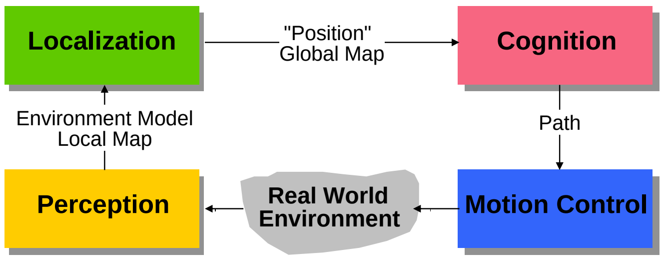
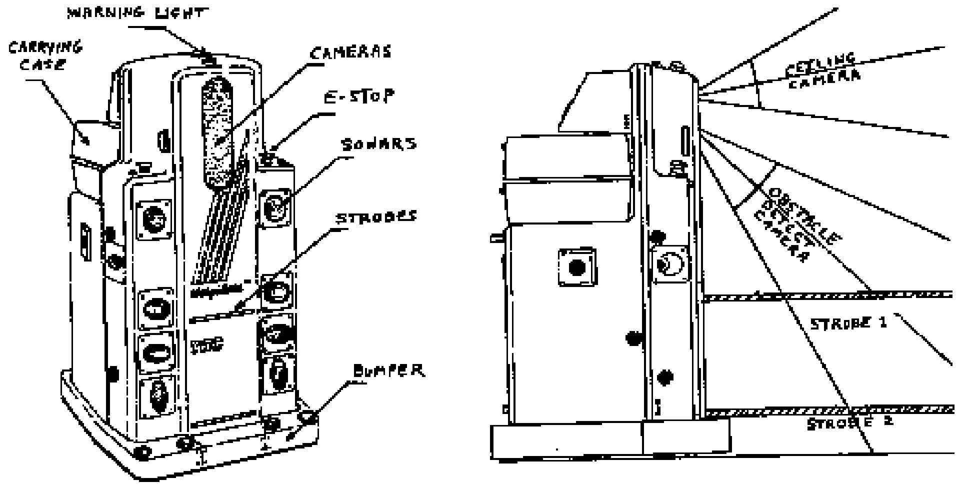
Perception

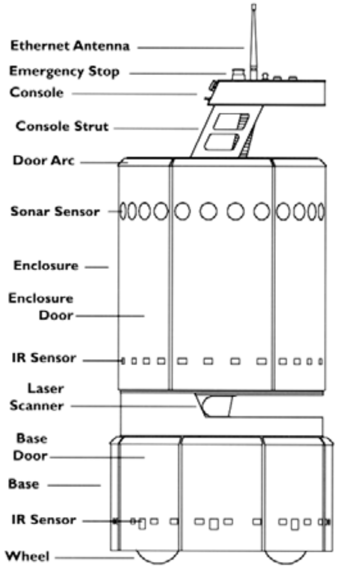
- Sensors
- Uncertainty
- Features



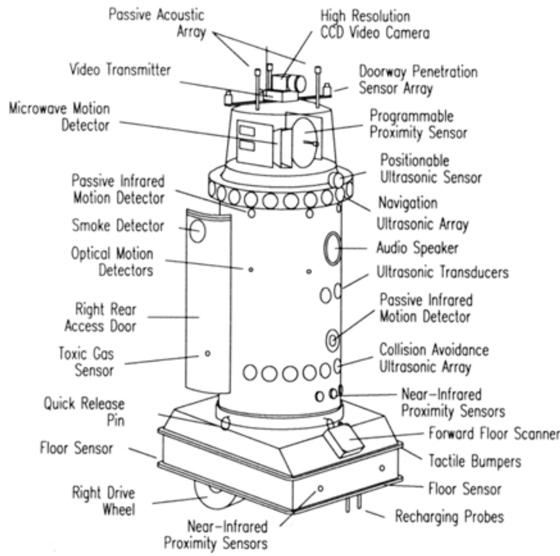
Example HelpMate, Transition Research Corp.



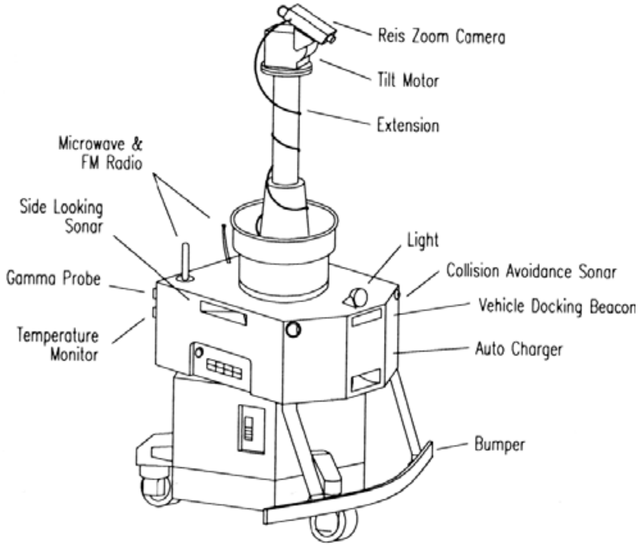
Example B21, Real World Interface



Example Robart II, H.R. Everett



Savannah, River Site Nuclear Surveillance Robot



BibaBot, BlueBotics SA, Switzerland

IMU
Inertial Measurement Unit

Emergency Stop Button

Wheel Encoders

Omnidirectional Camera

Pan-Tilt Camera

Sonar Sensors

Laser Range Scanner

Bumper

Classification of Sensors

- Proprioceptive sensors
  - *measure values internally to the system (robot),*
  - *e.g. motor speed, wheel load, heading of the robot, battery status*
- Exteroceptive sensors
  - *information from the robots environment*
  - *distances to objects, intensity of the ambient light, unique features.*
- Passive sensors
  - *energy coming for the environment*
- Active sensors
  - *emit their proper energy and measure the reaction*
  - *better performance, but some influence on envrionment*

General Classification (1)

General classification (typical use)	Sensor Sensor System	PC or EC	A or P
Tactile sensors (detection of physical contact or closeness; security switches)	Contact switches, bumpers Optical barriers Noncontact proximity sensors	EC EC EC	P A A
Wheel/motor sensors (wheel/motor speed and position)	Brush encoders Potentiometers Synchros, resolvers Optical encoders Magnetic encoders Inductive encoders Capacitive encoders	PC PC PC PC PC PC PC	P P A A A A A
Heading sensors (orientation of the robot in relation to a fixed reference frame)	Compass Gyroscopes Inclinometers	EC PC EC	P P A/P

A, active; P, passive; P/A, passive/active; PC, proprioceptive; EC, exteroceptive.

General Classification (2)

General classification (typical use)	Sensor Sensor System	PC or EC	A or P
Ground-based beacons (localization in a fixed reference frame)	GPS Active optical or RF beacons Active ultrasonic beacons Reflective beacons	EC EC EC EC	A A A A
Active ranging (reflectivity, time-of-flight, and geometric triangulation)	Reflectivity sensors Ultrasonic sensor Laser rangefinder Optical triangulation (1D) Structured light (2D)	EC EC EC EC EC	A A A A A
Motion/speed sensors (speed relative to fixed or moving objects)	Doppler radar Doppler sound	EC EC	A A
Vision-based sensors (visual ranging, whole-image analysis, segmentation, object recognition)	CCD/CMOS camera(s) Visual ranging packages Object tracking packages	EC	P

Characterizing Sensor Performance (1)

Measurement in real world environment is error prone

- Basic sensor response ratings
  - *Dynamic range*
    - *ratio between lower and upper limits, usually in decibels (dB, power)*
    - *e.g. power measurement from 1 Milliwatt to 20 Watts*

$$20 \cdot \log\left[\frac{20}{0.001}\right] = 86dB$$

    - *e.g. voltage measurement from 1 Millivolt to 20 Volt*
    - 20 instead of 10 because square of voltage is equal to power!!
  - *Range*
    - *upper limit*

Characterizing Sensor Performance (2)

- Basic sensor response ratings (cont.)
  - *Resolution*
    - *minimum difference between two values*
    - *usually: lower limit of dynamic range = resolution*
    - *for digital sensors it is usually the A/D resolution.*
      e.g. 5V / 255 (8 bit)
  - *Linearity*
    - *variation of output signal as function of the input signal*
    - *linearity is less important when signal is after treated with a computer*
  - *Bandwidth or Frequency*
    - *the speed with which a sensor can provide a stream of readings*
    - *usually there is an upper limit depending on the sensor and the sampling rate*
    - *Lower limit is also possible, e.g. acceleration sensor*

In Situ Sensor Performance (1)

Characteristics that are especially relevant for real world environments

- Sensitivity
  - *ratio of output change to input change*
  - *however, in real world environment, the sensor has very often high sensitivity to other environmental changes, e.g. illumination*
- Cross-sensitivity
  - *sensitivity to environmental parameters that are orthogonal to the target parameters*
- Error / Accuracy
  - *difference between the sensor's output and the true value*

error

$$\left(accuracy = 1 - \frac{|m - v|}{v}\right)$$

***m* = measured value**
***v* = true value**

In Situ Sensor Performance (2)

Characteristics that are especially relevant for real world environments

- Systematic error -> deterministic errors
  - *caused by factors that can (in theory) be modeled -> prediction*
  - *e.g. calibration of a laser sensor or of the distortion cause by the optic of a camera*
- Random error -> non-deterministic
  - *no prediction possible*
  - *however, they can be described probabilistically*
  - *e.g. Hue instability of camera, black level noise of camera ..*
- Precision
  - reproducibility *of sensor results*

$$precision = \frac{range}{\sigma}$$

Characterizing Error: The Challenges in Mobile Robotics

- Mobile Robot has to perceive, analyze and interpret the state of the surrounding
- Measurements in real world environment are dynamically changing and error prone.
- Examples:
  - *changing illuminations*
  - *specular reflections*
  - *light or sound absorbing surfaces*
  - *cross-sensitivity of robot sensor to robot pose and robot-environment dynamics*
    - *rarely possible to model -> appear as random errors*
    - *systematic errors and random errors might be well defined in controlled environment.* This is not the case for mobile robots !!

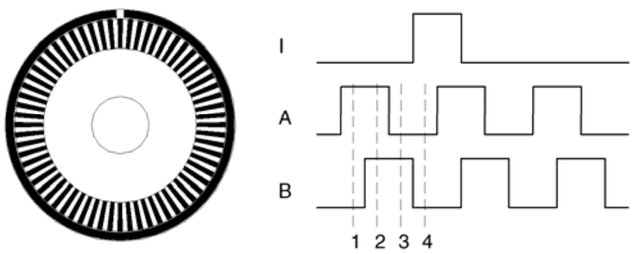
Multi-Modal Error Distributions: The Challenges in …

- Behavior of sensors modeled by probability distribution (random errors)
  - *usually very little knowledge about the causes of random errors*
  - *often probability distribution is assumed to be symmetric or even Gaussian*
  - *however, it is important to realize how wrong this can be!*
  - *Examples:*
    - *Sonar (ultrasonic) sensor might overestimate the distance in real environment and is therefore not symmetric*

      Thus the sonar sensor might be best modeled by two modes:
      - mode for the case that the signal returns directly
      - mode for the case that the signals returns after multi-path reflections.
    - *Stereo vision system might correlate to images incorrectly, thus causing results that make no sense at all*

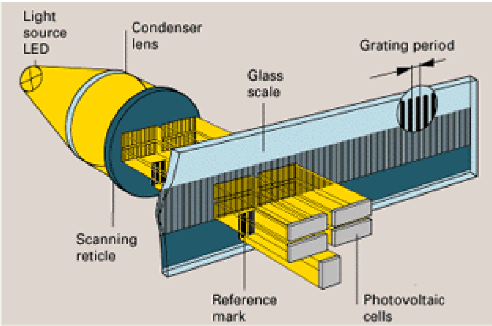
Wheel / Motor Encoders (1)

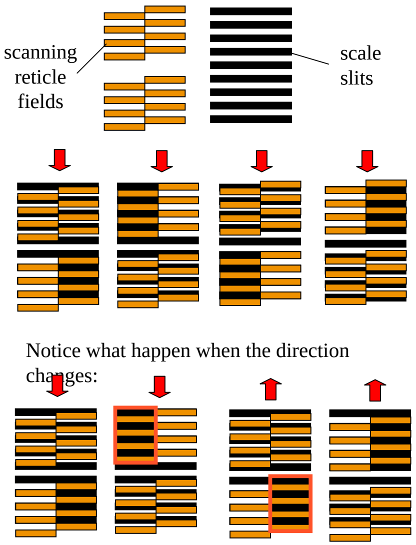
- measure position or speed of the wheels or steering
- wheel movements can be integrated to get an estimate of the robots position -> odometry
- optical encoders are proprioceptive sensors
  - *thus the position estimation in relation to a fixed reference frame is only valuable for short movements.*
- typical resolutions: 2000 increments per revolution.
  - *for high resolution: interpolation*



State	Ch A	Ch B
S_1	High	Low
S_2	High	High
S_3	Low	High
S_4	Low	Low

Wheel / Motor Encoders (2)





Heading Sensors

- Heading sensors can be proprioceptive (gyroscope, inclinometer) or exteroceptive (compass).
- Used to determine the robots orientation and inclination.
- Allow, together with an appropriate velocity information, to integrate the movement to an position estimate.
  - *This procedure is called dead reckoning (ship navigation)*

Compass

- Since over 2000 B.C.
  - *when Chinese suspended a piece of naturally magnetite from a silk thread and used it to guide a chariot over land.*
- Magnetic field on earth
  - *absolute measure for orientation.*
- Large variety of solutions to measure the earth magnetic field
  - *mechanical magnetic compass*
  - *direct measure of the magnetic field (Hall-effect, magnetoresistive sensors)*
- Major drawback
  - *weakness of the earth field*
  - *easily disturbed by magnetic objects or other sources*
  - *not feasible for indoor environments*

Gyroscope

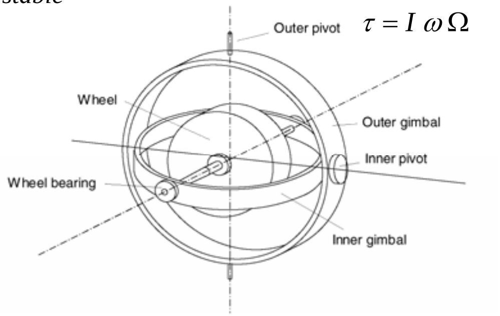
- Heading sensors, that keep the orientation to a fixed frame
  - *absolute measure for the heading of a mobile system.*
- Two categories, the mechanical and the optical gyroscopes
  - *Mechanical Gyroscopes*
    - *Standard gyro*
    - *Rated gyro*
  - *Optical Gyroscopes*
    - *Rated gyro*

Mechanical Gyroscopes

- Concept: inertial properties of a fast spinning rotor
  - *gyroscopic precession*
- Angular momentum associated with a spinning wheel keeps the axis of the gyroscope inertially stable.
- Reactive torque t (tracking stability) is proportional to the spinning speed w, the precession speed W and the wheels inertia I.
- No torque can be transmitted from the outer pivot to the wheel axis
  - *spinning axis will therefore be space-stable*
- Quality: 0.1° in 6 hours

- If the spinning axis is aligned with the north-south meridian, the earth's rotation has no effect on the gyro's horizontal axis
- If it points east-west, the horizontal axis reads the earth rotation

$$\tau = I\,\omega\,\Omega$$



Rate gyros

- Same basic arrangement shown as regular mechanical gyros

- But: gimble(s) are restrained by a torsional spring
  - *enables to measure angular speeds instead of the orientation.*

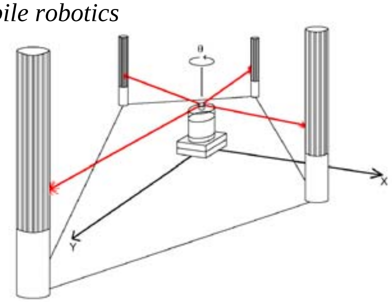
- Others, more simple gyroscopes, use Coriolis forces to measure changes in heading.

Optical Gyroscopes

- First commercial use started only in the early 1980 when they where first installed in airplanes.
- Optical gyroscopes
  - *angular speed (heading) sensors using two monochromic light (or laser) beams from the same source.*
- On is traveling in a fiber clockwise, the other counterclockwise around a cylinder
- Laser beam traveling in direction of rotation
  - *slightly shorter path -> shows a higher frequency*
  - *difference in frequency Δf of the two beams is proportional to the angular velocity Ω of the cylinder*
- New solid-state optical gyroscopes based on the same principle are build using microfabrication technology.

Ground-Based Active and Passive Beacons

- Elegant way to solve the localization problem in mobile robotics
- Beacons are signaling guiding devices with a precisely known position
- Beacon base navigation is used since the humans started to travel
  - *Natural beacons (landmarks) like stars, mountains or the sun*
  - *Artificial beacons like lighthouses*
- The recently introduced Global Positioning System (GPS) revolutionized modern navigation technology
  - *Already one of the key sensors for outdoor mobile robotics*
  - *For indoor robots GPS is not applicable,*
- Major drawback with the use of beacons in indoor:
  - *Beacons require changes in the environment -> costly.*
  - *Limit flexibility and adaptability to changing environments.*

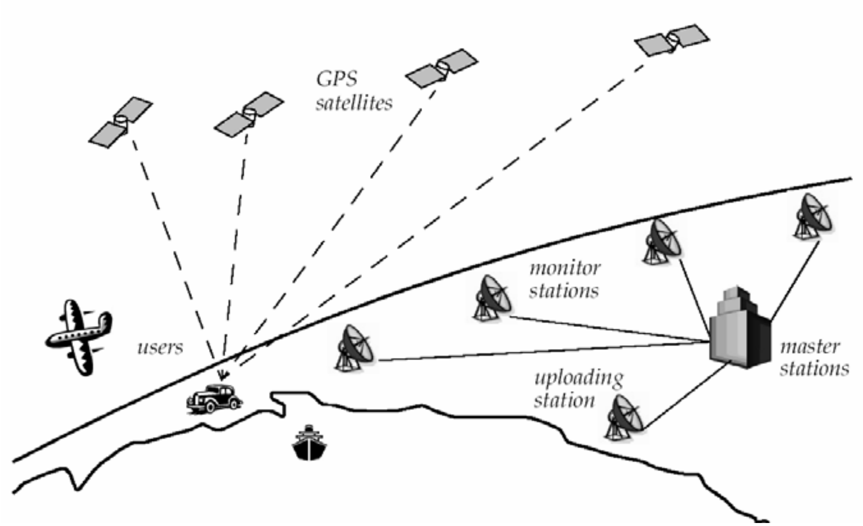


Global Positioning System (GPS) (1)

- *Developed for military use*
- *Recently it became accessible for commercial applications*
- *24 satellites (including three spares) orbiting the earth every 12 hours at a height of 20.190 km.*
- *Four satellites are located in each of six planes inclined 55 degrees with respect to the plane of the earth's equators*
- *Location of any GPS receiver is determined through a time of flight measurement*

- Technical challenges:
  - *Time synchronization between the individual satellites and the GPS receiver*
  - *Real time update of the exact location of the satellites*
  - *Precise measurement of the time of flight*
  - *Interferences with other signals*

Global Positioning System (GPS) (2)



Global Positioning System (GPS) (3)

- Time synchronization:
  - *atomic clocks on each satellite*
  - *monitoring them from different ground stations.*
- Ultra-precision time synchronization is extremely important
  - *electromagnetic radiation propagates at light speed,*
- Roughly 0.3 m per nanosecond.
  - *position accuracy proportional to precision of time measurement.*
- Real time update of the exact location of the satellites:
  - *monitoring the satellites from a number of widely distributed ground stations*
  - *master station analyses all the measurements and transmits the actual position to each of the satellites*
- Exact measurement of the time of flight
  - *the receiver correlates a pseudocode with the same code coming from the satellite*
  - *The delay time for best correlation represents the time of flight.*
  - *quartz clock on the GPS receivers are not very precise*
  - *the range measurement with four satellite*
  - *allows to identify the three values (x, y, z) for the position and the clock correction ΔT*
- Recent commercial GPS receiver devices allows position accuracies down to a couple meters.

Range Sensors (time of flight) (1)

- Large range distance measurement -> called range sensors
- Range information:
  - *key element for localization and environment modeling*
- Ultrasonic sensors as well as laser range sensors make use of propagation speed of sound or electromagnetic waves respectively. The traveled distance of a sound or electromagnetic wave is given by

$$d = c \cdot t$$

- Where
  - *d = distance traveled (usually round-trip)*
  - *c = speed of wave propagation*
  - *t = time of flight.*

Range Sensors (time of flight) (2)

- It is important to point out
  - *Propagation speed v of sound: 0.3 m/ms*
  - *Propagation speed v of of electromagnetic signals: 0.3 m/ns,*
    - *one million times faster.*
  - *3 meters*
    - *is 10 ms ultrasonic system*
    - *only 10 ns for a laser range sensor*
    - *time of flight t with electromagnetic signals is not an easy task*
    - *laser range sensors expensive and delicate*
- The quality of time of flight range sensors manly depends on:
  - *Uncertainties about the exact time of arrival of the reflected signal*
  - *Inaccuracies in the time of fight measure (laser range sensors)*
  - *Opening angle of transmitted beam (ultrasonic range sensors)*
  - *Interaction with the target (surface, specular reflections)*
  - *Variation of propagation speed*
  - *Speed of mobile robot and target (if not at stand still)*

Ultrasonic Sensor (time of flight, sound) (1)

- transmit a packet of (ultrasonic) pressure waves
- distance d of the echoing object can be calculated based on the propagation speed of sound c and the time of flight t.

$$d = \frac{c.t}{2}$$

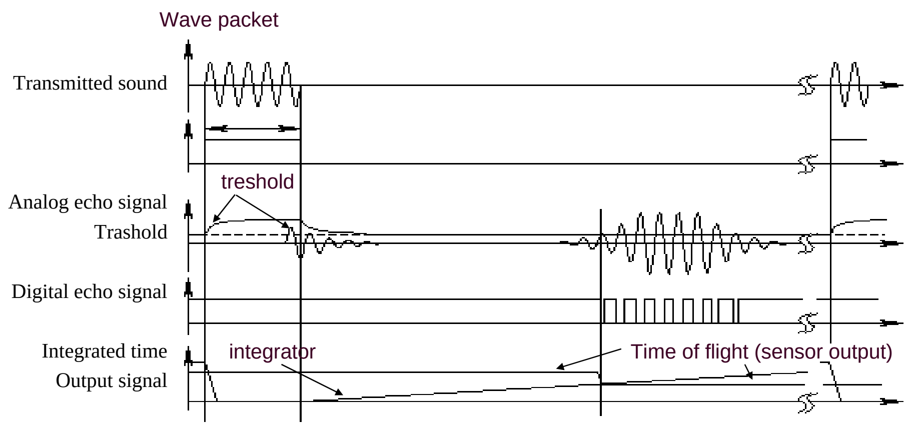
- The speed of sound c (340 m/s) in air is given by

$$c = \sqrt{\gamma . R . T}$$

where

γ: ration of specific heats
R: gas constant
T: temperature in degree Kelvin

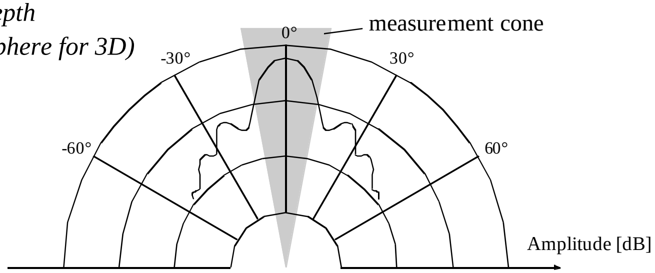
Ultrasonic Sensor (time of flight, sound) (2)



Signals of an ultrasonic sensor

Ultrasonic Sensor (time of flight, sound) (3)

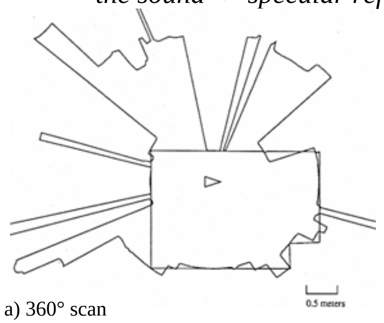
- typically a frequency: 40 - 180 kHz
- generation of sound wave: piezo transducer
  - *transmitter and receiver separated or not separated*
- sound beam propagates in a cone like manner
  - *opening angles around 20 to 40 degrees*
  - *regions of constant depth*
  - *segments of an arc (sphere for 3D)*



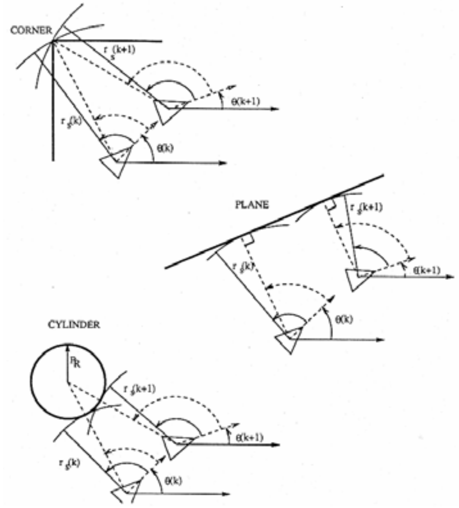
Typical intensity distribution of a ultrasonic sensor

Ultrasonic Sensor (time of flight, sound) (4)

- Other problems for ultrasonic sensors
  - *soft surfaces that absorb most of the sound energy*
  - *surfaces that are fare from being perpendicular to the direction of the sound -> specular reflection*

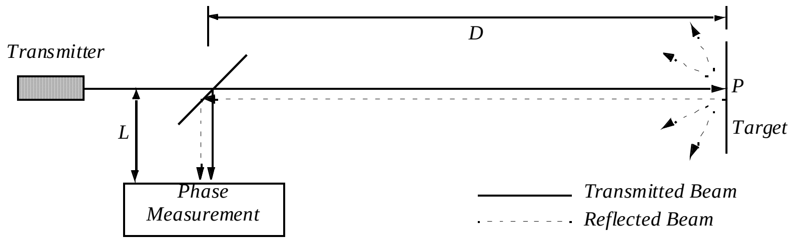


a) 360° scan



b) results from different geometric primitives

Laser Range Sensor (time of flight, electromagnetic) (1)



- Transmitted and received beams coaxial
- Transmitter illuminates a target with a collimated beam
- Received detects the time needed for round-trip
- A mechanical mechanism with a mirror sweeps
  - *2 or 3D measurement*

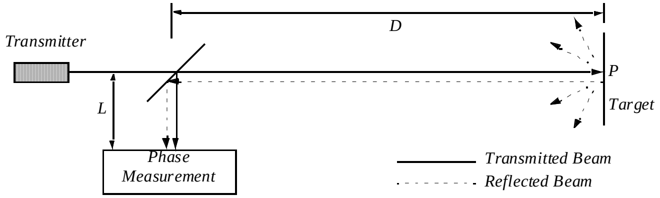
Laser Range Sensor (time of flight, electromagnetic) (2)

Time of flight measurement

- Pulsed laser
  - *measurement of elapsed time directly*
  - *resolving picoseconds*
- Beat frequency between a frequency modulated continuous wave and its received reflection
- Phase shift measurement to produce range estimation
  - *technically easier than the above two methods.*

Laser Range Sensor (time of flight, electromagnetic) (3)

- *Phase-Shift Measurement*



$\lambda = c/f \qquad D' = L + 2D = L + \frac{\theta}{2\pi}\lambda$

Where

c: is the speed of light; *f* the modulating frequency; *D'* covered by the emitted light is

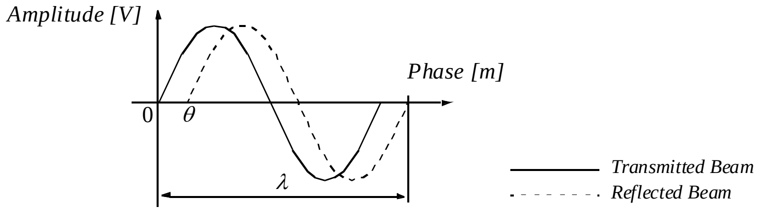
- *for f = 5 Mhz (as in the A.T&T. sensor), λ = 60 meters*

Laser Range Sensor (time of flight, electromagnetic) (4)

- Distance D, between the beam splitter and the target

$$D = \frac{\lambda}{4\pi}\theta \qquad (2.33)$$

- where
  - *θ: phase difference between the transmitted*
- Theoretically ambiguous range estimates
  - *since for example if λ = 60 meters, a target at a range of 5 meters = target at 35 meters*



Laser Range Sensor (time of flight, electromagnetic) (5)

- Confidence in the range (phase estimate) is inversely proportional to the square of the received signal amplitude.

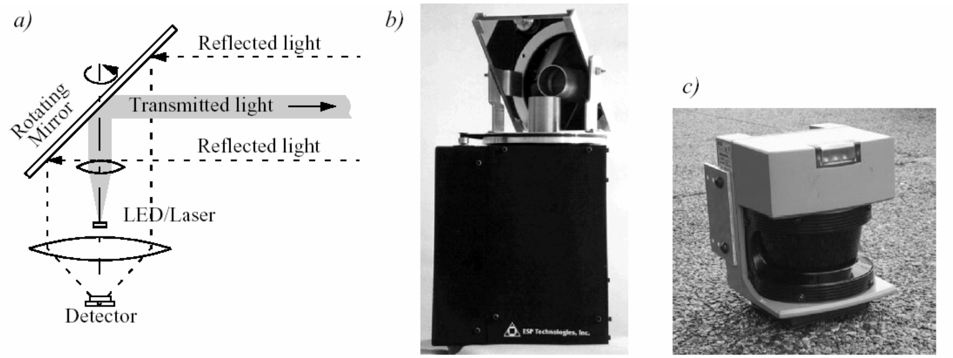
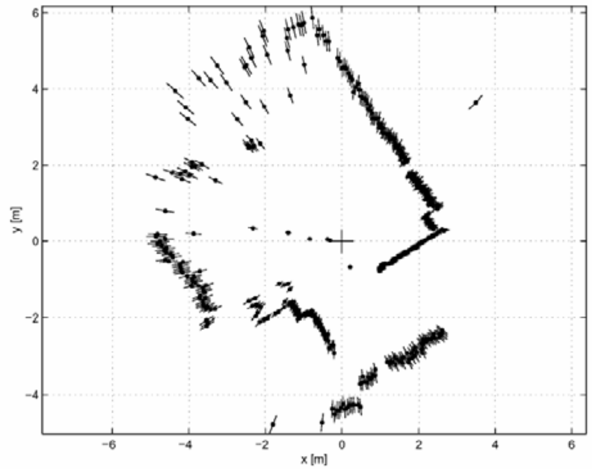


Figure 4.11
(a) Schematic drawing of laser range sensor with rotating mirror; (b) Scanning range sensor from EPS Technologies Inc.; (c) Industrial 180 degree laser range sensor from Sick Inc., Germany

Laser Range Sensor (time of flight, electromagnetic)

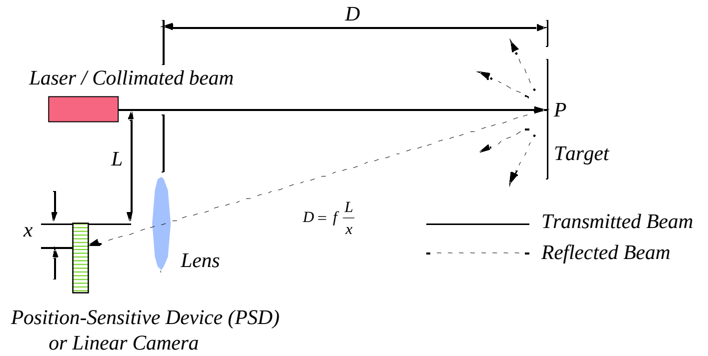
- Typical range image of a 2D laser range sensor with a rotating mirror. The length of the lines through the measurement points indicate the uncertainties.



Triangulation Ranging

- geometrical properties of the image to establish a distance measurement
- e.g. project a well defined light pattern (e.g. point, line) onto the environment.
  - *reflected light is than captured by a photo-sensitive line or matrix (camera) sensor device*
  - *simple triangulation allows to establish a distance.*
- e.g. size of an captured object is precisely known
  - *triangulation without light projecting*

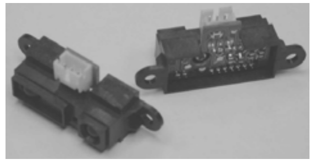
Laser Triangulation (1D)



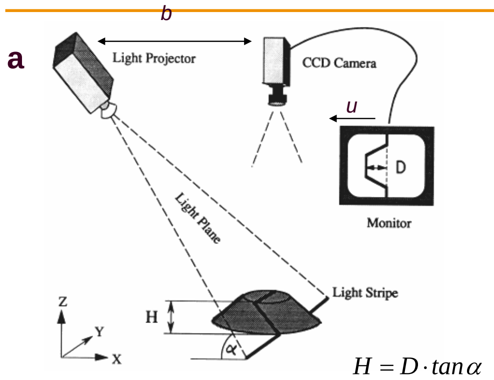
Principle of 1D laser triangulation.

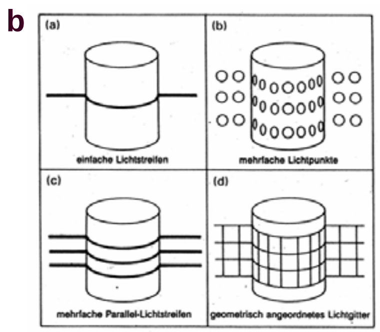
- *distance is proportional to the 1/x*

$$D = f\,\frac{L}{x}$$



Structured Light (vision, 2 or 3D)



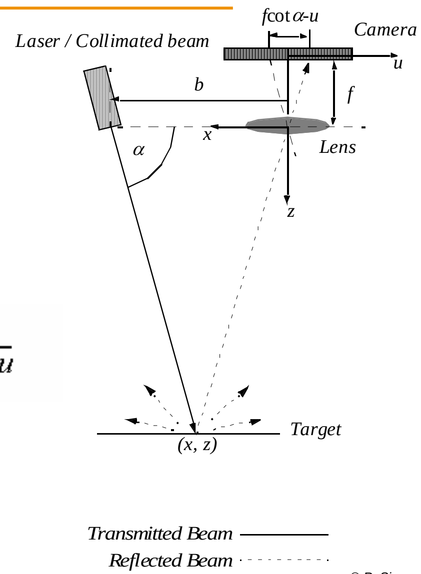


- Eliminate the correspondence problem by projecting structured light on the scene.
- Slits of light or emit collimated light (possibly laser) by means of a rotating mirror.
- Light perceived by camera
- Range to an illuminated point can then be determined from simple geometry.

Structured Light (vision, 2 or 3D)

- One dimensional schematic of the principle
- From the figure, simple geometry shows that:

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



Structured Light (vision, 2 or 3D)

- Range resolution is defined as the triangulation gain G_p :

$$\frac{\partial u}{\partial z} = G_p = \frac{b \cdot f}{z^2}$$

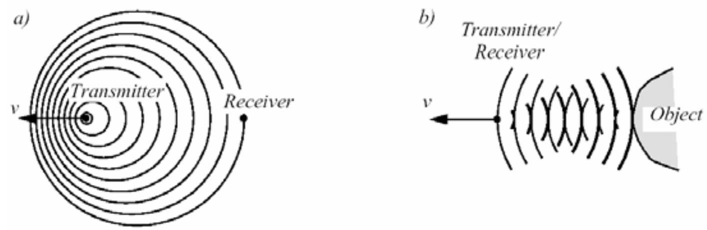
- Influence of α:

$$\frac{\partial \alpha}{\partial z} = G_\alpha = \frac{b \sin \alpha^2}{z^2}$$

- Baseline length b:
  - *the smaller b is the more compact the sensor can be.*
  - *the larger b is the better the range resolution is.*

  Note: for large b, the chance that an illuminated point is not visible to the receiver increases.
- Focal length f:
  - *larger focal length f can provide*
    - *either a larger field of view*
    - *or an improved range resolution*
  - *however, large focal length means a larger sensor head*

Doppler Effect Based (Radar or Sound)



a) between two moving objects   b) between a moving and a stationary object

$f_r = f_t(1 + v/c)$ if transmitter is moving $f_r = f_t \dfrac{1}{1 + v/c}$ if receiver is moving

$\Delta f = f_t - f_r = \dfrac{2 f_t v \cos\theta}{c}$ Doppler frequency shift $v = \dfrac{\Delta f \cdot c}{2 f_t \cos\theta}$ relative speed

- *Sound waves: e.g. industrial process control, security, fish finding, measure of ground speed*
- *Electromagnetic waves: e.g. vibration measurement, radar systems, object tracking*

Vision-based Sensors: Hardware

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

- CMOS (*Complementary Metal Oxide Semiconductor technology*)

Vision in General

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!)

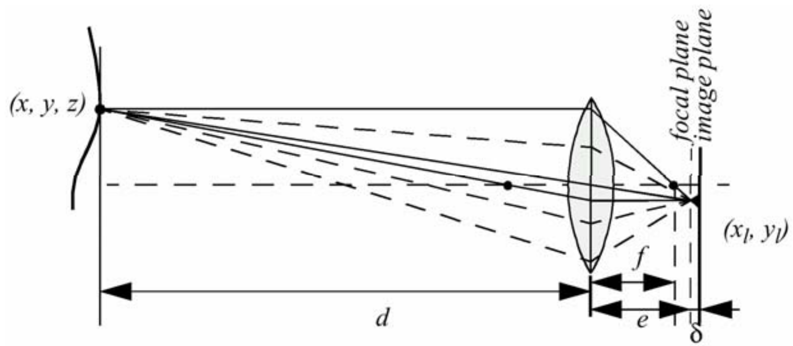
Vision is also our most complicated sense. Whilst we can reconstruct views with high resolution on photographic paper, the next step of understanding how the brain processes the information from our eyes is still in its infancy.

When an image is recorded through a camera, a 3 dimensional scene is projected onto a 2 dimensional plane (the film or a light sensitive photo sensitive array). In order to try and recover some “useful information” from the scene, usually edge detectors are used to find the contours of the objects. From these edges or edge fragments, much research time has to been spent attempting to produce fool proof algorithms which can provide all the necessary information required to reconstruct the 3-D scene which produced the 2-D image. Even in this simple situation, the edge fragments found are not perfect, and will require careful processing if they are to be integrated into a clean line drawing representing the edges of objects. The interpretation of 3-D scenes from 2-D images is not a trivial task. However, using stereo imaging or triangulation methods, vision can become a powerful tool for environment capturing.

Vision-based Sensors: Sensing

- Visual Range Sensors
  - *Depth from focus*
  - *Stereo vision*
- Motion and Optical Flow
- Color Tracking Sensors

Depth from Focus (1)



$$\frac{1}{f} = \frac{1}{d} + \frac{1}{e} \qquad R = \frac{L\delta}{2e}$$

Depth from Focus (2)


- Measure of sub-image gradient

$$sharpness_1 = \sum_{x,y} |I(x,y) - I(x-1,y)|$$

$$sharpness_2 = \sum_{x,y} (I(x,y) - I(x-2,y-2))^2$$

Depth from Focus (3)

- Point Spread Function *h*

$$h(x_g, y_g, x_f, y_f, R_{x,y}) = \begin{bmatrix} \frac{1}{\pi R^2} & if\ ((x_g - x_f)^2 + (y_g - y_f)^2) \le R^2 \\ 0 & if\ ((x_g - x_f)^2 + (y_g - y_f)^2) > R^2 \end{bmatrix}$$

$$g(x_g, y_g) = \sum_{x,y} h(x_g, y_g, x, y, R_{x,y}) f(x,y)$$

Stereo Vision

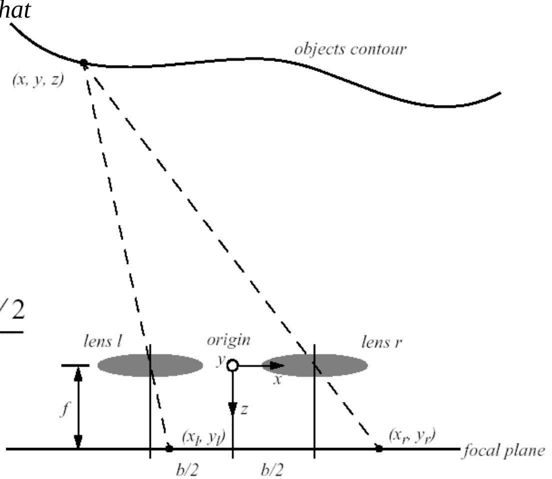
- Idealized camera geometry for stereo vision
  - *Disparity between two images -> Computing of depth*
  - *From the figure it can be seen that*

$$\frac{x_l}{f} = \frac{x + b/2}{z} \quad \text{and} \quad \frac{x_r}{f} = \frac{x - b/2}{z}$$

$$\frac{x_l - x_r}{f} = \frac{b}{z}$$

$$x = b\frac{(x_l + x_r)/2}{x_l - x_r} \; ; \quad y = b\frac{(y_l + y_r)/2}{x_l - x_r}$$

$$z = b\frac{f}{x_l - x_r}$$



Stereo Vision

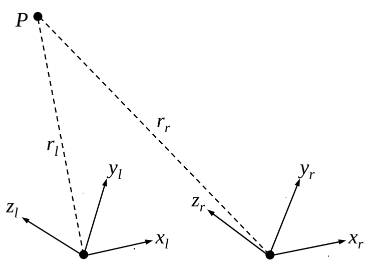
1. Distance is inversely proportional to disparity
   - *closer objects can be measured more accurately*
2. Disparity is proportional to b.
   - *For a given disparity error, the accuracy of the depth estimate increases with increasing baseline b.*
   - *However, as b is increased, some objects may appear in one camera, but not in the other.*
3. A point visible from both cameras produces a conjugate pair.
   - *Conjugate pairs lie on epipolar line (parallel to the x-axis for the arrangement in the figure above)*

Stereo Vision – the general case

- The same point P is measured differently in the left camera image :

$$r'_r = R \cdot r'_l + r_0 \qquad \begin{bmatrix} x'_r \\ y'_r \\ z'_r \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} x'_l \\ y'_l \\ z'_l \end{bmatrix} + \begin{bmatrix} r_{01} \\ r_{02} \\ r_{03} \end{bmatrix}$$

- where
  - R *is a 3 x 3 rotation matrix*
  - r_0 *= offset translation matrix*
- The above equations have two uses:
  - *We can find r_r if we knew R and r_l and r_0. Note: For perfectly aligned cameras R=I (unity matrix)*
  - *We can calibrate the system and find r_{11}, r_{12} … given corresponding values of x_l, y_l, z_l, x_r, y_r and z_r.*
- We have 12 unknowns and require 12 equations:
  - *we require 4 conjugate points for a complete calibration.*
  - *Note: Additionally there is a optical distortion of the image*



left camera coordinate system

right camera coordinate system

Stereo Vision

Calculation of Depth

- *The key problem in stereo is now how do we solve the correspondence problem?*

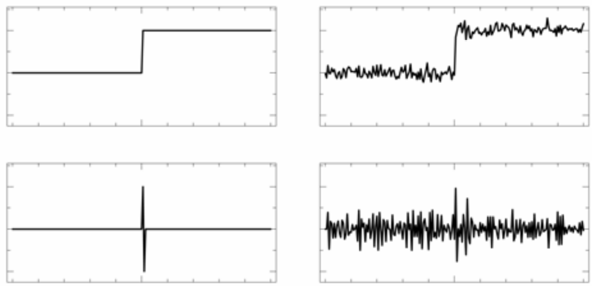
Gray-Level Matching

- *match gray-level wave forms on corresponding epipolar lines*
- *"brightness" = image irradiance I(x,y)*
- *Zero Crossing of Laplacian of Gaussian is a widely used approach for identifying feature in the left and right image*

Zero Crossing of Laplacian of Gaussian

- Identification of features that are stable and match well
- Laplacian of intensity image $$L(x,y) = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2}$$
- Convolution with P: $$L = P \otimes I \qquad P = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$
- Step / Edge Detection in Noisy Image
  - *filtering through Gaussian smoothing*

$$\begin{bmatrix} \frac{1}{16} & \frac{2}{16} & \frac{1}{16} \\ \frac{2}{16} & \frac{4}{16} & \frac{2}{16} \\ \frac{1}{16} & \frac{2}{16} & \frac{1}{16} \end{bmatrix}$$



Stereo Vision Example

- Extracting depth information from a stereo image
  - *a1 and a2: left and right image*
  - *b1 and b2: vertical edge filtered left and right image; filter = [1 2 4 -2 -10 -2 4 2 1]*
  - *c: confidence image: bright = high confidence* (good texture)
  - *d: depth image: bright = close; dark = far*

SVM Stereo Head Mounted on an All-terrain Robot

- Stereo Camera
  - *Videre Design*
  - *www.videredesign.com*
- Robot
  - *Shrimp, EPFL*
- Application of Stereo Vision
  - *Traversability calculation based on stereo images for outdoor navigation*
  - *Motion tracking*

Optical Flow (1)

- *E (x, y, t)* = irradiance at time *t* at the image point (*x*, *y*).
- *u (x, y)* and *v (x, y)* = optical flow vector at that point
  - *find a new image for a point where the irradiance will be the same at time t+ δt*

$$E(x + u\delta t, y + v\delta t, t + \delta t) = E(x, y, t)$$

- If brightness varies smoothly with *x*, *y* and *t* we can expand the left hand side as a Taylor series to obtain:

$$E(x, y, t) + \delta x\frac{\partial E}{\partial x} + \delta y\frac{\partial E}{\partial y} + \delta t\frac{\partial E}{\partial t} + e = E(x, y, t)$$

- *e* = second and higher order terms in δx …
  - *With δt -> 0*

$$\frac{\partial E}{\partial x}\frac{dx}{dt} + \frac{\partial E}{\partial y}\frac{dy}{dt} + \frac{\partial E}{\partial t} = 0 \qquad u = \frac{dx}{dt}\ ; \quad v = \frac{dy}{dt}$$

Optical Flow (2)

- from which we can abbreviate:

$$E_x u + E_y v + E_t = 0$$

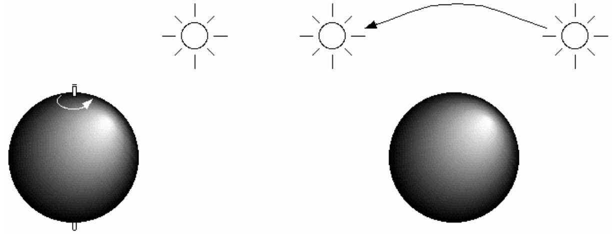
optical flow constraint equation

  - *The derivatives Ex, Ey and Et are estimated from the image.*

- From this equation we can only get the direction of the velocity (u, v) and not unique values for u and v.
  - *One therefore introduces additional constraint, smoothness of optical flow (see lecture notes)*

Problems with Optical Flow

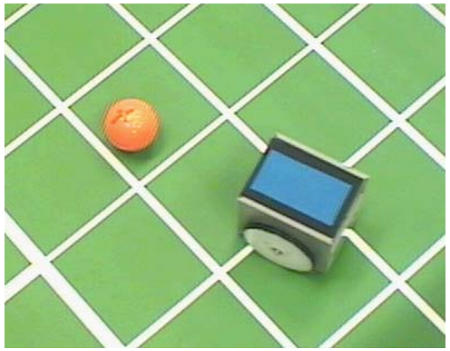
- Motion of the sphere or the light source here demonstrates that optical flow is not always the same as the motion field.

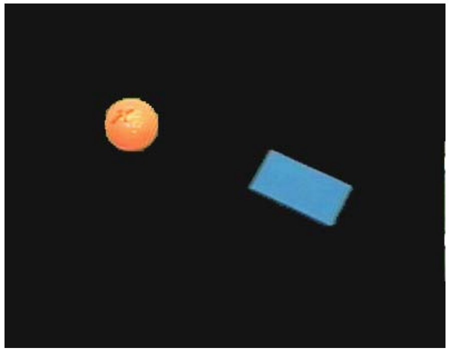


- Left: Discontinuities in Optical Flow
  - *silhouettes (one object occluding another)*
    - *discontinuities in optical flow*
  - *find these points*
    - *stop joining with smooth solution.*
- Right: Motion of sphere, moving light sources

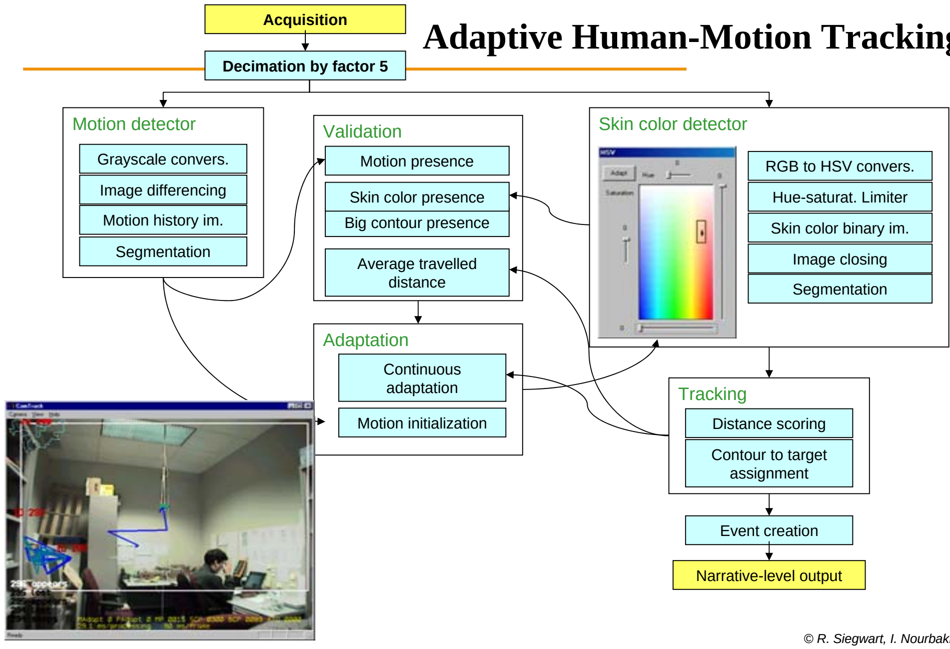
Color Tracking Sensors

- Motion estimation of ball and robot for soccer playing using color tracking





Adaptive Human-Motion Tracking



Adaptive Human-Motion Tracking

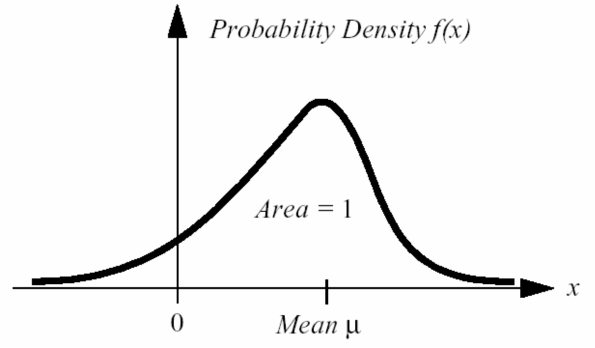
Uncertainty Representation

- Sensing is always related to uncertainties.
  - *What are the sources of uncertainties?*
  - *How can uncertainty be represented or quantified?*
  - *How do they propagate - uncertainty of a function of uncertain values?*
  - *How do uncertainties combine if different sensor reading are fused?*
  - *What is the merit of all this for mobile robotics?*
- Some definitions:
  - *Sensitivity:* *G=out/in*
  - *Resolution:* *Smallest change which can be detected*
  - *Dynamic Range:* *$value_{max}$/ resolution (10^4 -10^6)*
  - *Accuracy:* *$error_{max}$= (measured value) - (true value)*
- Errors are usually unknown:

  deterministic ⟷ *non deterministic (random)*

Uncertainty Representation (2)

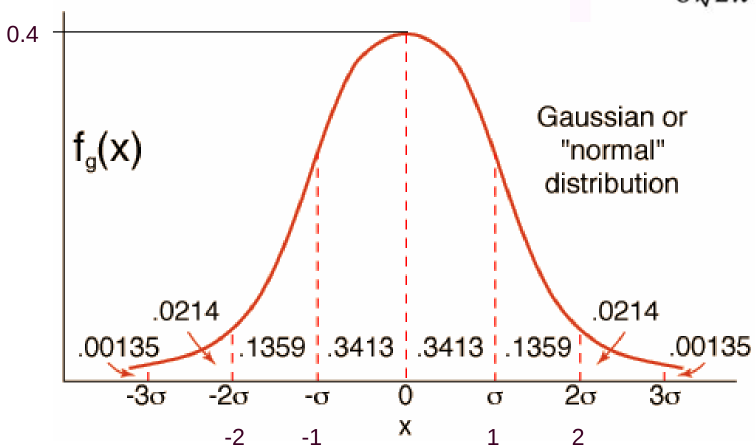
- Statistical representation and independence of random variables on blackboa-d



Gaussian Distribution

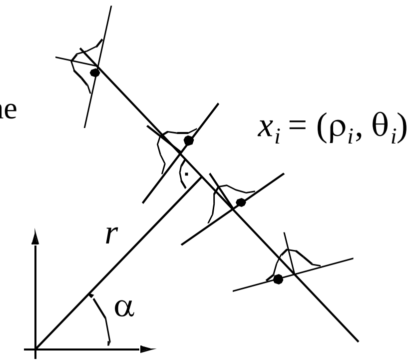
$\mu = 0$ and $\sigma = 1$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$



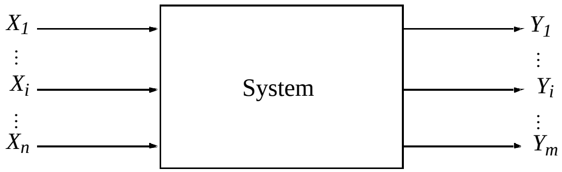
The Error Propagation Law: Motivation

- Imagine extracting a line based on point measurements with uncertainties.
- The model parameters ρ_i (length of the perpendicular) and θ_i (its angle to the abscissa) describe a line uniquely.



- The question:
  - *What is the uncertainty of the extracted line knowing the uncertainties of the measurement points that contribute to it ?*

The Error Propagation Law



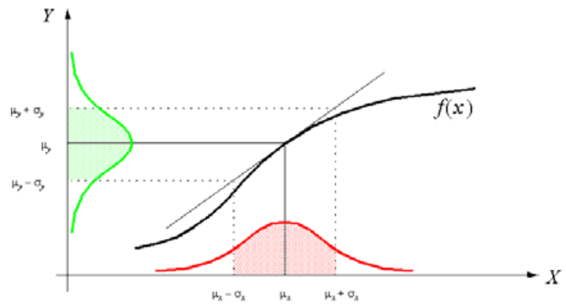
- Error propagation in a multiple-input multi-output system with n inputs and m outputs.

$$Y_j = f_j\,(X_1 \ldots X_n)$$

The Error Propagation Law

- One-dimensional case of a nonlinear error propagation problem
- It can be shown, that the output covariance matrix C_Y is given by the error propagation law:

$$C_Y = F_X C_X F_X^T$$

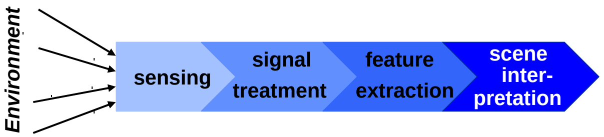


- where
  - ➢ *C_X: covariance matrix representing the input uncertainties*
  - ➢ *C_Y: covariance matrix representing the propagated uncertainties for the outputs.*
  - ➢ *F_X: is the* ***Jacobian*** *matrix defined as:*

$$F_X = \nabla f = \left[\nabla_X \cdot f(X)^T\right]^T = \begin{bmatrix} f_1 \\ \vdots \\ f_m \end{bmatrix}\begin{bmatrix} \frac{\partial}{\partial X_1} & \cdots & \frac{\partial}{\partial X_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial X_1} & \cdots & \frac{\partial f_1}{\partial X_n} \\ \vdots & \cdots & \vdots \\ \frac{\partial f_m}{\partial X_1} & \cdots & \frac{\partial f_m}{\partial X_n} \end{bmatrix}$$

  - ➢ *which is the transposed of the gradient of $f(X)$.*

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 as we have seen in previous section.*
  - ➢ *However, the main difficulty lies in interpreting these data, that is, in deciding what the sensor signals tell us about the environment.*
  - ➢ *Choice of sensors (e.g. in-door, out-door, walls, free space …)*
  - ➢ *Choice of the environment model*

Feature

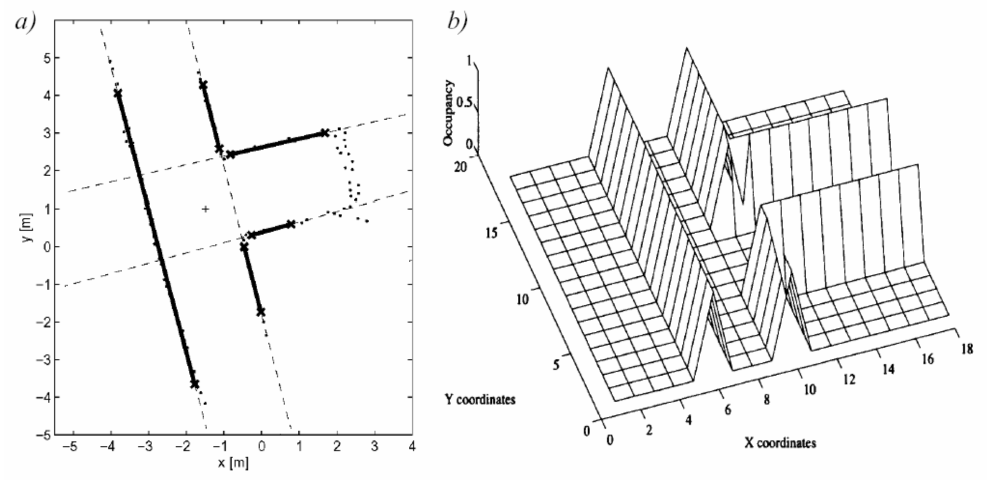
- Features are distinctive elements or geometric primitives of the environment.
- They usually can be extracted from measurements and mathematically described.
  - ➢ *low-level features (geometric primitives) like lines, circles*
  - ➢ *high-level features like edges, doors, tables or trash cans.*

In mobile robotics features help for localization and map building.

Environment Representation and Modeling → Features

- Environment Representation
  - *Continuous Metric* → *x,y,θ*
  - *Discrete Metric* → *metric grid*
  - *Discrete Topological* → *topological grid*
- Environment Modeling
  - *Raw sensor data, e.g. laser range data, grayscale images*
    - *large volume of data, low distinctiveness*
    - *makes use of all acquired information*
  - *Low level features, e.g. line other geometric features*
    - *medium volume of data, average distinctiveness*
    - *filters out the useful information, still ambiguities*
  - *High level features, e.g. doors, a car, the Eiffel tower*
    - *low volume of data, high distinctiveness*
    - *filters out the useful information, few/no ambiguities, not enough information*

Environment Models: Examples



A: Feature base Model

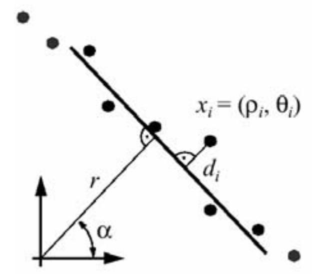
B: Occupancy Grid

Feature extraction base on range images

- Geometric primitives like line segments, circles, corners, edges
- Most other geometric primitives the parametric description of the features becomes already to complex and no closed form solutions exist.
- However, lines segments are very often sufficient to model the environment, especially for indoor applications.

Features Based on Range Data: Line Extraction (1)

$$\rho_i \cos(\theta_i - \alpha) - r = d_i$$



- Least Square

$$S = \sum_i d_i^2 = \sum_i (\rho_i \cos(\theta_i - \alpha) - r)^2$$

$$\frac{\partial S}{\partial \alpha} = 0 \qquad \frac{\partial S}{\partial r} = 0$$

- Weighted Least Square

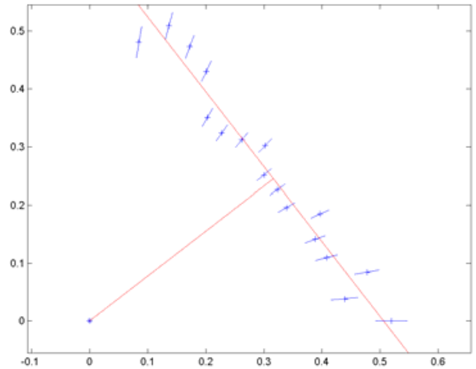
$$w_i = 1/\sigma_i^2$$

$$S = \sum w_i d_i^2 = \sum w_i (\rho_i \cos(\theta_i - \alpha) - r)^2$$

Features Based on Range Data: Line Extraction (2)

- 17 measurement
- error (σ) proportional to ρ^2
- weighted least square:

$$w_i = 1/\sigma_i^2$$



$$\alpha = \frac{1}{2}\operatorname{atan}\left(\frac{\sum w_i\rho_i^2\sin 2\theta_i - \frac{2}{\sum w_i}\sum\sum w_i w_j \rho_i\rho_j\cos\theta_i\sin\theta_j}{\sum w_i\rho_i^2\cos 2\theta_i - \frac{1}{\sum w_i}\sum\sum w_i w_j\rho_i\rho_j\cos(\theta_i+\theta_j)}\right)$$

$$r = \frac{\sum w_i\rho_i\cos(\theta_i-\alpha)}{\sum w_i}$$

Propagation of uncertainty during line extraction

$$C_{AR} = \begin{bmatrix} \sigma_A^2 & \sigma_{AR} \\ \sigma_{AR} & \sigma_R^2 \end{bmatrix} \quad ?$$

(output covariance matrix)

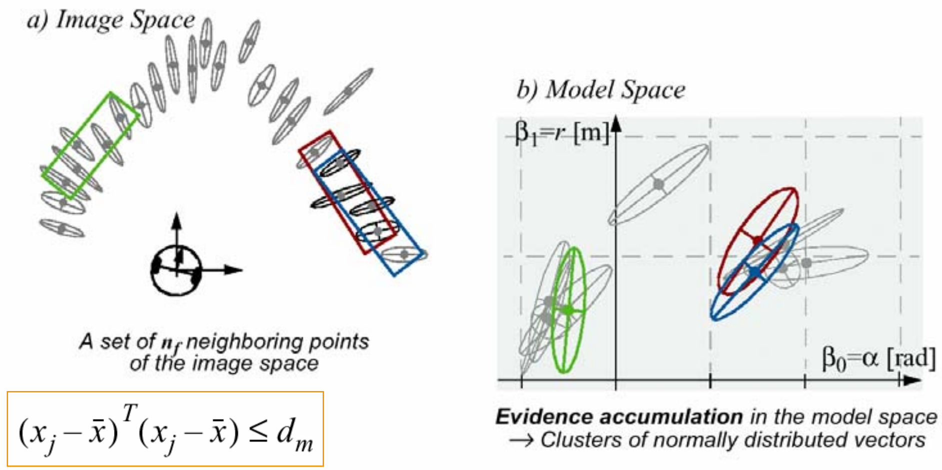
$$C_X = \begin{bmatrix} C_P & \mathbf{0} \\ \mathbf{0} & C_Q \end{bmatrix} = \begin{bmatrix} diag(\sigma_{\rho_i}^2) & \mathbf{0} \\ \mathbf{0} & diag(\sigma_{\theta_i}^2) \end{bmatrix} \qquad 2n \times 2n$$

Jacobian:

$$F_{PQ} = \begin{bmatrix} \frac{\partial\alpha}{\partial P_1} & \frac{\partial\alpha}{\partial P_2} & \cdots & \frac{\partial\alpha}{\partial P_n} & \frac{\partial\alpha}{\partial Q_1} & \frac{\partial\alpha}{\partial Q_2} & \cdots & \frac{\partial\alpha}{\partial Q_n} \\ \frac{\partial r}{\partial P_1} & \frac{\partial r}{\partial P_2} & \cdots & \frac{\partial r}{\partial P_n} & \frac{\partial r}{\partial Q_1} & \frac{\partial r}{\partial Q_2} & \cdots & \frac{\partial r}{\partial Q_n} \end{bmatrix}$$

$$C_{AR} = F_{PQ} C_X F_{PQ}^T$$

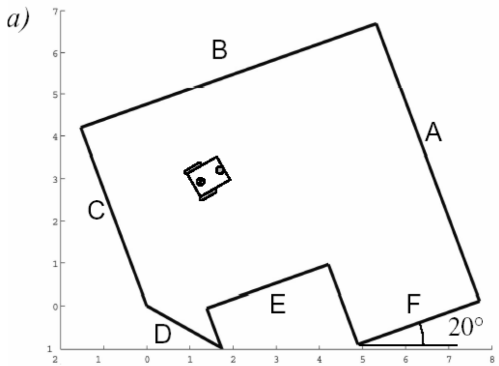
Segmentation for Line Extraction

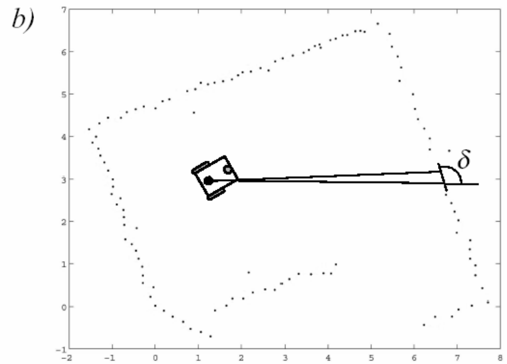


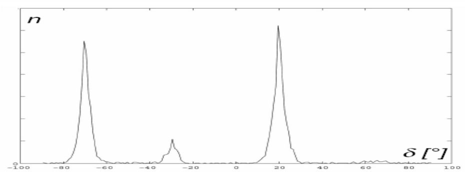
$$\boxed{(x_j - \bar{x})^T (x_j - \bar{x}) \le d_m}$$

Fig 4.36 Clustering: Finding neighboring segments of a common line

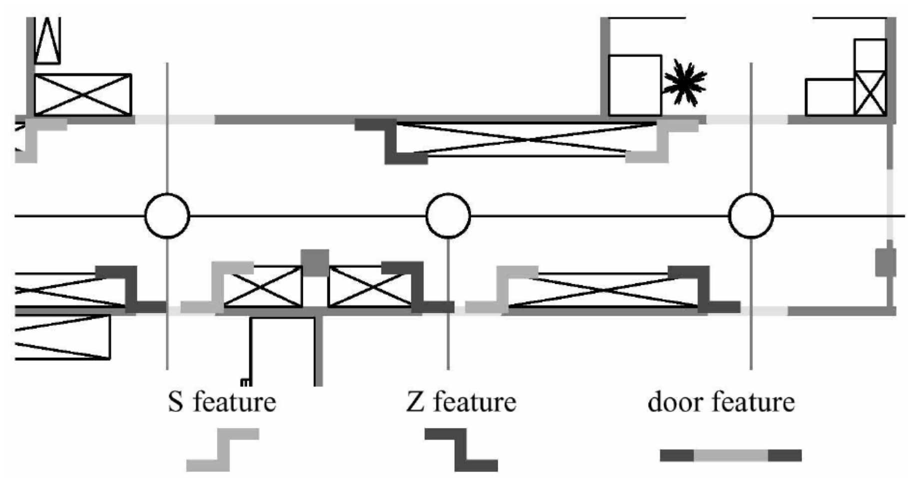
Angular Histogram (range)







Extracting Other Geometric Features

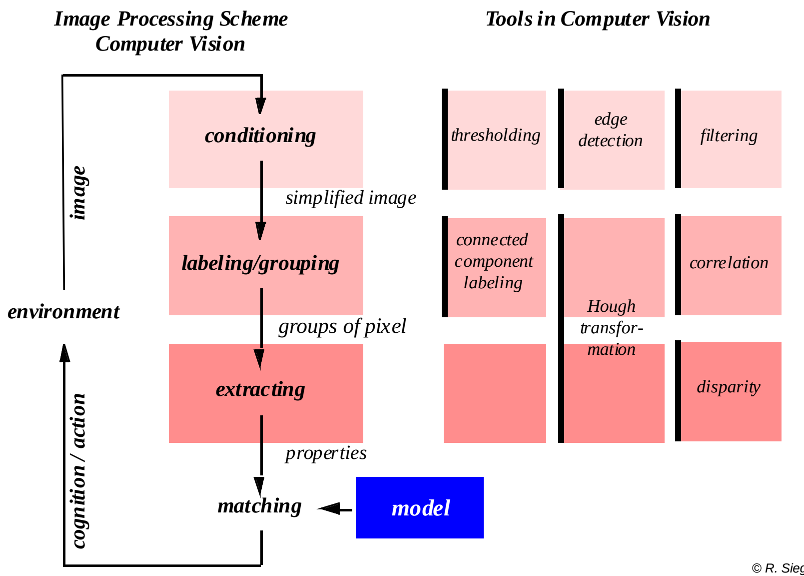


Feature extraction

Scheme and tools in computer vision

- Recognition of features is, in general, a complex procedure requiring a variety of steps that successively transform the iconic data to recognition information.
- Handling unconstrained environments is still very challenging problem.

Visual Appearance-base Feature Extraction (Vision)



Feature Extraction (Vision): Tools

- Conditioning
  - *Suppresses noise*
  - *Background normalization by suppressing uninteresting systematic or patterned variations*
  - *Done by:*
    - *gray-scale modification (e.g. trasholding)*
    - *(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 pixel participating in the same kind of event*
- Extracting
  - *Compute a list of properties for each group*
- Matching (see chapter 5)

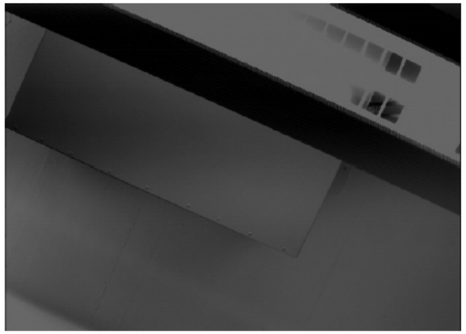
Filtering and Edge Detection

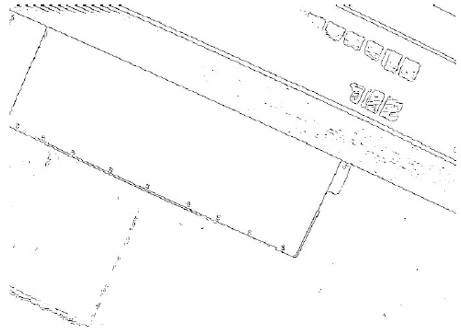
- Gaussian Smoothing
  - *Removes high-frequency noise*
  - *Convolution of intensity image I with G:* $\hat{I} = G \otimes I$

    with: $G = \frac{1}{16}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$
- 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.*
  - *Noise, thus first filtering/smoothing required before edge detection*

Edge Detection

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



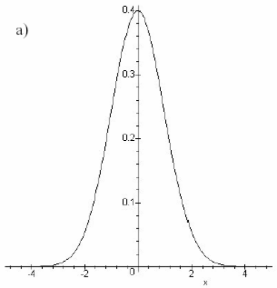


Optimal Edge Detection: Canny

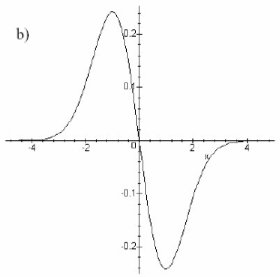
- The processing steps
  - *Convolution of image with the Gaussian function G*
  - *Finding maxima in the derivative*
- Canny combines both in one operation

$$(G \otimes I)' = G' \otimes I$$

(a) A Gaussian

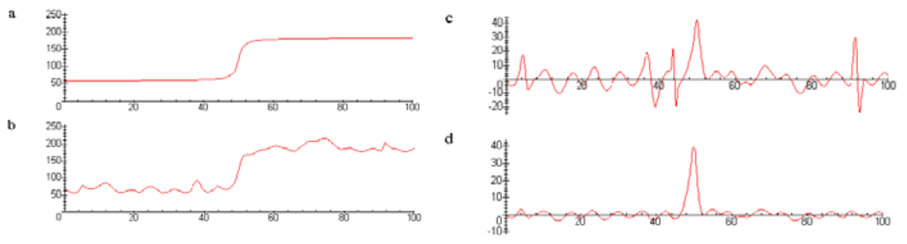


$$G_\sigma(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$$



$$G_\sigma'(x) = \frac{-x}{\sqrt{2\pi}\sigma^3} e^{-\frac{x^2}{2\sigma^2}}$$

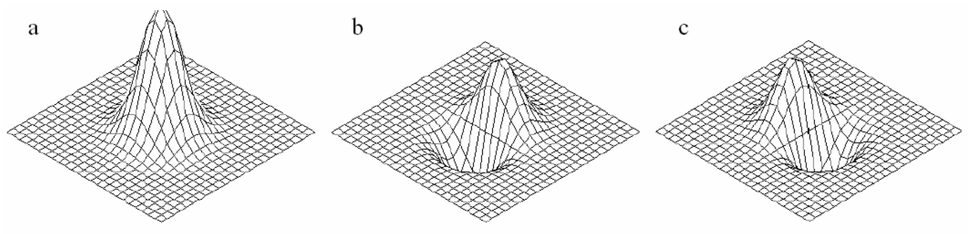
Optimal Edge Detection: Canny 1D example



(a) Intensity 1-D profile of an ideal step edge.
(b) Intensity profile *I(x)* of a real edge.
(c) Its derivative *I'(x)*.
(d) The result of the convolution *R(x) = G'* ⊗ *I*, where *G'* is the first derivative of a Gaussian function.

Optimal Edge Detection: Canny

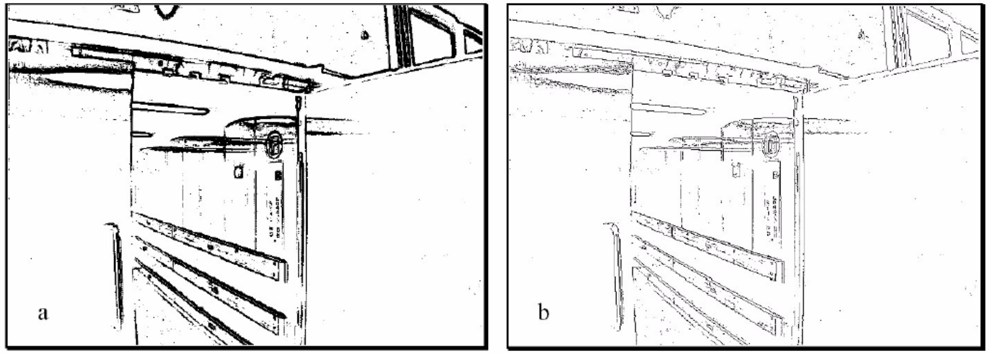
- 1-D edge detector can be defined with the following steps:
  1. *Convolute the image* I *with* G' *to obtain* R.
  2. *Find the absolute value of* R.
  3. *Mark those peaks* |R| *that are above some predefined threshold* T. *The threshold is chosen to eliminate spurious peaks due to noise.*
- 2D → Two dimensional Gaussian function



$$G_\sigma(x, y) = G_\sigma(x)G_\sigma(y) \qquad f_V(x, y) = G'_\sigma(x)G_\sigma(y) \qquad f_H(x, y) = G'_\sigma(y)G_\sigma(x)$$

Optimal Edge Detection: Canny Example

a) Example of Canny edge detection
b) After nonmaxima suppression



Gradient Edge Detectors

- **Roberts**

$$|G| \cong \sqrt{r_1^2 + r_2^2} \; ; \quad r_1 = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \; ; \quad r_2 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

- **Prewitt**

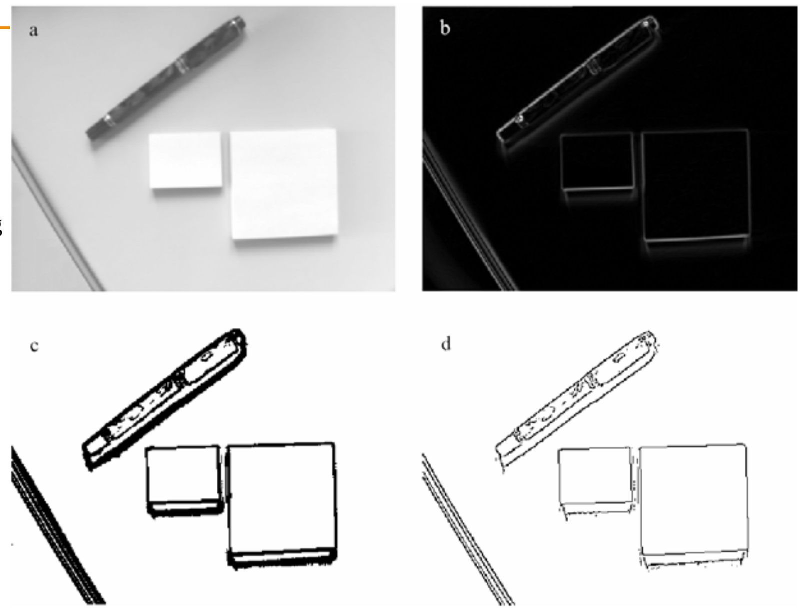
$$|G| \cong \sqrt{p_1^2 + p_2^2} \; ; \quad \theta \cong \operatorname{atan}\left(\frac{p_1}{p_2}\right) \; ; \quad p_1 = \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} \; ; \quad p_2 = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix}$$

- **Sobel**

$$|G| \cong \sqrt{s_1^2 + s_2^2} \; ; \quad \theta \cong \operatorname{atan}\left(\frac{s_1}{s_2}\right) \; ; \quad s_1 = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \; ; \quad s_2 = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

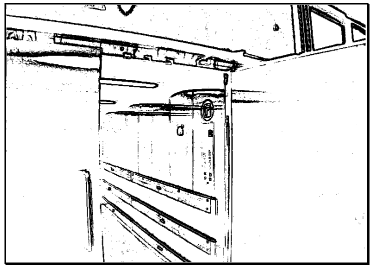
Example

a) Raw image
b) Filtered (Sobel)
c) Thresholding
d) Nonmaxima suppression

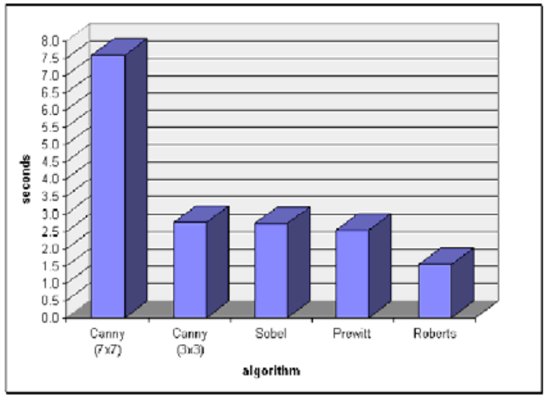


Nonmaxima Suppression

- Output of an edge detector is usually a b/w image where the pixels with gradient magnitude above a predefined threshold are black and all the others are white
- Nonmaxima suppression generates contours described with only one pixel thinness



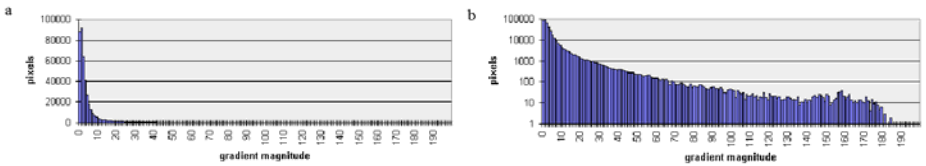
Comparison of Edge Detection Methods



- Average time required to compute the edge figure of a 780 x 560 pixels image.
- The times required to compute an edge image are proportional with the accuracy of the resulting edge images

Dynamic Thresholding

- Changing illumination
  - *Constant threshold level in edge detection is not suitable*
- Dynamically adapt the threshold level
  - *consider only the n pixels with the highest gradient magnitude for further calculation steps.*

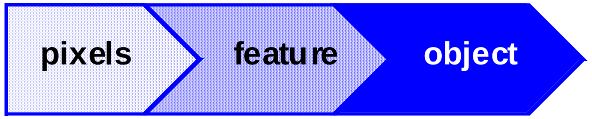


(a) Number of pixels with a specific gradient magnitude in the image of Figure 1.2(b).
(b) Same as (a), but with logarithmic scale

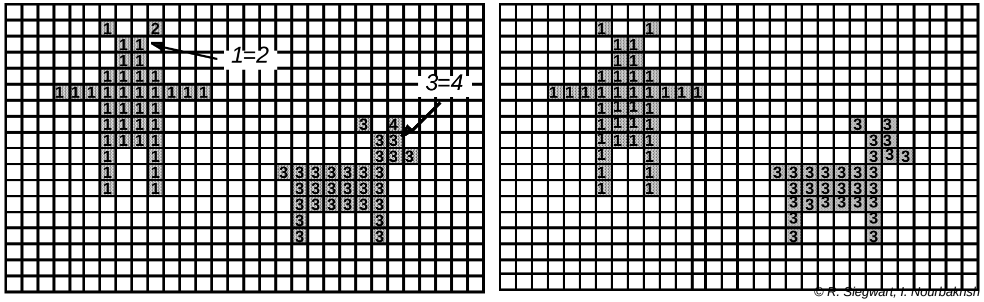
Hough Transform: Straight Edge Extraction

- All points p on a straight-line edge must satisfy $y_p = m_1 x_p + b_1$.
- Each point (x_p, y_p) that is part of this line constraints the parameter m_1 and b_1.
- The Hough transform finds the line (line-parameters m, b) that get most "votes" from the edge pixels in the image.
- This is realized by four stepts
  1. *Create a 2D array A [m,b] with axes that tessellate the values of m and b.*
  2. *Initialize the array A to zero.*
  3. *For each edge pixel (x_p, y_p) in the image, loop over all values of m and b: if $y_p = m_1 x_p + b_1$ then $A[m,b]+=1$*
  4. *Search cells in A with largest value. They correspond to extracted straight-line edge in the image.*

Grouping, Clustering: Assigning Features to Features



- Connected Component Labeling



Floor Plane Extraction

- Vision based identification of traversable
- The processing steps
  - *As pre-processing, smooth I_f using a Gaussian smoothing operator*
  - *Initialize a histogram array H with* n *intensity values:* *for* $H[i] = 0$ for $i = 1, \ldots, n$
  - *For every pixel (x,y) in I_f increment the histogram:* $H[I_f(x, y)] += 1$

Whole-Image Features

- OmniCam

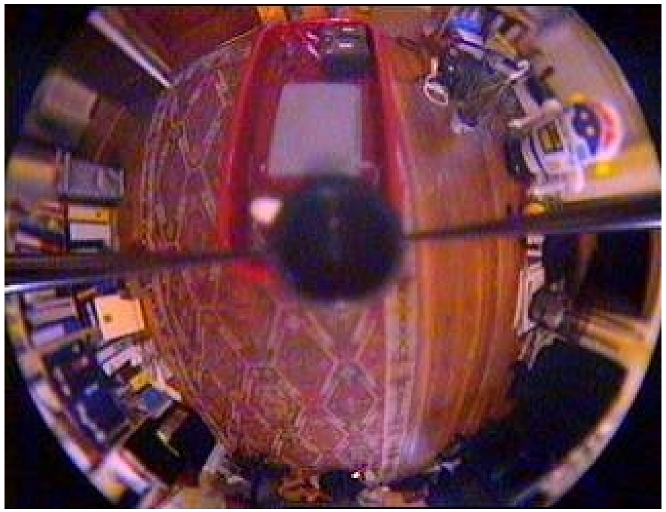


Image Histograms

- The processing steps
  - *As pre-processing, smooth G_i using a Gaussian smoothing operator*
  - *Initialize H_i with n levels:* $H[j] = 0$ for $j = 1, \ldots, n$
  - *For every pixel (x,y) in G_i increment the histogram:* $H_i[G_i[x, y]] += 1$

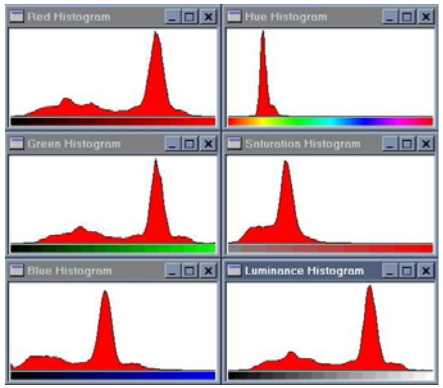


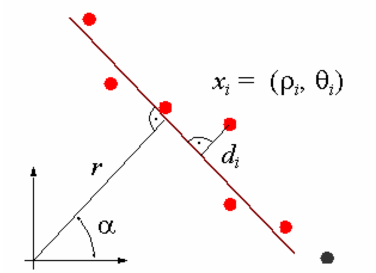
Image Fingerprint Extraction

- Highly distinctive combination of simple features

Example:

Probabilistic Line Extraction from Noisy 1D Range Data

- Suppose:
  - *the segmentation problem has already been solved,*
  - *regression equations for the model fit to the points have a closed-form solution – which is the case when fitting straight lines.*
  - *that the measurement uncertainties of the data points are known*



Line Extraction

- Estimating a line in the least squares sense. The model parameters (length of the perpendicular) and (its angle to the abscissa) describe uniquely a line.

- n measurement points in polar coordinates $x_i = (\rho_i, \theta_i)$

- modeled as random variables $X_i = (P_i, Q_i)$

- Each point is independently affected by Gaussian noise in both coordinates.

$$P_i \sim N(\rho_i, \sigma^2_{\rho_i}) \quad (2.49)$$

$$Q_i \sim N(\theta_i, \sigma^2_{\theta_i}) \quad (2.50)$$

Line Extraction

- Task: find the line

$$x\cos\alpha + y\sin\alpha - r = 0$$

where $x = \rho\cos(\theta)$; $y = \rho\sin(\theta)$

-> $\rho\cos\theta\cos\alpha + \rho\sin\theta\sin\alpha - r = 0$

$$\rho\cos(\theta - \alpha) - r = 0 \quad (2.51)$$

- This model minimizes the *orthogonal* distances d_i of a point (ρ_i, θ_i) to the line

$$\rho_i\cos(\theta_i - \alpha) - r = d_i \quad (2.52)$$

- Let *S* be the (unweighted) sum of squared errors.

$$S = \sum_i d_i^2 = \sum_i (\rho_i\cos(\theta_i - \alpha) - r)^2 \quad (2.53)$$

Line Extraction

- The model parameters (α, r) are now found by solving the nonlinear equation system

$$\frac{\partial S}{\partial \alpha} = 0 \qquad \frac{\partial S}{\partial r} = 0. \tag{2.54}$$

- Suppose each point a known variance σ_i^2 modelling the uncertainty in radial and angular.
  - ➢ *variance is used to determine a weight w_i for each single point, e.g.*

$$w_i = 1/\sigma_i^2. \tag{2.55}$$

- Then, equation (2.53) becomes

$$S = \sum w_i d_i^2 = \sum w_i (\rho_i \cos(\theta_i - \alpha) - r)^2. \tag{2.56}$$

Line Extraction

- It can be shown that the solution of (2.54) in the weighted least square sense is

$$\alpha = \frac{1}{2}\operatorname{atan2}\left(\frac{\sum w_i \rho_i^2 \sin 2\theta_i - \frac{2}{\sum w_i}\sum\sum w_i w_j \rho_i \rho_j \cos\theta_i \sin\theta_j}{\sum w_i \rho_i^2 \cos 2\theta_i - \frac{1}{\sum w_i}\sum\sum w_i w_j \rho_i \rho_j \cos(\theta_i + \theta_j)}\right) \tag{2.57}$$

$$r = \frac{\sum w_i \rho_i \cos(\theta_i - \alpha)}{\sum w_i}. \tag{2.58}$$

- How the uncertainties of the measurements propagate through 'the system' (eq. 2.57, 2.58)?

Line Extraction→ Error Propagation Law

$$C_{AR} = F_{PQ} C_X F_{PQ}^T, \tag{2.59}$$

- given the ***2n* x *2n*** input covariance matrix:

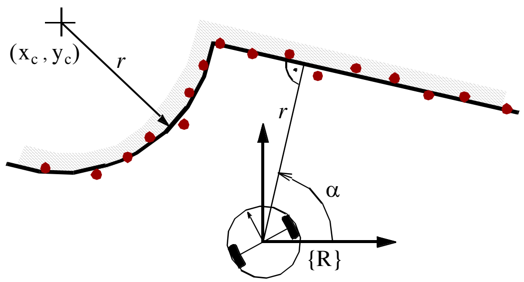
$$C_X = \begin{bmatrix} C_P & \mathbf{0} \\ \mathbf{0} & C_Q \end{bmatrix} = \begin{bmatrix} diag(\sigma_{\rho_i}^2) & \mathbf{0} \\ \mathbf{0} & diag(\sigma_{\theta_i}^2) \end{bmatrix} \tag{2.60}$$

- and the system relationships (2.57) and (2.58). Then by calculating the Jacobian

$$F_{PQ} = \begin{bmatrix} \frac{\partial\alpha}{\partial P_1} & \frac{\partial\alpha}{\partial P_2} & \cdots & \frac{\partial\alpha}{\partial P_n} & \frac{\partial\alpha}{\partial Q_1} & \frac{\partial\alpha}{\partial Q_2} & \cdots & \frac{\partial\alpha}{\partial Q_n} \\ \frac{\partial r}{\partial P_1} & \frac{\partial r}{\partial P_2} & \cdots & \frac{\partial r}{\partial P_n} & \frac{\partial r}{\partial Q_1} & \frac{\partial r}{\partial Q_2} & \cdots & \frac{\partial r}{\partial Q_n} \end{bmatrix} \tag{2.61}$$

- we can instantly form the error propagation equation () yielding the sought C_{AR}

Feature Extraction: The Simplest Case – Linear Regression



Linear Regression

$Y_i = \alpha + \beta X_i + \varepsilon_i$

ε_i : in *y* direction ~$N(0,\sigma^2)$

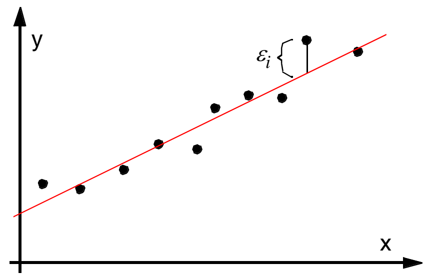
-> α, β such that $\Sigma\varepsilon_i^2$ is minimal.

Model for *straight lines:*

$$x\cos(\alpha) + y\sin(\alpha) - r = 0$$

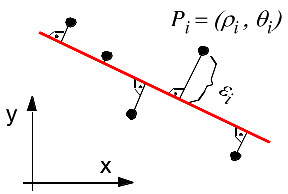
Model for *circles:*

$$(x + x_c)^2 + (y + y_c)^2 - r^2 = 0$$



Feature Extraction: Nonlinear Linear Regression

1) For **straight lines**



$$x_i \cos(\alpha) + y_i \sin(\alpha) - r = \varepsilon_i$$

in order that $\sum_i \varepsilon_i^2$ is minimal:

$$\frac{\partial}{\partial\alpha}\sum_i (x_i \cos(\alpha) + y_i \sin(\alpha) - r)^2 = 0$$

$$\frac{\partial}{\partial r}\sum_i (x_i \cos(\alpha) + y_i \sin(\alpha) - r)^2 = 0$$

A nonlinear equation system $\rightarrow \alpha, r$

A measure of the estimate's **uncertainty: covariance matrix**

$$C_{\alpha r} = \begin{bmatrix} \sigma_{\alpha\alpha} & \sigma_{\alpha r} \\ \sigma_{\alpha r} & \sigma_{rr} \end{bmatrix}$$

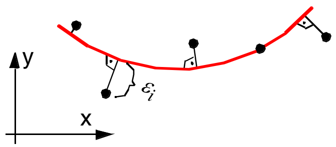
$$\sigma_{\alpha\alpha} \approx \sum_i \left[\frac{\partial\alpha}{\partial\rho_i}(\underline{\rho})\right]^2 \sigma_i^2$$

$$\sigma_{rr} \approx \sum_i \left[\frac{\partial r}{\partial\rho_i}(\underline{\rho})\right]^2 \sigma_i^2$$

$$\sigma_{\alpha r} = \mathbf{COV}[A, R]$$

Apply **error propagation law**

2) For **circles**



$$(x_i + x_c)^2 + (y_i + y_c)^2 - r^2 = \varepsilon_i$$

Feature Extraction / Sensory 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 as we have seen in previous section.
- However, the main difficulty lies in interpreting these data, that is, in deciding what the sensor signals tell us about the environment.
  - *Choice of sensors (e.g. in-door, out-door, walls, free space …)*
  - *choice of the environment model*