Robot Perception

Noppanut Thongton











Robot perception

• One of the most important tasks of an autonomous system of any kind is to acquire **knowledge** about its environment. This is done by taking **measurements** using various sensors and then extracting meaningful information from those measurements.











The mobile robot's sensors.











Sensor classification

- **Proprioceptive** sensors *measure values internal to the system* (robot); for example, motor speed, wheel load, robot arm joint angles, battery voltage.
- Exteroceptive sensors acquire information from the robot's environment; for example, distance measurements, light intensity, sound amplitude. Hence exteroceptive sensor measurements are interpreted by the robot in order to extract meaningful environmental features.











Sensor classification

- Passive sensors measure ambient environmental energy entering the sensor. Examples of passive sensors include temperature probes, microphones, and CCD or CMOS cameras.
- Active sensors *emit energy into the environment*, then measure the environmental reaction. Because active sensors can manage more controlled interactions with the environment, they often achieve superior performance. However, active sensing introduces several risks: the outbound energy may affect the very characteristics that the sensor is attempting to measure.











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	P P 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
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
Active ranging (reflectivity, time-of-flight, and geo- metric triangulation)	Reflectivity sensors Ultrasonic sensor Laser rangefinder Optical triangulation (1D) Structured light (2D)	EC EC EC EC EC	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











Basic sensor response ratings

- The dynamic range is the ratio of the maximum input value to the minimum measurable input value. Because this raw ratio can be unwieldy, it is usually measured in decibels.
- Range is also an important rating in mobile robot applications because often robot sensors operate in environments where they are frequently exposed to input values beyond their working range.
- It is critical to understand how the sensor will respond.











Basic sensor response ratings

- *Resolution* is the minimum difference between two values that can be detected by a sensor
- *Linearity* is an important measure governing the behavior of the sensor's output signal as the input signal varies.
- Bandwidth or frequency is used to measure the speed with which a sensor can provide a stream of readings.











In situ sensor performance

- Sensitivity is the ratio of output change to input change.
- *Cross-sensitivity* is the technical term for sensitivity to environmental parameters that are orthogonal to the target parameters for the sensor
- **Error** of a sensor is defined as the difference between the sensor's output measurements and the true values being measured, within some specific operating context. Given a true value v and a measured value m, we can define error as

$$error = m - v$$











In situ sensor performance

• **Accuracy** is defined as the degree of conformity between the sensor's measurement and the true value, and is often expressed as a proportion of the true value (e.g., 97.5% accuracy). Thus small error corresponds to high accuracy and vice versa:

$$accuracy = 1 - \frac{|error|}{v}$$

• **Systematic errors** are caused by factors or processes that can in theory be modeled.











In situ sensor performance

- **Random errors** cannot be predicted using a sophisticated model nor can they be mitigated by more precise sensor machinery.
- **Precision** is often confused with accuracy, and now we have the tools to clearly distinguish these two terms. Intuitively, high precision relates to reproducibility of the sensor results.

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











Wheel/motor sensors

Wheel/motor sensors are devices used to measure the internal state and dynamics of a mobile robot. These sensors have vast applications outside of mobile robotics and, as a result, mobile robotics has enjoyed the benefits of high-quality, low-cost wheel and motor sensors that offer excellent resolution







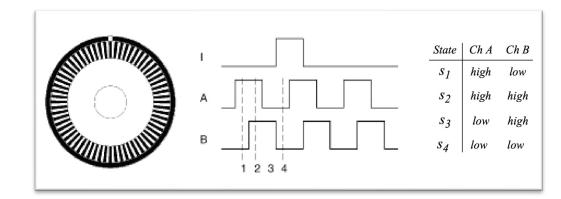




Wheel/motor sensors



Optical incremental encoders have become the most popular device for measuring angular speed and position within a motor drive or at the shaft of a wheel or steering mechanism. In mobile robotics, encoders are used to control the position or speed of wheels and other motor-driven joints. In robotics, the resulting sine wave is transformed into a discrete square wave using a threshold to choose between *light* and *dark* states. Resolution is measured in cycles per revolution (CPR).













Heading sensors can be *proprioceptive* (gyroscope, inclinometer) or *exteroceptive* (compass). They are used to determine the **robot's orientation and inclination**. They allow us, together with appropriate velocity information, to integrate the movement to a position estimate.











Compasses

The two most common modern sensors for measuring the direction of a magnetic field are **the Hall effect** and **flux gate compasses**.

- The Hall effect describes the behavior of electric potential in a semiconductor when in the
 presence of a magnetic field. When a constant current is applied across the length of a
 semiconductor, there will be a voltage difference in the perpendicular direction, across the
 semiconductor's width, based on the relative orientation of the semiconductor to magnetic
 flux lines. In addition, the sign of the voltage potential identifies the direction of the magnetic
 field.
- The flux gate compass operates on a different principle. Two small coils are wound on ferrite
 cores and are fixed perpendicular to one another. When alternating current is activated in
 both coils, the magnetic field causes shifts in the phase depending on its relative alignment
 with each coil. By measuring both phase shifts, the direction of the magnetic field in two
 dimensions can be computed.











Gyroscopes

Gyroscopes are heading sensors which preserve their **orientation in relation to a fixed reference frame**. Thus, they provide an **absolute measure** for the heading of a mobile system. Gyroscopes can be classified in two categories, **mechanical gyroscopes** and **optical gyroscopes**.

- Mechanical gyroscopes. The concept of a mechanical gyroscope relies on the inertial properties of a fast-spinning rotor.
- **Optical gyroscopes.** Optical gyroscopes are a relatively new innovation. Commercial use began in the early 1980s when they were first installed in **aircraft**. Optical gyroscopes are angular speed sensors that use two monochromatic light beams, or lasers, emitted from the same source, instead of moving, mechanical parts.



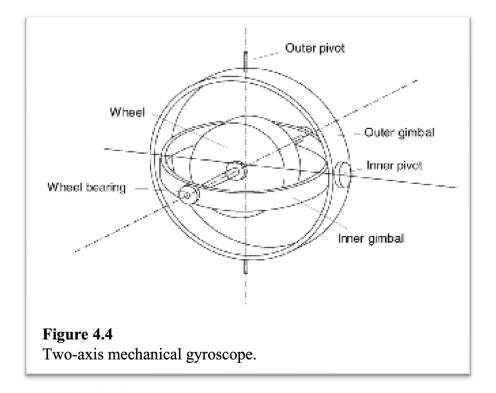








Gyroscopes













Ground-based beacons

One elegant approach to solving the **localization problem** in mobile robotics is to use *active* or *passive* beacons. Using the interaction of on-board sensors and the environmental beacons, the robot can **identify its position** precisely. Although the general intuition is identical to that of early human navigation beacons, such as stars, mountains, and lighthouses, modern technology has enabled sensors to localize an outdoor robot with accuracies of better than 5 cm within areas that are kilometers in size.











Ground-based beacons

The global positioning system

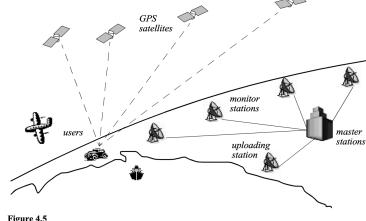


Figure 4.5
Calculation of position and heading based on GPS.

The global positioning system (GPS) was initially developed for military use but is now freely available for civilian navigation. There are at least twenty-four operational GPS satellites at all times. The satellites orbit 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 equator. Each satellite continuously transmits data that indicate its location and the current time. Therefore, GPS receivers are completely passive but exteroceptive sensors. The GPS satellites synchronize their transmissions so that their signals are sent at the same time. When a GPS receiver reads the transmission of two or more satellites, the arrival time differences inform the receiver as to its relative distance to each satellite. By combining information regarding the arrival time and instantaneous location of four satellites, the receiver can infer its own position.











Active ranging

Active ranging sensors continue to be the most popular sensors in mobile robotics. Many ranging sensors have a low price point, and, most importantly, all ranging sensors provide easily interpreted outputs: direct measurements of distance from the robot to objects in its vicinity. For obstacle detection and avoidance, most mobile robots rely heavily on active ranging sensors.











Active ranging

Time-of-flight active ranging

Time-of-flight ranging makes use of the propagation speed of sound or an electromagnetic wave. In general, the travel distance of a sound of 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.











Active ranging

• The ultrasonic sensor (time-of-flight, sound)

The basic principle of an ultrasonic sensor is to transmit a packet of (ultrasonic) pressure waves and to measure the time it takes for this wave packet to reflect and return to the receiver. The distance d of the object causing the reflection can be calculated based on the propagation speed of sound c and the time-of-flight t.

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

$$c = \sqrt{\gamma Rt}$$

where

 γ = ratio of specific heats;

R = gas constant;

T = temperature in degrees Kelvin.





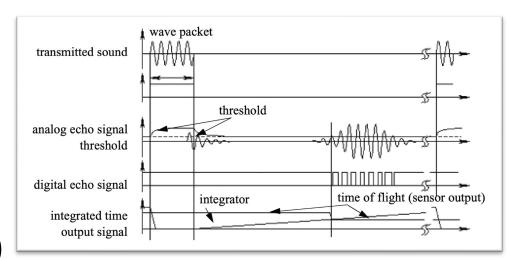






Active ranging

The ultrasonic sensor (time-of-flight, sound)



The figure shows the different signal output and input of an ultrasonic sensor. First, a series of sound pulses are emitted, comprising the wave packet. An integrator also begins to linearly climb in value, measuring the time from the transmission of these sound waves to detection of an echo. A threshold value is set for triggering an incoming sound wave as a valid echo. This threshold is often decreasing in time, because the amplitude of the expected echo decreases over time based on dispersal as it travels longer. But during transmission of the initial sound pulses and just afterward, the threshold is set very high to suppress triggering the echo detector with the outgoing sound pulses. A transducer will continue to ring for up to several milliseconds after the initial transmission, and this governs the blanking time of the sensor.





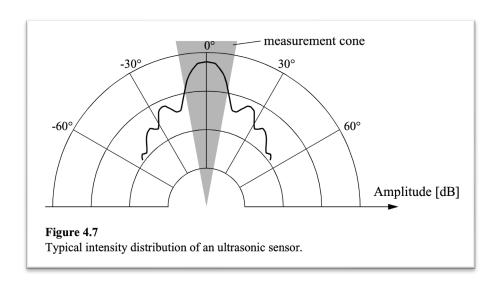






Active ranging

The ultrasonic sensor (time-of-flight, sound)



The ultrasonic wave typically has a frequency **between 40 and 180 kHz** and is usually generated by a piezo or electrostatic transducer. Most ultrasonic sensors used by mobile robots have an effective range of roughly 12 cm to 5 m. In mobile robot applications, specific implementations generally achieve a **resolution of approximately 2 cm**. In most cases one may want a narrow opening angle for the sound beam in order to also obtain precise directional information about objects that are encountered. This is a major limitation since sound propagates in a cone-like manner with opening angles around 20 to 40 degrees.





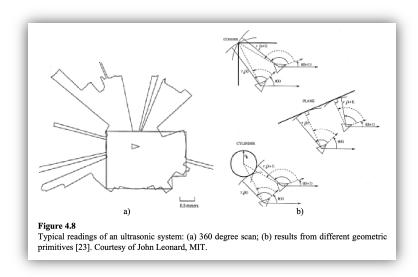






Active ranging

The ultrasonic sensor (time-of-flight, sound)



The sensor readings must be plotted as segments of an arc (sphere for 3D) and not as point measurements. However, recent research developments show significant improvement of the measurement quality in using sophisticated echo processing.











Active ranging

The ultrasonic sensor (time-of-flight, sound)

A final limitation of ultrasonic ranging relates to **bandwidth**. Particularly in moderately open spaces, a single ultrasonic sensor has a relatively slow cycle time. For example, measuring the distance to an object that is 3 m away will take such a sensor 20 ms, limiting its operating speed to **50 Hz**.











Active ranging

Laser rangefinder (time-of-flight, electromagnetic)

The laser rangefinder is a time-of-flight sensor that achieves significant improvements over the ultrasonic range sensor owing to the use of **laser light instead of sound**. This type of sensor consists of a transmitter which illuminates a target with a collimated beam (e.g., laser), and a receiver capable of detecting the component of light which is essentially coaxial with the transmitted beam. Often referred to as **optical radar or lidar (light detection and ranging)**, these devices produce a range estimate based on the time needed for the light to reach the target and return. A mechanical mechanism with a mirror sweeps the light beam to cover the required scene in a plane or even in three dimensions, using a rotating, nodding mirror.











Active ranging

Laser rangefinder (time-of-flight, electromagnetic)

One way to measure the time of flight for the light beam is to **use a pulsed laser** and then **measure the elapsed time** directly, just as in the ultrasonic solution described earlier. Electronics capable of resolving picoseconds are required in such devices and they are therefore very expensive. A second method is to measure the beat frequency between a frequency-modulated continuous wave (FMCW) and its received reflection. Another, even easier method is to measure the phase shift of the reflected light.











Transmitter Phase Measurement Phase Measurement Phase Measurement Phase Measurement Phase Measurement

Figure 4.9 Schematic of laser rangefinding by phase-shift measurement.

Active ranging

Phase-shift measurement

Near-infrared light (from a light-emitting diode [LED] or laser) is collimated and transmitted from the transmitter in this figure and hits a point P in the environment. For surfaces having a roughness greater than the wavelength of the incident light, diffuse reflection will occur, meaning that the light is reflected almost isotropically. The wavelength of the infrared light emitted is 824 nm and so most surfaces, with the exception of only highly polished reflecting objects, will be diffuse reflectors. The component of the infrared light which falls within the receiving aperture of the sensor will return almost parallel to the transmitted beam for distant objects.











Active ranging

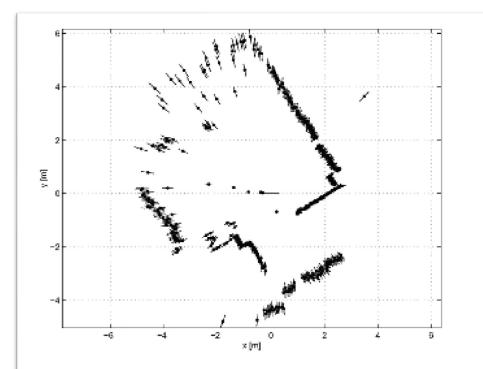


Figure 4.12
Typical range image of a 2D laser range sensor with a rotating mirror. The length of the lines through the measurement points indicate the uncertainties.











Active ranging

Triangulation-based active ranging

Triangulation-based ranging sensors use geometric properties manifest in their measuring strategy to establish distance readings to objects. The simplest class of triangulation-based rangers are active because they project a known light pattern (e.g., a point, a line, or a texture) onto the environment. The reflection of the known pattern is captured by a receiver and, together with known geometric values, the system can use simple triangulation to establish range measurements. If the receiver measures the position of the reflection along a single axis, we call the sensor an optical triangulation sensor in 1D. If the receiver measures the position of the reflection along two orthogonal axes, we call the sensor a structured light sensor. These two sensor types are described in the two sections below.





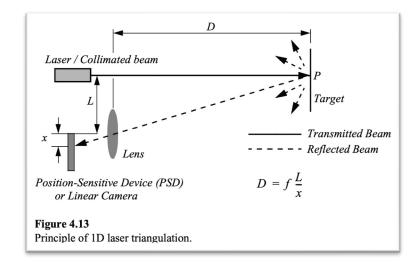






Active ranging





The principle of optical triangulation in 1D is straightforward, as depicted in this figure. A collimated beam (e.g., focused infrared LED, laser beam) is transmitted toward the target. The reflected light is collected by a lens and projected onto a position-sensitive device (PSD) or linear camera. Given the geometry of the figure, the distance is given by

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

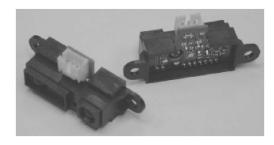


Figure 4.14

A commercially available, low-cost optical triangulation sensor: the Sharp GP series infrared rangefinders provide either analog or digital distance measures and cost only about \$ 15.











Active ranging

Structured light (2D sensor)

If one replaces the linear camera or PSD of an optical triangulation sensor with a 2D receiver such as a CCD or CMOS camera, then one can recover distance to a large set of points instead of to only one point. The emitter must project a known pattern, or structured light, onto the environment.





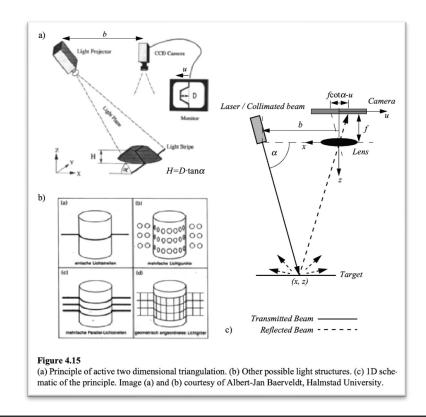






Active ranging

Structured light (2D sensor)









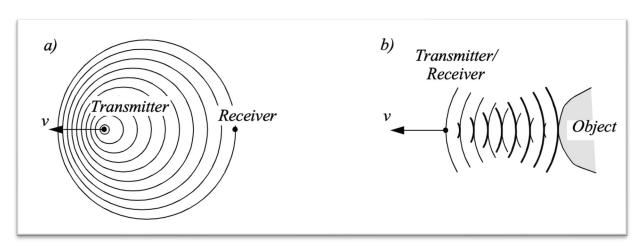




Motion/speed sensors

Doppler effect-based sensing (radar or sound)

A transmitter emits an electromagnetic or sound wave with a frequency. It is either received by a receiver (figure a) or reflected from an object (figure b).













Motion/speed sensors

Doppler effect-based sensing (radar or sound)

The measured frequency at the receiver is a function of the relative speed between transmitter and receiver according to

$$f_r = f_t \frac{1}{1 + v/c}$$

if the transmitter is moving and

$$f_r = f_t (1 + v/c)$$

if the receiver is moving.





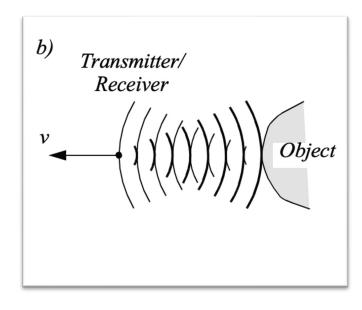






Motion/speed sensors

Doppler effect-based sensing (radar or sound)



In the case of a reflected wave (figure b) there is a factor of 2 introduced, since any change x in relative separation affects the round-trip path length by 2x. Furthermore, in such situations it is generally more convenient to consider the change in frequency Δf , known as the *Doppler shift*, as opposed to the *Doppler frequency* notation above.

$$\Delta f = f_t - f_r = \frac{2f_t v \cos \theta}{c}$$

$$v = \frac{\Delta f \cdot c}{2f_t \cos \theta}$$

where

 Δf = Doppler frequency shift;

 θ = relative angle between direction of motion and beam axis.











Vision-based sensors

Vision is our most powerful sense. It provides us with an enormous amount of information about the environment and enables rich, intelligent interaction in **dynamic environments**. It is therefore not surprising that a great deal of effort has been devoted to providing machines with sensors that mimic the capabilities of the human vision system. The first step in this process is the creation of sensing devices that capture the same raw information light that the human vision system uses.











Vision-based sensors

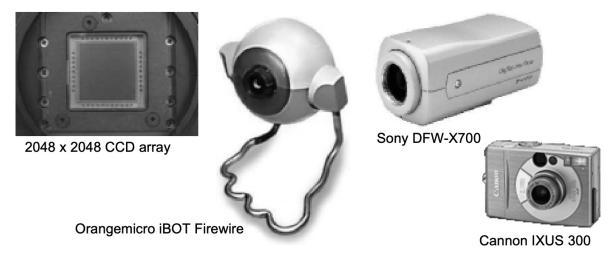


Figure 4.17

Commercially available CCD chips and CCD cameras. Because this technology is relatively mature, cameras are available in widely varying forms and costs (http://www.howstuffworks.com/digital-camera2.htm).











Vision-based sensors

Motion and optical flow

A great deal of information can be recovered by recording time-varying images from a fixed (or moving) camera. First, we distinguish between the motion field and optical flow

- Motion field: this assigns a velocity vector to every point in an image. If a point in the environment moves with velocity v_0 , then this induces a velocity v_i in the image plane. It is possible to determine mathematically the relationship between v_0 and v_i .
- Optical flow: it can also be true that brightness patterns in the image move as the object that causes them moves (light source). Optical flow is the apparent motion of these brightness patterns.











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

Vision-based sensors

Optical flow

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

$$\frac{\partial E}{\partial x}\frac{dx}{dt} + \frac{\partial E}{\partial y}\frac{dy}{dt} + \frac{\partial E}{\partial t} = 0$$

$$u = \frac{dx}{dt}$$
; $v = \frac{dy}{dt}$ $E_x = \frac{\partial E}{\partial x}$; $E_y = \frac{\partial E}{\partial y}$; $E_t = \frac{\partial E}{\partial t} = 0$

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











Vision-based sensors

Color-tracking sensors

Color represents an **environmental characteristic** that is orthogonal to range, and it represents both a natural cue and an artificial cue that can provide new information to a mobile robot. For example, the annual robot soccer events make extensive use of color both for environmental marking and for **robot localization**.









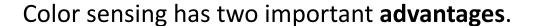






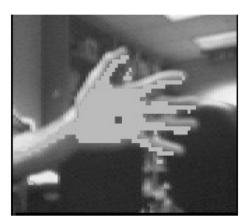
Vision-based sensors

Color-tracking sensors



- 1. Detection of color is a straightforward function of a single image; therefore, no correspondence problem need be solved in such algorithms.
- 2. Because color sensing provides a new, independent environmental cue, if it is **combined** (i.e., *sensor fusion*) **with existing cues**, such as data from stereo vision or laser range finding, we can expect significant information gains.















An autonomous mobile robot must be able to determine its relationship to the environment by **making measurements with its sensors** and then using those measured signals. A wide variety of sensing technologies are available, as shown in the previous section. But every sensor we have presented is imperfect: measurements always have error and, therefore, uncertainty associated with them. Therefore, sensor inputs must be used in a way that enables the robot to interact with its environment successfully in spite of measurement uncertainty.

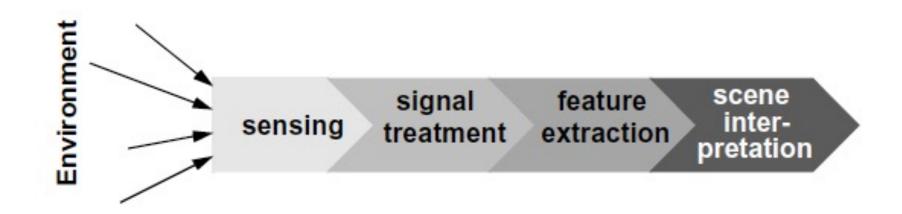






















In practical terms, mobile robots do not necessarily use feature extraction and scene interpretation for every activity. Instead, robots will interpret sensors to varying degrees depending on each specific functionality. For example, in order to guarantee emergency stops in the face of immediate obstacles, the robot may make direct use of raw forward-facing range readings to stop its drive motors. For local obstacle avoidance, raw ranging sensor strikes may be combined in an occupancy grid model, enabling smooth avoidance of obstacles meters away. For map-building and precise navigation, the range sensor values and even vision sensor measurements may pass through the complete perceptual pipeline, being subjected to feature extraction followed by scene interpretation to minimize the impact of individual sensor uncertainty on the robustness of the robot's mapmaking and navigation skills. The pattern that thus emerges is that, as one moves into more sophisticated, long-term perceptual tasks, the feature extraction and scene interpretation aspects of the perceptual pipeline become essential.











Feature definition

Features are recognizable structures of elements in the environment. They usually can be extracted from measurements and mathematically described. **Good features** are always **perceivable** and **easily** detectable from the environment.

• Low-level features (geometric primitives) like lines, circles, or polygons.

Low-level features are abstractions of raw data, and as such provide a lower volume of data while increasing the distinctiveness of each feature. The hope, when one incorporates low-level features, is that the features are filtering out poor or useless data, but of course it is also likely that some valid information will be lost as a result of the feature extraction process.

• **High-level features** (objects) such as *edges, doors, tables, or a trash can*.

High-level features provide maximum abstraction from the raw data, thereby reducing the volume of data as much as possible while providing highly distinctive resulting features. Once again, the abstraction process has the risk of filtering away important information, potentially lowering data utilization.











Feature definition

In mobile robotics, features play an especially **important role** in the creation of **environmental models**. They enable more compact and robust descriptions of the environment, helping a mobile robot during both **map-building and localization**. When designing a mobile robot, a critical decision revolves around choosing the appropriate features for the robot to use.











Target environment

For geometric features to be useful, the target geometries must be readily detected in the actual environment. For example, line features are extremely useful in office building environments due to the abundance of straight wall segments, while the same features are virtually useless when navigating Mars.

Available sensors

Obviously, the specific sensors and sensor uncertainty of the robot impacts the appropriateness of various features. Armed with a laser rangefinder, a robot is well qualified to use geometrically detailed features such as corner features owing to the high-quality angular and depth resolution of the laser scanner. In contrast, a sonar-equipped robot may not have the appropriate tools for corner feature extraction.











Computational power

Vision-based feature extraction can effect a significant computational cost, particularly in robots where the vision sensor processing is performed by one of the robot's main processors.

Environment representation

Feature extraction is an important step toward scene interpretation, and by this token the features extracted must provide information that is consonant with the representation used for the environmental model. For example, nongeometric vision-based features are of little value in purely geometric environmental models but can be of great value in topological models of the environment



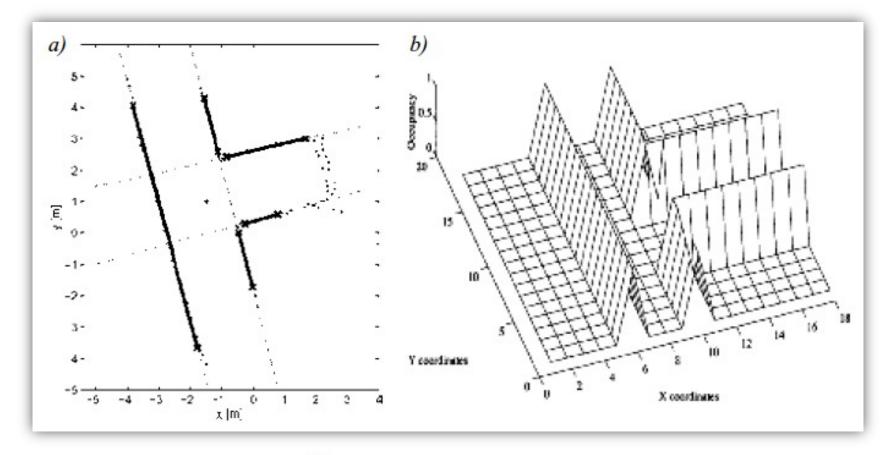








Feature Extraction













Feature extraction based on range data (laser, ultrasonic, vision-based ranging)

Most of today's features extracted from ranging sensors are geometric primitives such as line segments or circles. The main reason for this is that for most other geometric primitives the parametric description of the features becomes **too complex**, and no closed-form solution exists. Here we describe line extraction in detail, demonstrating how the uncertainty models presented above can be applied to the problem of combining multiple sensor measurements. Afterward, we briefly present another very successful feature of **indoor mobile robots**, the corner feature, and demonstrate how these features can be combined in a single representation.











Feature extraction based on range data (laser, ultrasonic, vision-based ranging)

Line extraction

Geometric feature extraction is usually the process of **comparing** and **matching** measured sensor data against a predefined description, or template, of the expect feature. Usually, the system is overdetermined in that the number of sensor measurements exceeds the number of feature parameters to be estimated. Since the sensor measurements all have some error, there is no perfectly consistent solution and, instead, the problem is one of optimization. One can, for example, extract the feature that minimizes the discrepancy with all sensor measurements used (e.g., least-squares estimation).











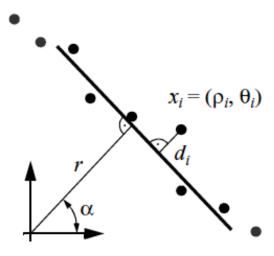
Feature extraction based on range data (laser, ultrasonic, vision-based ranging)

Line extraction

Probabilistic line extraction from uncertain range sensor data.

$$E[P_i \cdot P_j] = E[P_i]E[P_j] \qquad \forall i, j = 1, ..., n$$

$$\forall i,j=1,...,n$$









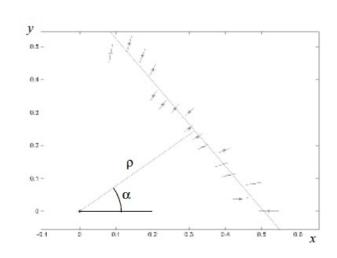




Feature extraction based on range data (laser, ultrasonic, vision-based ranging)

Line extraction

Probabilistic line extraction from uncertain range sensor data.



pointing angle of sensor θ_i [deg]	range ρ _i [m]
0	0.5197
5	0.4404
10	0.4850
15	0.4222
20	0.4132
25	0.4371
30	0.3912
35	0.3949
40	0.3919
45	0.4276
50	0.4075
55	0.3956
60	0.4053
65	0.4752
70	0.5032
75	0.5273
80	0.4879



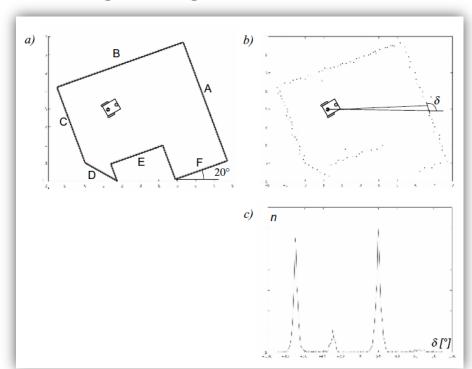


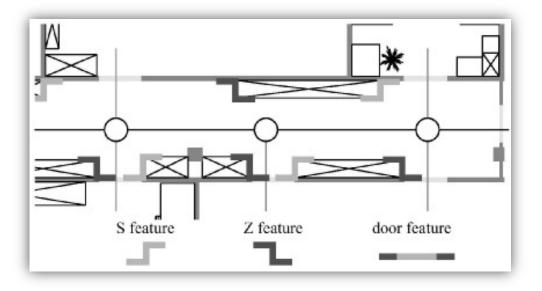






Range histogram features







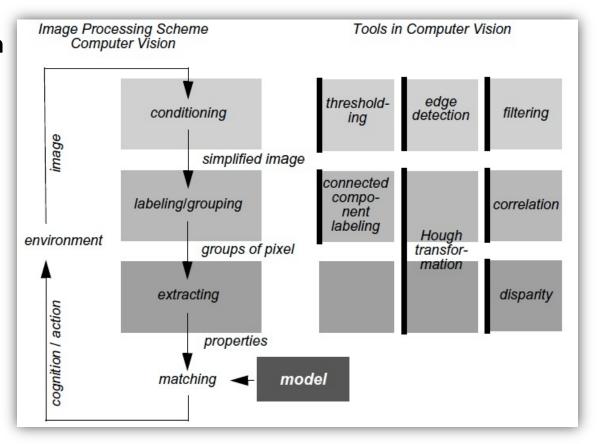








Visual appearance-based feature extraction













Visual appearance-based feature extraction

Spatially localized features

In the computer vision community, many algorithms assume that the object of interest occupies only a sub-region of the image, and therefore the features being sought are localized spatially within images of the scene. Local image-processing techniques find features that are local to a subset of pixels, and such local features map to specific locations in the physical world. This makes them particularly applicable to geometric models of the robot's environment.











Visual appearance-based feature extraction

Spatially localized features

Edge detection

