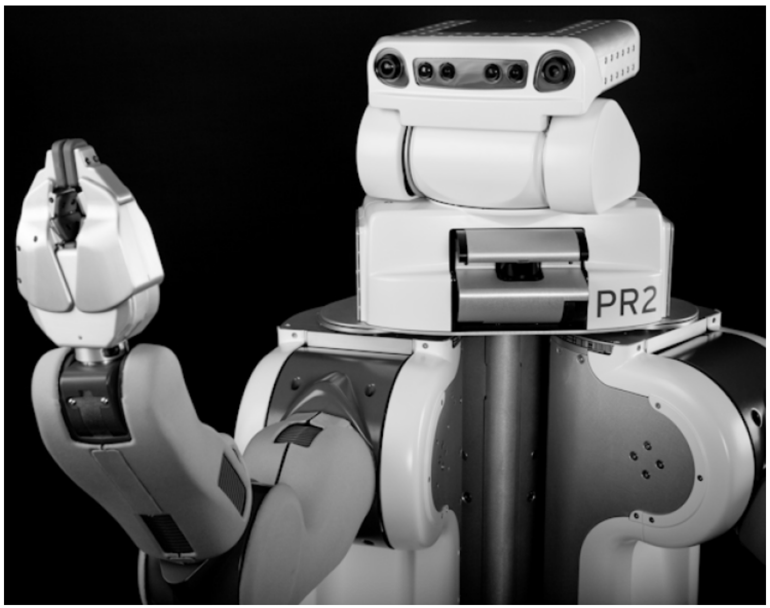


Chapter 8
Perception

Part 4

8.4 Aspects of Geometric and Semantic Computer Vision

Outline

- 8.4 Aspects of Geometric and Semantic Computer Vision
  - 8.4.1 Pixel Classification
  - 8.3.2 Computational Computer Vision
  - 8.3.3 Obstacle Detection
  - Summary

Outline

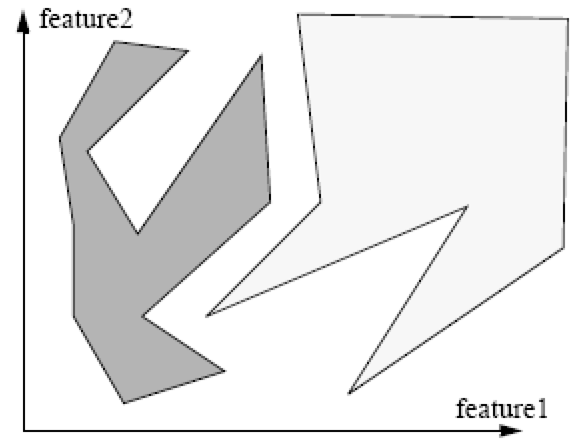
- 8.4 Aspects of Geometric and Semantic Computer Vision
  - <u>8.4.1 Pixel Classification</u>
  - 8.3.2 Computational Computer Vision
  - 8.3.3 Obstacle Detection
  - Summary

8.4.1 Pixel Classification

- Pixel Classification - assigns each pixel to one of a number of “classes”:
  - such as road, rock, bush, grass, yellow paint, etc.
  - useful for picking out the road for road following and obstacles for obstacle avoidance.

8.4.1 Pixel Classification

- Divide the environment into some number of classes and try to place a pixel in its appropriate class.
- Examples of sets of classes:
  - road/nonroad
  - vegetation/mineral/animal
  - hazard/nonhazard
  - hazard/nonhazard/not sure

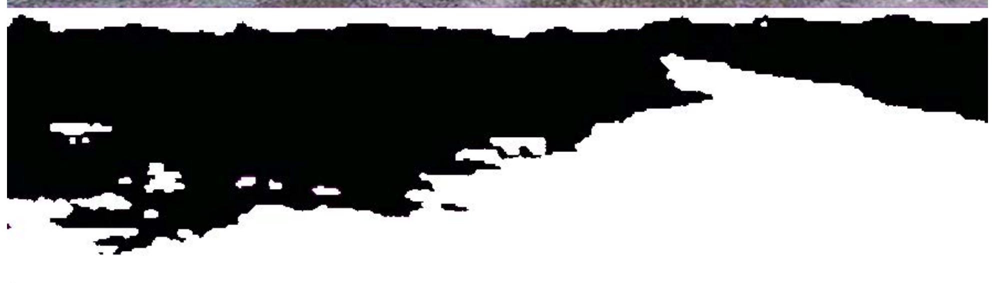


- Each pixel is considered to be a vector of attributes or “features” which lives in a multidimensional space.
- Regions in this space are supposed (conjectured) to correspond to classes.
- Its typical to have to preprocess images to remove effects of shadows, to normalize for texture etc.
- Sometimes each pixel is ascribed the properties of the region around it (e.g. texture).

8.4.1 Pixel Classification

- For example, suppose that green stuff is soft vegetation, brown stuff is hard vegetation, and grey stuff is dirt road.

Input

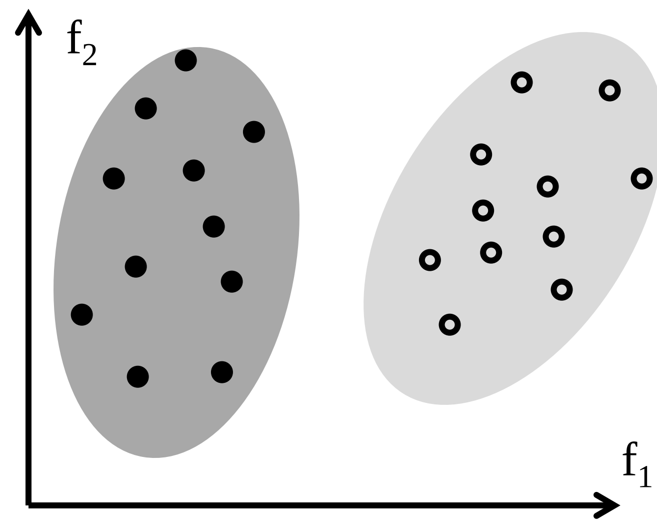


Output
(2 classes)

- This result could be passed into a trail following control algorithm

8.4.1.1 Training the Classifier

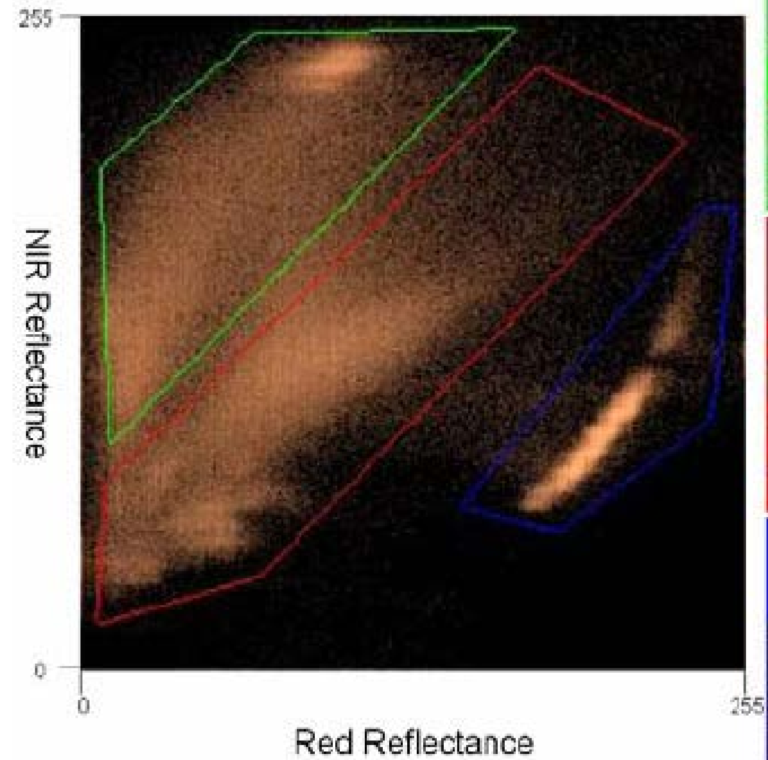
- Often, the regions are not known beforehand
  - Must be "learned".
- In supervised learning, we hand label (provide classes for) portions of images and use this data to determine the characteristics of the class.

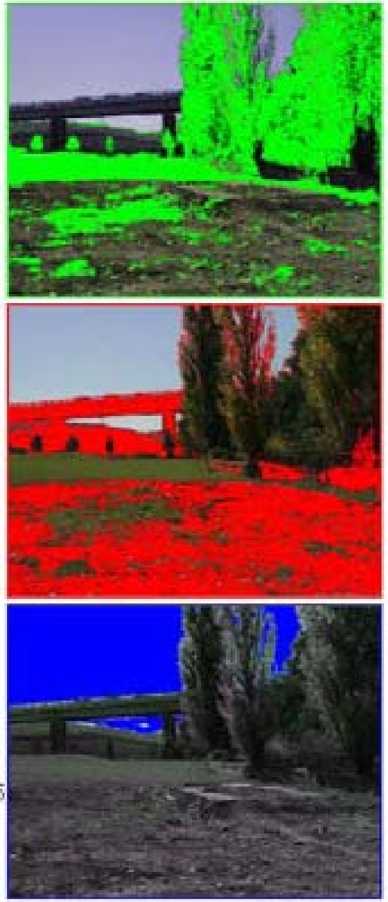


- One way to represent a class is in terms of its covariance matrix..

8.4.1.1 Training the Classifier
(Regions in Feature Space)

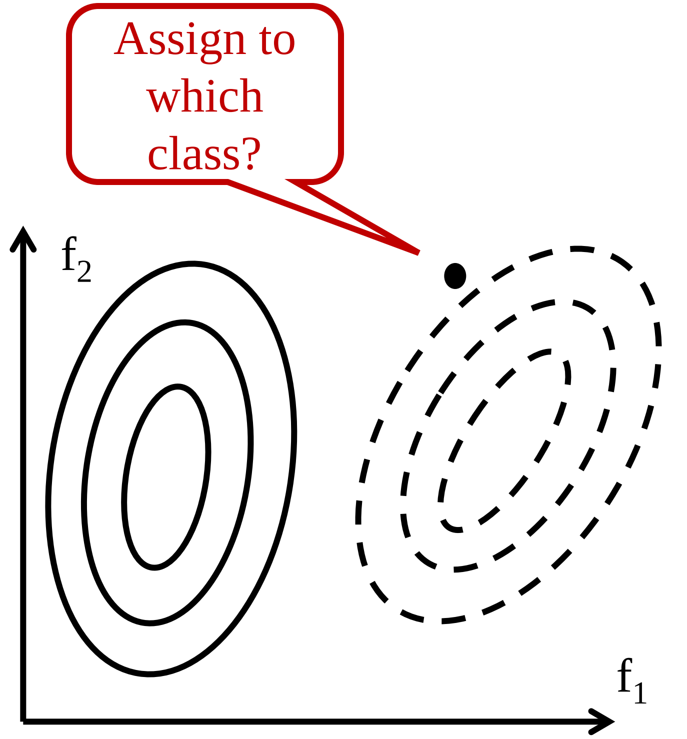
- Training can be accomplished by labeling regions in either feature space or image space.
- Near IR (NIR) is valuable for distinguishing vegetation.





8.4.1.2 Decision Surfaces

- Defines a rule to decide which class to which a pixel value belongs.
- If Gaussians are used to define P(class|rgb) then…
- MHD is a reasonable measure of proximity to the class mean.
- These days, a matrix multiply for each pixel is feasible.



$$d^2 = (\underline{x} - \underline{m})^T S^{-1} (\underline{x} - \underline{m})$$

$\underline{x}$ = pixel feature vector
$\underline{m}$ = class mean
S = covariance matrix

8.4.1.3 Fisher's Linear Discriminant

- Linear decision surface.
- Fast to compute.
- Define "within class" scatter:

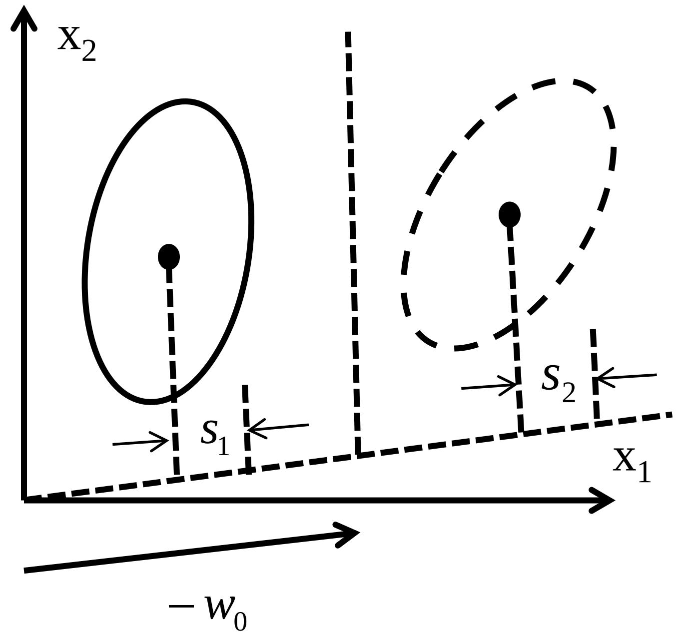
Measures spread within classes

$$S_W = S_1 + S_2$$

- Define the rank 1 "between class" scatter

$$S_B = (\underline{m}_1 - \underline{m}_2)(\underline{m}_1 - \underline{m}_2)^T$$

Measures spread between classes

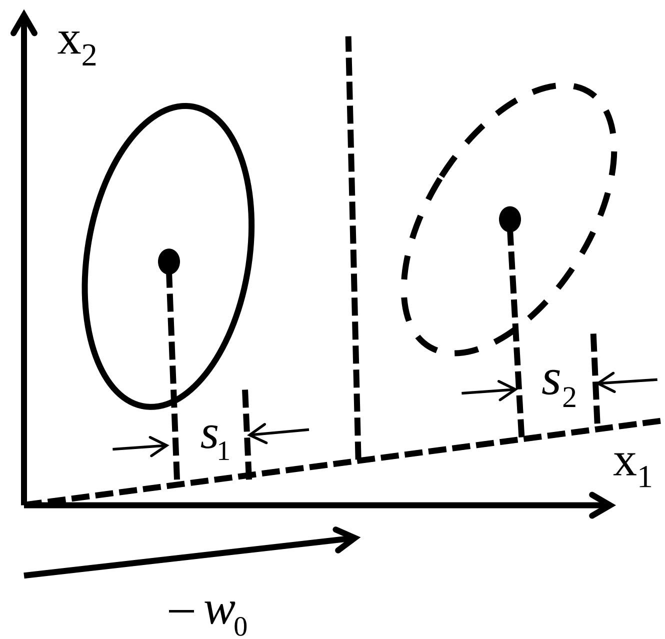


8.4.1.3 Fisher's Linear Discriminant

- Want to maximize the ratio: $J(\underline{w}) = \dfrac{\underline{w}^T S_B \underline{w}}{\underline{w}^T S_W \underline{w}}$
- The solution is:

$$\underline{w} = S_W^{-1}(\underline{m}_1 - \underline{m}_2)$$

- To classify a pixel, compute: $g(\underline{x}) = \underline{w}^T \underline{x} + \underline{w}_0$



- $\underline{w}_0$ represents the threshold on $\underline{w}^T \underline{x}$ which must be exceeded to make $g(x) > 0$ and cause a choice of class 2 over class 1

Outline

- 8.4 Aspects of Geometric and Semantic Computer Vision
  - 8.4.1 Pixel Classification
  - 8.3.2 Computational Computer Vision
  - 8.3.3 Obstacle Detection
  - Summary

Shape Inference

- These methods compute the range and/or shape of objects in the environment.
- Stereo - computes the range to all or some pixels in one of a number of images.
- Structured Light - same as stereo but light is projected onto the scene.
- Known Object - use known dimensions of object to determine range.
- Exotics such as range from focus, photometric stereo etc have seen little use.

8.4.2 Stereo Vision

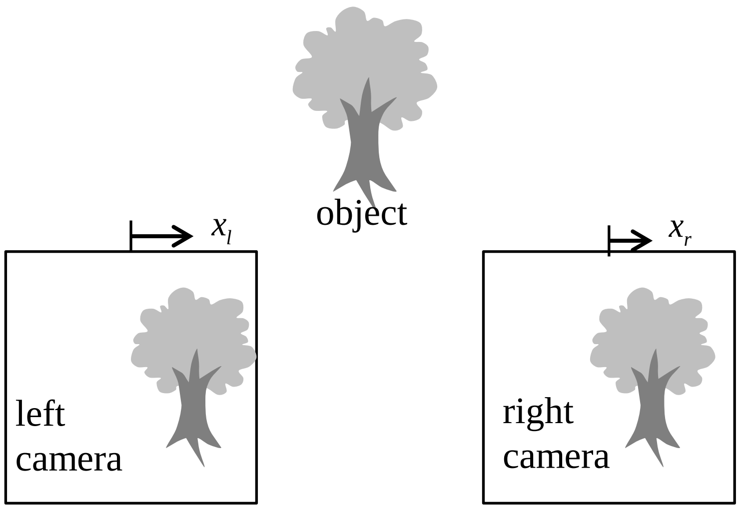
- While there are other options for passive ranging, it usually takes the form of stereo vision on mobile robots today.
  – MER Rovers Spirit and Opportunity use stereo.
- Structured light is a distant second.
  – The Mars Pathfinder Rover had a structured light system on board.
- Ranging may be performed only at specific features (say, at vertical lines) or everywhere in the image (“dense” stereo).
- Two-eyed (“binocular”) stereo is common but there are advantages to having more than two eyes.
- Cameras normally have <u>parallel</u> orientations (no “vergence”)

8.4.2.1 Principle of Operation

- Analogous to vision in primates.
- Nearer objects have greater disparity.
- Exploit this in reverse.

Eqn A

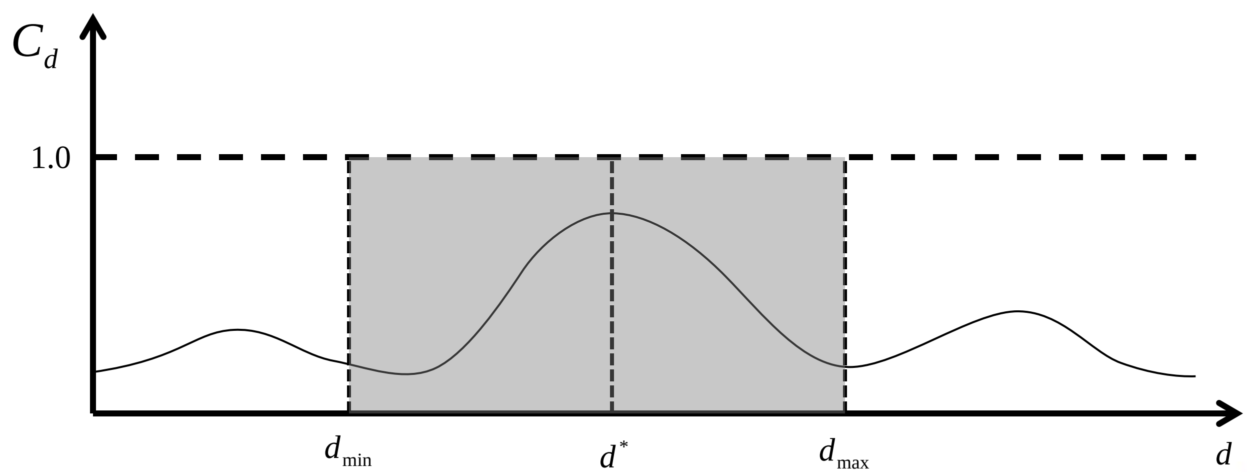
$$R = bf/d$$



- The hard part is the correspondences needed to get the disparity.

8.4.2.2 Search for Pixel Correspondences

- For each pixel in left image, correlate region around it with a line of pixels in the right image.
  - Generates a curve of similarity versus disparity.
- Find disparity of maximum correlation.
- Image noise, distortions, poor calibration, and many other error sources conspire to make the correlation calculations unreliable.



8.4.2.2 Search for Pixel Correspondences

(Horopter Stereo)

- Scene shape affects the distortion of regions from eye to eye.
  - Highest when disparity gradient in image is highest.
- Horopter technique assumes a reasonable disparity gradient based on flat terrain.

Input Intensity Image

Output Range Image

8.4.2.4 Advantages and Disadvantages

(Advantages)

- Passive.
- Solid state.
- Density.
  - Data is relatively dense
  - Though not necessarily of high angular resolution.
- Cost. Now relatively inexpensive.
- Appearance Registration.
  - Appearance and range data are <u>inherently aligned.</u>
- Frame capture.
  - <u>No distortion</u> within an image due to vehicle motion.

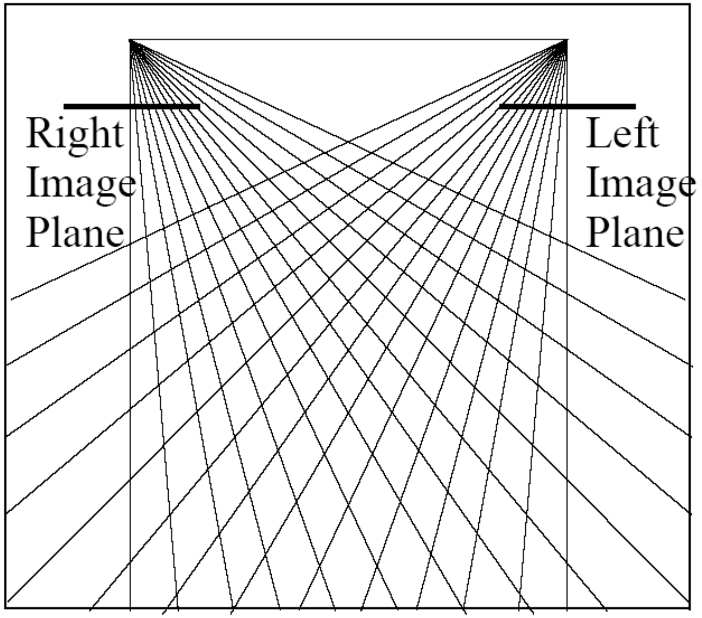
8.4.2.4 Advantages and Disadvantages

(Disadvantages)

- Calibration.
  - Relies on pixel to pixel alignment of imagery.
- Range resolution.
  - Resolution degrades quadratically with range.
- Angular resolution.
  - Correlation processing acts as a low pass filter. Reduces res by order of magnitude.
- Passive.
  - Stereo fails under near darkness, or no texture conditions.
- Triangulation.
  - Increased baseline leads to increased distortion and missing parts problems.
- Processing.
  - Requires a dedicated high performance processor. Ladars do all that in hardware. No longer such a big deal.

Resolution

- Downrange resolution can be determined by differentiating Eqn A:



$$\Delta R = (-bf/d^2)\Delta d = [-R^2/(bf)]\Delta d$$

- Define normalized disparity:

$$\delta = d/f$$

- Range resolution is now:

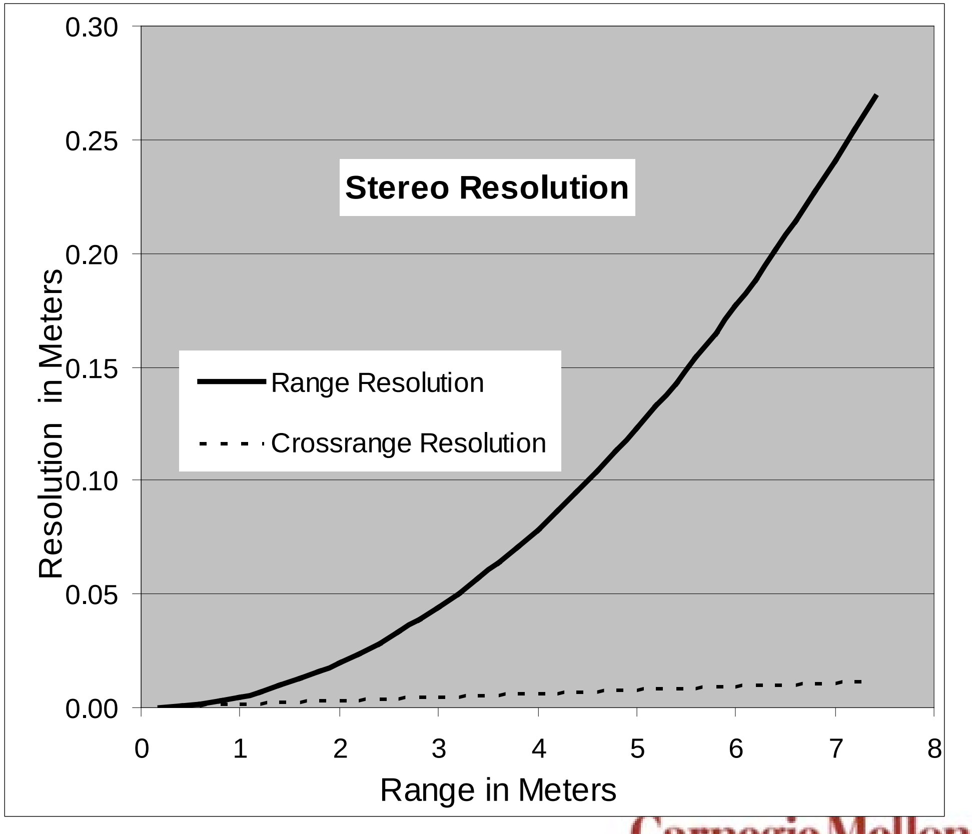
$$\Delta R = [-R^2/b]\Delta\delta$$

- Crossrange resolution is linear:

$$\Delta n = R\Delta\delta$$

Resolution

- In practice, stereo operates by matching regions of perhaps 10 X 10 pixels between imagery.
  - Range values become correlated

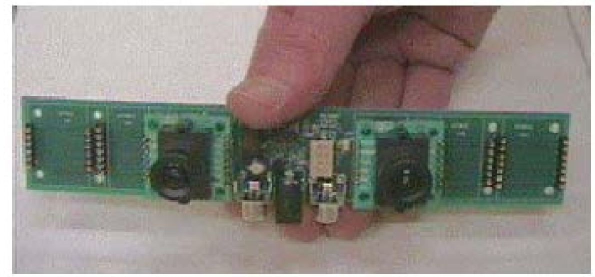


- These resolution numbers must be reduced by a factor of 10.

Commercial Stereo

Parameter	Point Grey BumbleBee & Triclops	SRI Small Vision System	Sarnoff Acadia Stereo
Image Dimen-sions	640 X 480 (higher pending)	320 X 240	640 X 480
Frame Rate	30 Hz	20 Hz	60 Hz
Size	16 X 4 X 4 cm	6 X 2 X 2 inches	(user sup-plies cam-eras)
Baseline	12 cm	6.2 inches	(user sup-plied)
Disparities	24	24	64

Point Grey Bumble Bee

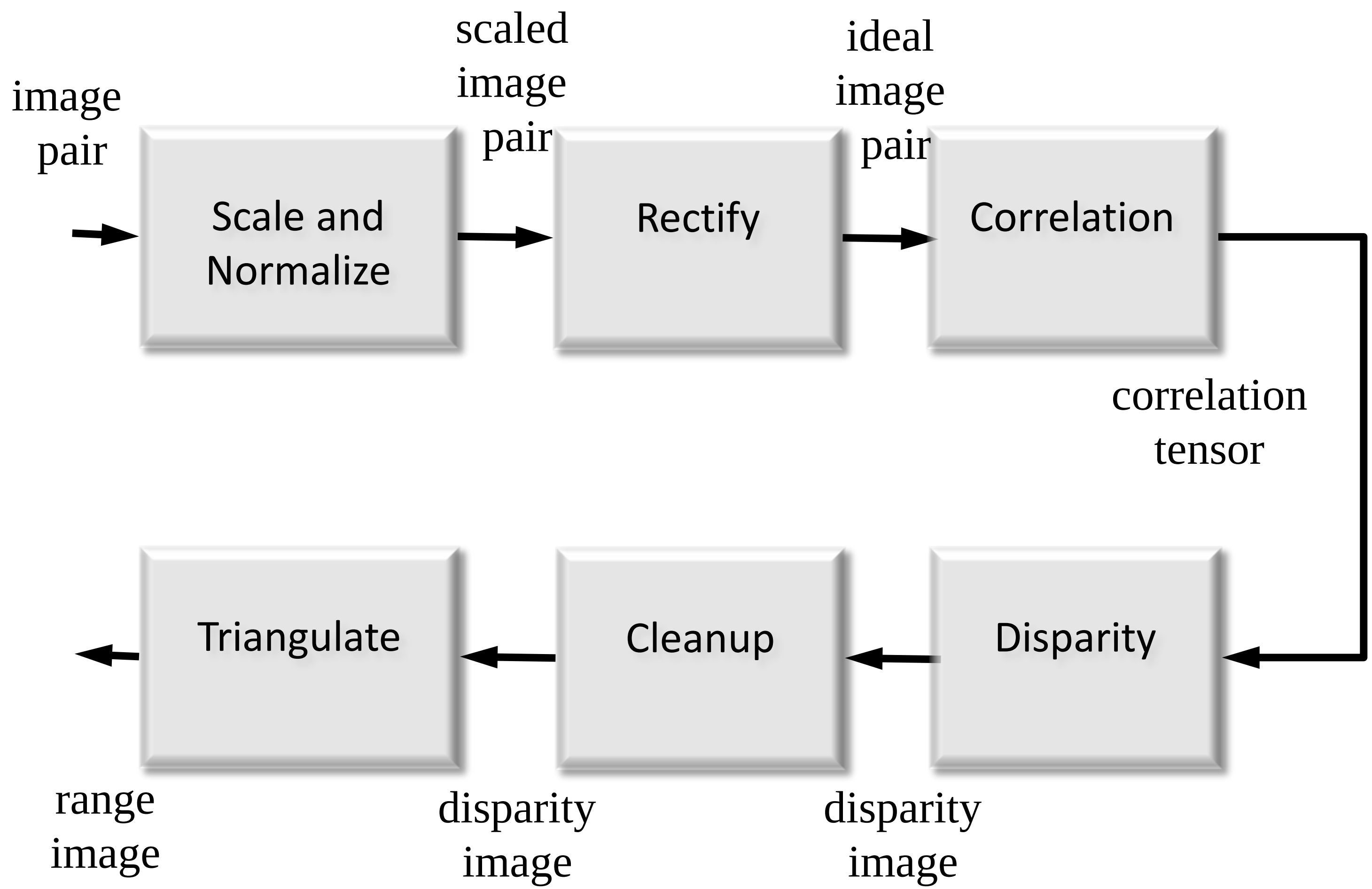


SRI Small Vision System

Sarnoff Acadia Board

- Target markets may favor higher frame rates but better data is more useful to a mobile robot.

8.4.2.5 Data Flow



Complexity

$$f_{stereo} = (2K_1 + 2K_2 + K_5 + K_6)RC + (K_3 + K_4)RCD$$

R = rows
C = cols
D = disparities

- Bottom Line: Its RCD

Outline

- 8.4 Aspects of Geometric and Semantic Computer Vision
  - 8.4.1 Pixel Classification
  - 8.3.2 Computational Computer Vision
  - <u>8.3.3 Obstacle Detection</u>
  - Summary

Obstacle Detection

- A wide variety of approaches have been tried.
  - Their success ranges from high to low based, mostly, on the difficulty of the environment.
- Under stationary environment assumption:
  - Dwell and evidence accumulation is possible if you measure ego-motion.
  - Evidence acquired from different perspectives may be important for resolving power (e.g. for wide beam sensors)
- When assumption is wrong, moving things are subject to motion smear and associated false positives and negatives.
- Mapping from sensor readings to map cells may be:
  - One to many – sonar
  - Many to one - ladar

Tradeoffs

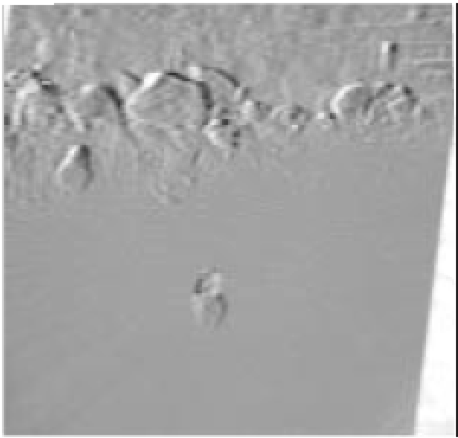
- When evidence is accumulated over time, some sort of map becomes necessary in order to:
  - Have a place to store (memory) intermediate data / results.
  - Compensate for the effects of vehicle motion (register readings).
- False negatives can usually only be reduced by increasing false positives.
  - In the limit, the stationary robot will hit nothing.
- Most approaches benefit from accumulating evidence but ...
  - There may be strong real-time constraints - related to deciding its an obstacle before you hit it.
  - Hence, there is a strong response-resolution tradeoff. Good answers or fast answers - pick any one.

8.4.3.1 Evidence

- The evidence of an obstacle can take many forms…….

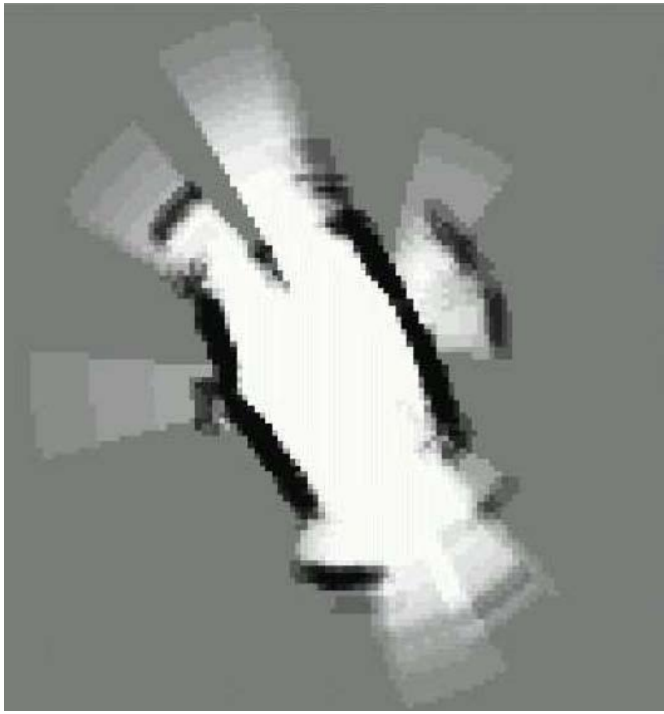
8.4.3.1.1 Deviation from Expectations

- When the world is boring, deviations from the norm are obstacles.
- Easy to do this indoors with range imagery.
  - Even right in the disparity image in stereo.



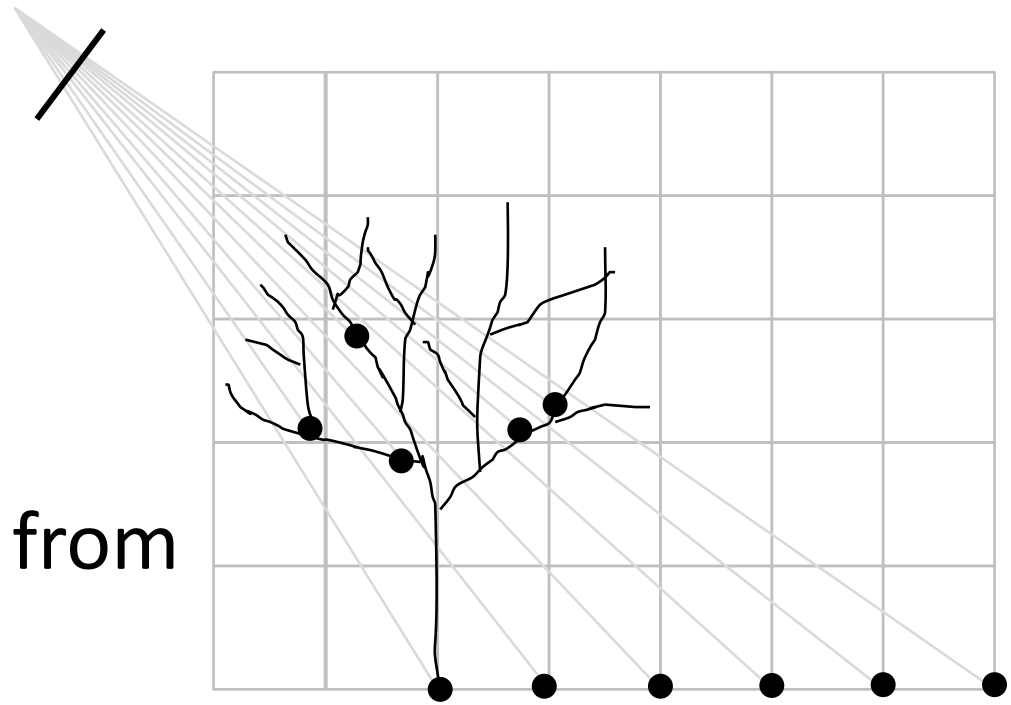
8.4.3.1.2 Occupancy / Presence

- Anything (other than me and floor) is bad.
- Only works in simplest of environments.
- Common approach when sensors have poor resolving power.
- 2D and 3D grids are commonly used to accumulate evidence.



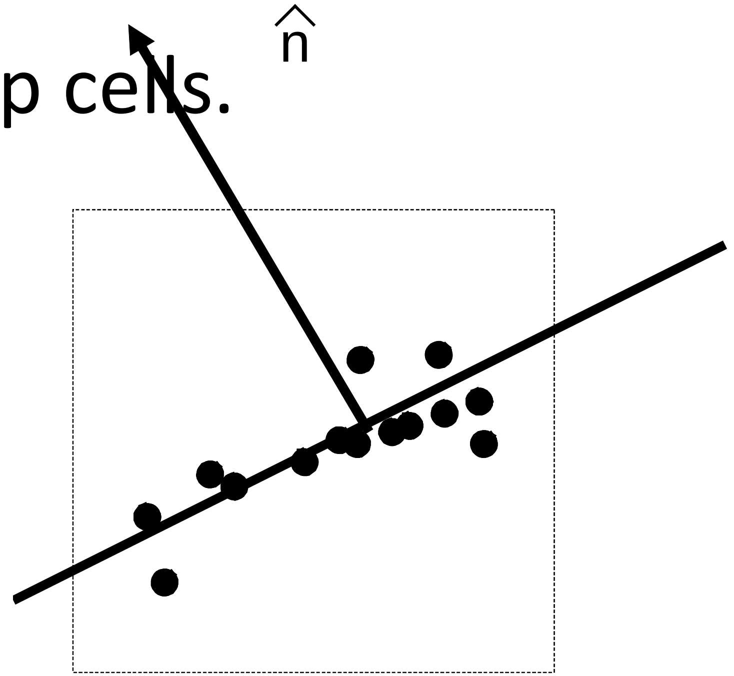
8.4.3.1.4 Density

- Track ratio of:
  - hits/misses
- Not truly density but related.
- Helps distinguish rock from bush.



8.4.3.1.5 Slope

- Sometimes slope is main attribute of interest.
- Compute scatter matrix in map cells.
- Best fit plane:



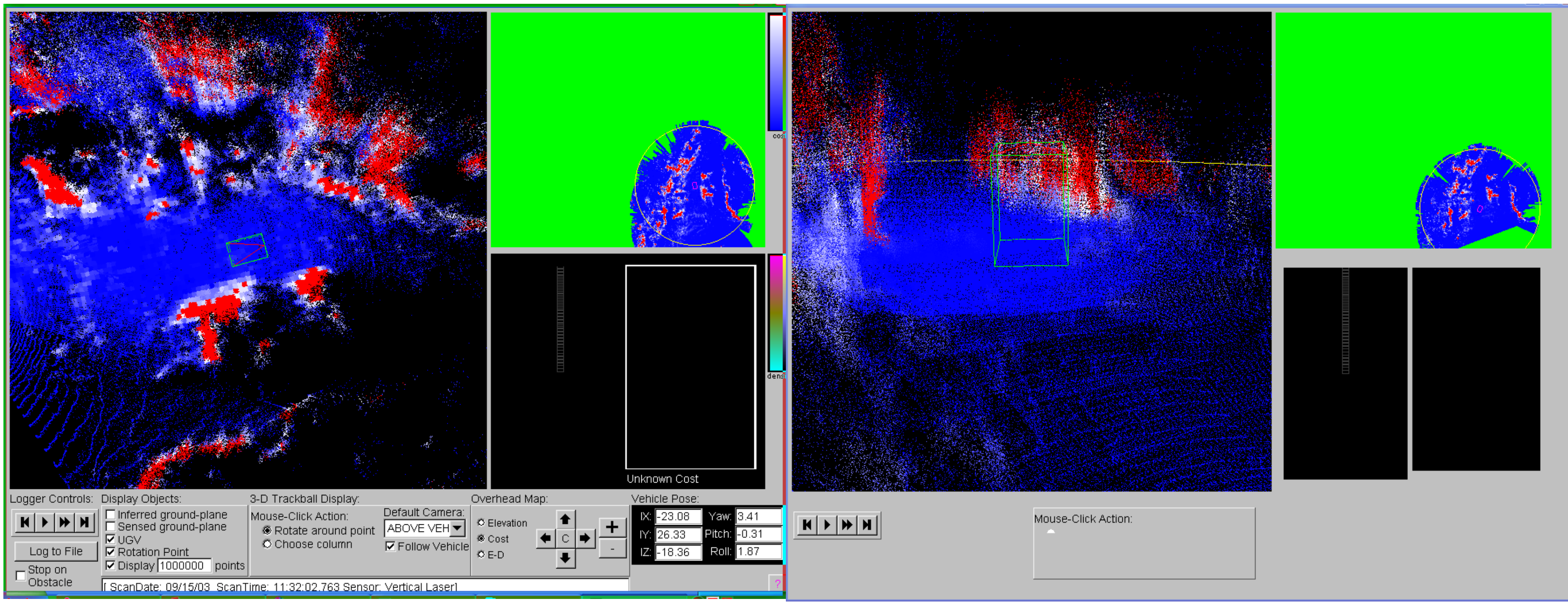
- Fit data with:

$$a'x + b'y + c'z = 1$$

$$\begin{bmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \\ \dots & \dots & \dots \end{bmatrix} \begin{bmatrix} a' \\ b' \\ c' \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \dots \end{bmatrix}$$

Check normal direction relative to vertical to determine if cell is an obstacle

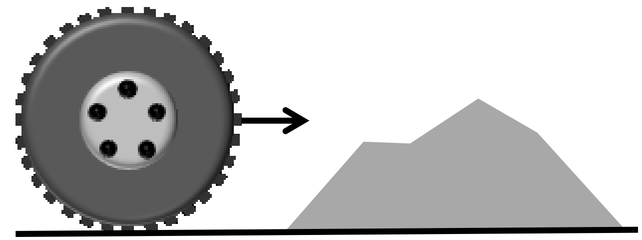
Video



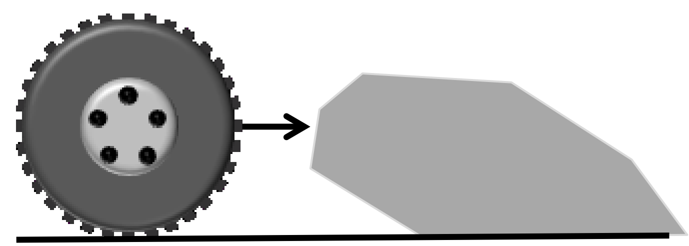
Carnegie Mellon
THE ROBOTICS INSTITUTE

8.4.3.1.6 Shape

- Virtually no work has been done on this problem.
- Contemporary solution is to check the slope or height change.



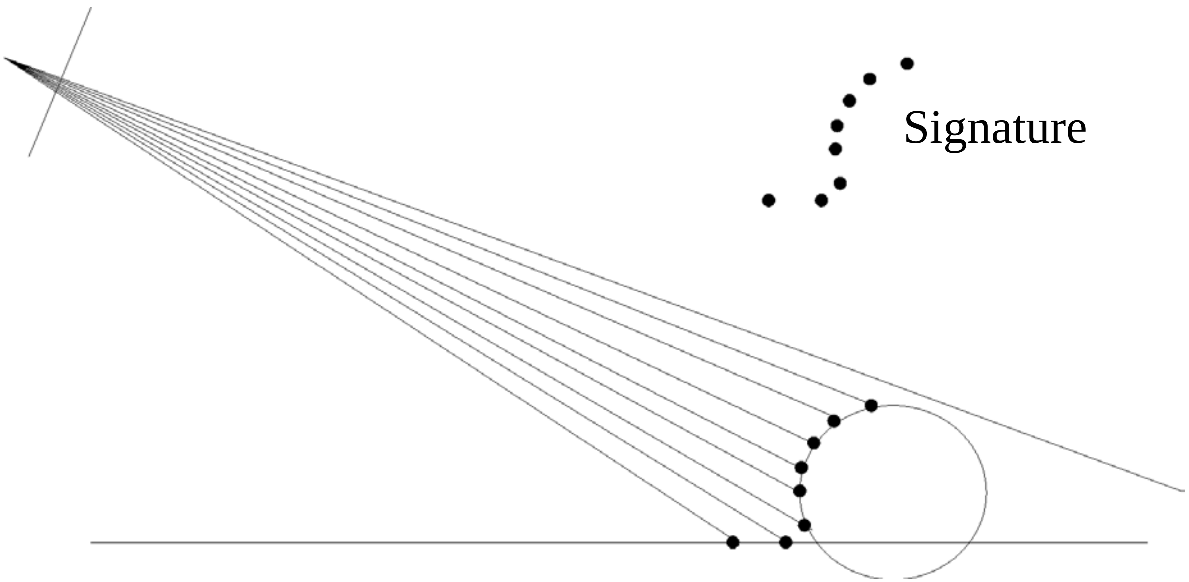
No Problem



Ouch!!!

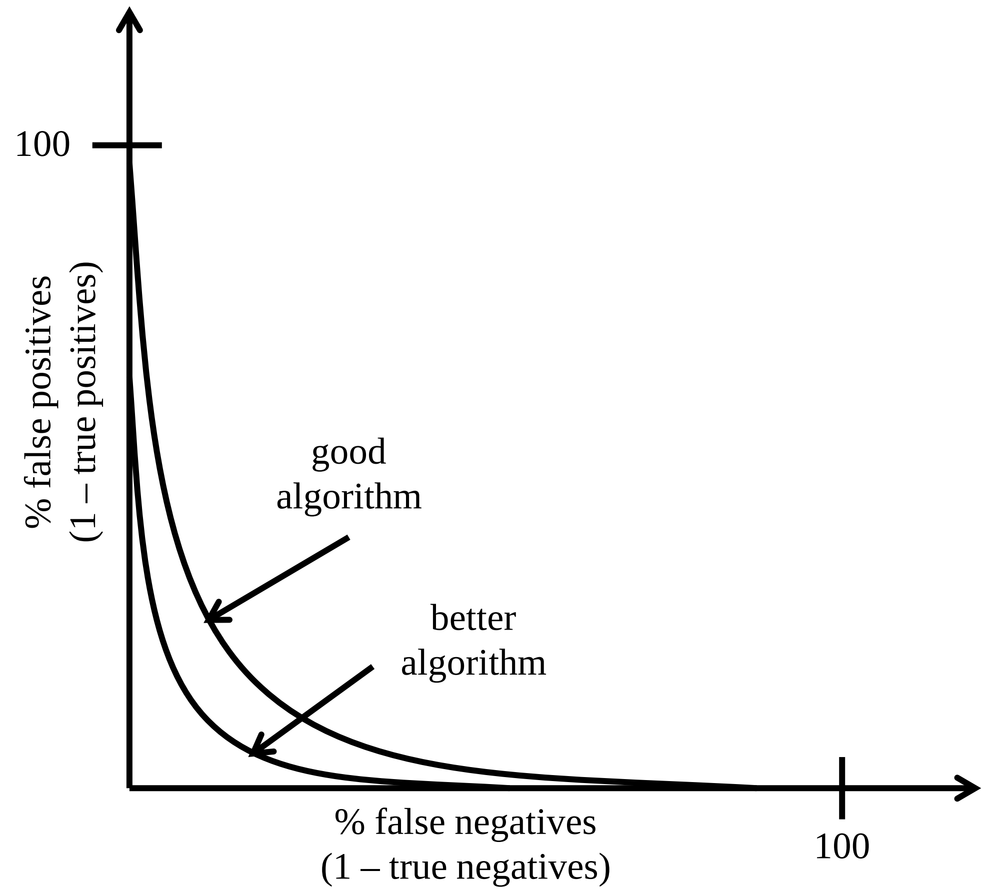
Signature Recognition

- If particular obstacles are prevalent, the problem can become simply recognition.
- Consider ladar signature of a fallen tree.



8.4.3.2 Performance

- Obstacles must be detected in time to react.
  – That means when they are far away.
- The tradeoff between false positives and negatives leads to a “pick your poison” trade.



8.4.3.2.2 Vehicle Speed

(Pathology: Speed Dependent Resoluti[on])

- Basic requirement:

$$\delta = \frac{[h/Y(V)]}{n}$$

n=4 detects presence only, takes more for resolution of smallest

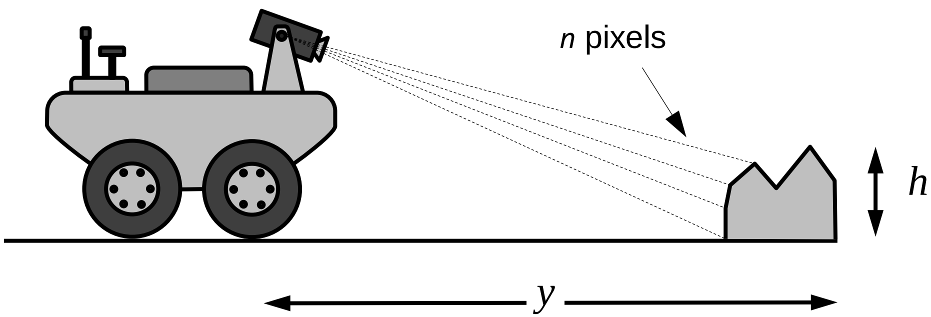
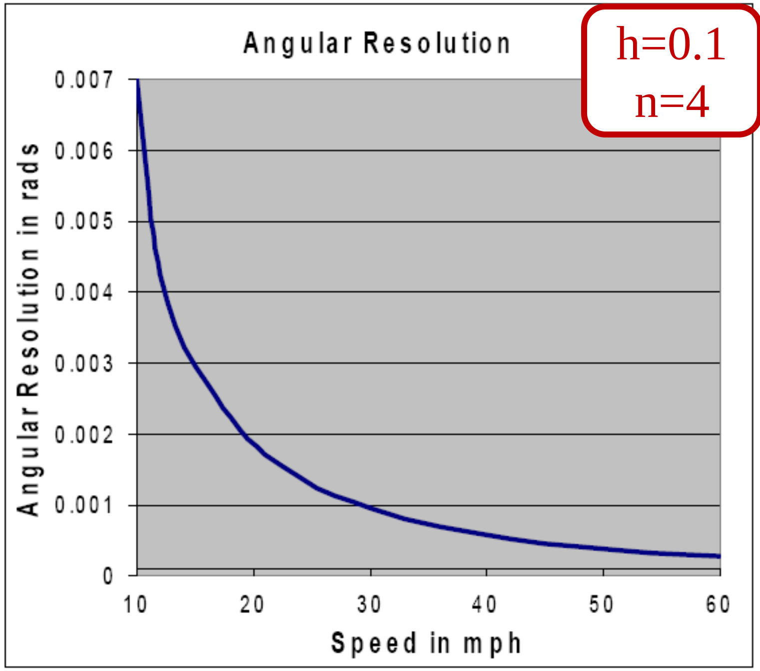


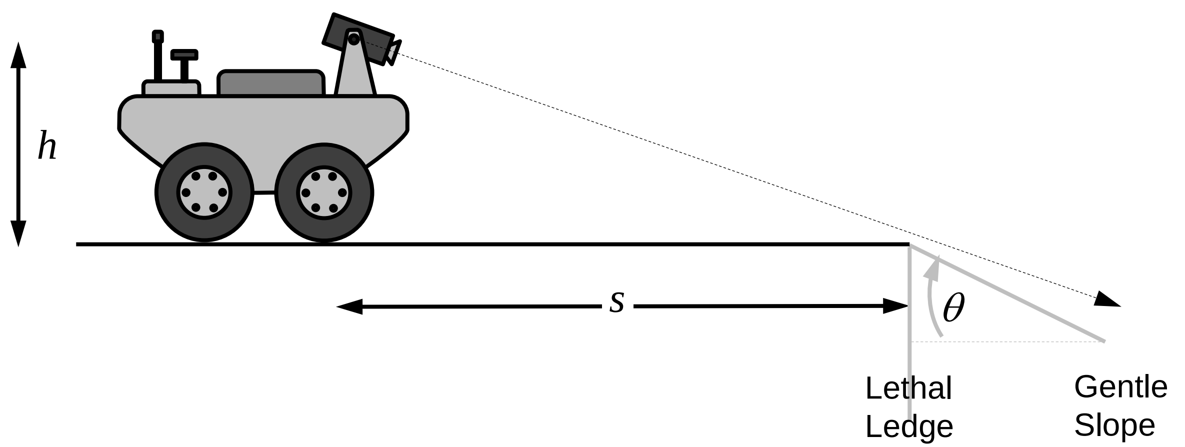
TABLE 2. Sensor technologies and their angular resolutions

Sensor	Resolution (mrads)
Laser	3 (Typical)
Stereo	80
Radar	worse



8.4.3.2.3 Pathological Obstacles
(Negative Obstacles)

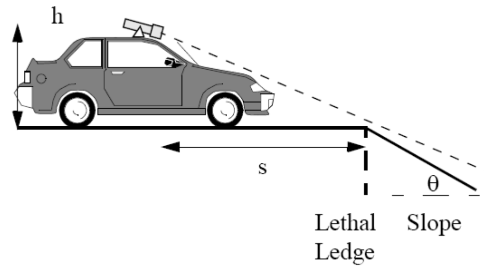
- In addition to obvious terrain self-occlusion, there are more subtle cases.
- The front edge of a negative obstacle occludes most of the information required to determine that it is a negative obstacle.



Distinguishing a downslope from a ledge cannot be done until close enough (maybe too close to stop)

8.4.3.2.3 Pathological Obstacles

(Negative Obstacles)



- Vehicle must be close enough to satisfy:

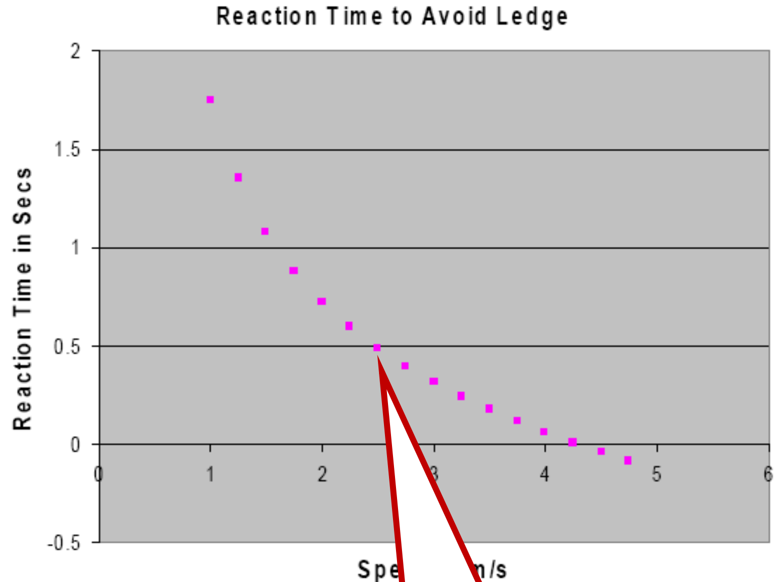
$$s = h / \tan\theta$$ Detection range

- Substitute for stopping distance and solve for reaction time:

$$T_{react} = \frac{(h/\tan\theta - V^2/(2\mu g))}{V}$$

- Increased resolution makes no difference.

At long range, a tiny bump can hide a vehicle sized hole!!



Physics induced speed limit

Outline

- 8.4 Aspects of Geometric and Semantic Computer Vision
  - 8.4.1 Pixel Classification
  - 8.3.2 Computational Computer Vision
  - 8.3.3 Obstacle Detection
  - Summary

Summary

- Computer vision is now a very deep field independent of robotics.
  - Only some of it is highly relevant to mobile robotics.
- Pixel classification is a rapid end-to-end transformation of sensed data onto domain relevant classes.
- Stereo is great because its passive and because it gives co-registered appearance data.
  - It requires a solution to the correspondence problem.
  - Dense stereo requires it at every point.
  - Computational load is proportional to rows X columns X depth levels.
- Stereo is starting to become a commercial commodity.
- Detecting features is an important process for mobile robots. It has uses in detecting shapes and for computing egomotion.
- In indoor range data, curvature features like corners are a good source of localization information.

Summary

- Many schemes have been used for obstacle detection with varying success.
- Detecting small obstacles at high speeds or negative ones at any speed is a daunting challenge.