro iBOT Firewire

Sony DFW-X700

Cannon IXUS 300

- CMOS (*Complementary Metal Oxide Semiconductor technology*)

Stereo Ranging Systems

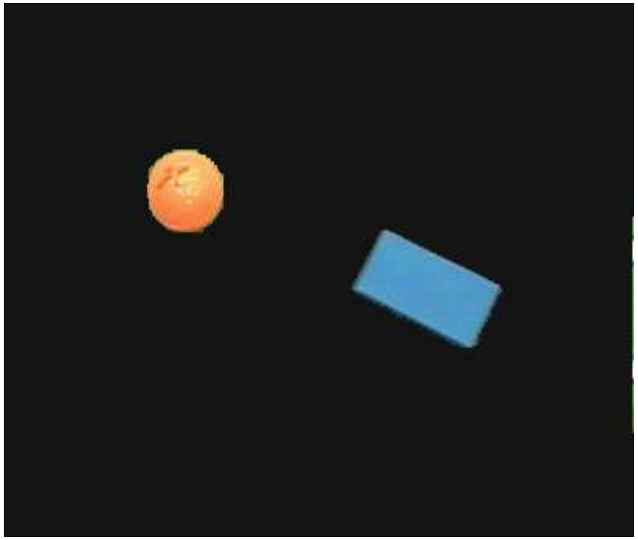
- Advantages
  - Better resolution than ultrasonic and infrared
  - Very reliable when environment is sufficiently cluttered
  - Often packaged with software to calculate depth
- Disadvantages
  - Cannot identify mirrors and/or glass
  - Sensitive to lighting conditions
  - Poor performance when environment lacks features
  - More expensive than ultrasonic and infrared
  - Larger than ultrasonic and infrared
  - Difficult to calibrate

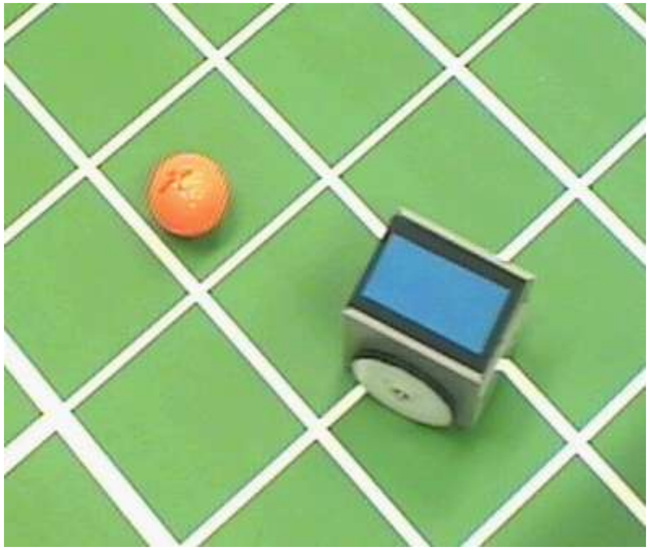
Robot Vision

- Vision processing can be a very complex problem
- Responding in real time to vision information is difficult
- Here are some ways to simplify the problem
  - Use color to identify uniquely colored objects
  - Use a combination of color and movement (color/blob tracking) . Mark important objects with salient or recognizable colors
  - Use a small image plane
  - Combine with simpler and faster sensors with vision
  - Use knowledge about the environment

Color Tracking Sensors

- Unlike ultrasonic and infrared range finders, vision systems can also detect and track color in the environment
- There is no correspondence problem to be solved in such algorithms (it only requires one image)
- By using sensor fusion, color tracking can produce significant information gains





Feature Extraction: Whole-Image Features

- Whole-Image features are not designed to identify specific spatial structures
- They serve as a compact representation of the entire local region
- Extract one or more features that are correlated with the robot's position for localization