Computer Science 773

Robotics and Real-time Control

PERCEPTION

Machines sense; robots perceive. The "promotion" is in accord with the greater generality of robots; while control systems for machines must have sensors to detect events in the area they control, the events themselves are generally predictable from the nature of the machine, and specific sensors can be provided to detect the events of interest. In contrast, a robot requires general-purpose sensory equipment to match its general-purpose nature, and will have to extract information of interest from the sensory information it receives. In other words, the machine's sensors are carefully designed so that the significance of the sensory information received is immediately apparent; the robot has to interpret its sensory input to find out what, if anything, it means.

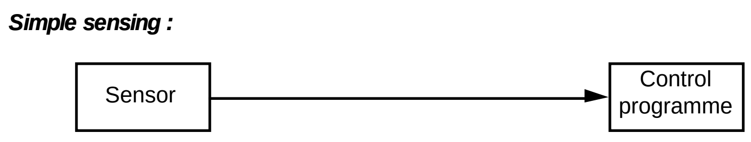
What sorts of general-purpose sensors are available ? The obvious ones, and those most widely used in robots, are the analogues of the human sensory organs. We get our information from the universe for the most part by sight, hearing, and touch; so do robots. While it is presumably possible in principle to invent others, the only obvious candidates for this category are sonar (not used much by people, but common in bats) and lidar, more commonly called laser rangefinding. Other individual sensors are sometimes used (temperature, altitude, magnetic field, etc.), but these can hardly be said to give general information about the universe.

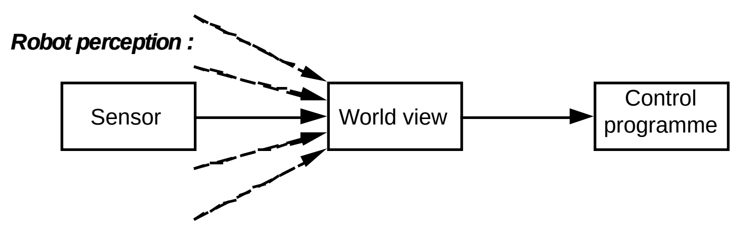
Robot sensors are usually classified as *contact* or *non-contact* sensors. The non-contact senses, based on light or sound, are typically effective at long range, and not invasive. Contact sensors can only be used when the object to be sensed is within reach, and – obviously – require contact with the object, but because of that can get some information which is not accessible from a distance. The two are complementary.

THE WORLD VIEW.

Information from the sensors is usually used to construct a model of the environment, called a *world view*. (There is debate on whether or not this is a good idea. Perhaps we'll get on to that later; for the moment, we carry on because it's a common thing to do, and is probably nothing like as bad as its opponents try to make out.)

The world view is something that doesn't really exist in simpler systems. With a conventional control system, the significance of each sensor is fairly precisely known, and it typically gives one piece of information which is used for one specific purpose. That being so, the sensor reading can simply be channelled to the procedure which uses it, and need not be noticed by anything else. In robot perception, in contrast, the sensors used are usually very general to match the generality of the robot's application. Rather than being specifically matched to some particular item which the control system must know, they give general information about the surrounding world. Such information must be digested before it yields the sort of information which is required by the robot software, and the world view is – to continue the metaphor – the system's stomach.





The world view is a representation of the robot's surroundings in the form of data structures of some sort. The nature of the representation, and the level of detail, can be whatever is suitable for the application; world views range from solid volume element models which record only whether each element is occupied or free to highly structured representations in which individual items are identified.

Apart from anything else, a world view acts as a useful interface between perception and action; all the sensory information goes into the world view, then when the rest of the system wants to know what's out there, it looks at the world view to find out. It also gives some flexibility in what sorts of sensor you use; provided that you get the right information in the world view, any sort of sensor will do, and you can – in principle – extend or alter the sensors without telling anyone about it. You can even use several sensors, and combine their information to give an improved world view. That's called *sensor integration*, and is a great deal harder than it sounds.

What you can't do, so far, unfortunately, is decide just what it is that you want to move from the world view to the control programme. Possibilities range from a depth map, which is a representation of the field of view with the distance to the first known solid object recorded at each ":pixel", to a full three-dimensional description of the surroundings, with individual objects distinguished and identified. Because of that, each investigating team builds its own world view system, so you can't (so far as I know) go out and buy one. As building any sort of world view is not an easy task, this is a significant deterrent to progress.

VISION.

Undoubtedly the major non-contact sense. It's accurate, gives lots of detail, and not very expensive. Unfortunately, the raw data are extremely hard to interpret. Given (say) a quarter of a million dots, how do you work out that this set of dots represents a table, this set represents a door, the door is open (so your robot can go through it), ... ? How do you even work out that the picture means that a particular point in space is occupied by a solid body ?

In fact, without some sort of additional information, you can't work out those things, but even given the information it isn't easy. To recognise the table, you might convert the input image to a set of lines by edge-detection, and compare subsets of the set of edges with stored images of the edges of tables; to find where the points are in space, you might acquire a second image from nearby and use stereoscopic analysis of the two images.

Generally, working out what's out there by just looking at the world under *ambient lighting* conditions is hard. Progress is being made, but it isn't fast; the spectacular successes (like a neural-network vision system that can drive a car at 100 kph) usually work by choosing a very specific problem, and developing special methods for that problem. (The neural-network driver wouldn't be likely to recognise a hitch-hiker except as an obstacle to avoid.)

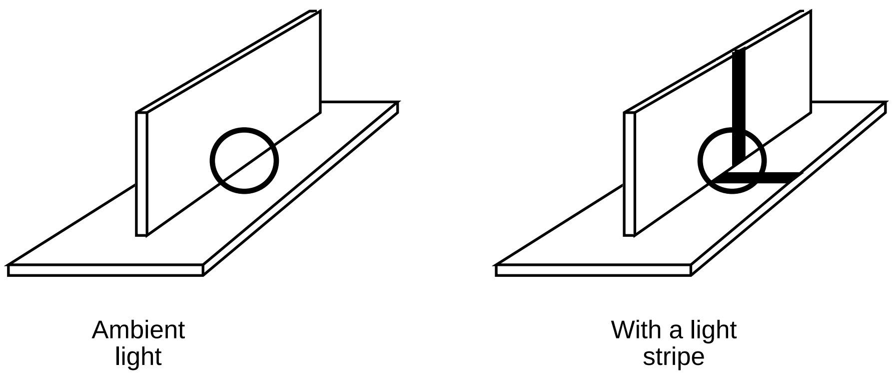
In industrial environment, ambient-light systems are sometimes used, but almost always in circumstances where the environment is artificially simplified. For example, a vision system might be used to recognise the position and orientation of parts coming along a conveyor belt so that a robot can pick them up, but to work fast enough the parts are likely to be separated on the belt, and contrasting with the belt in colour, and it will be assumed that there's nothing else on the belt.

Under such artificially simplified circumstances, rather crude, but fast, processing methods might be sufficient to do all that you need. An example is the calculation of *moments of intensity*, which are like moments of inertia (not exciting events in dramatic productions) and in effect give the parameters of the ellipse which best fits a distinguishable clump of points in the image. If you only have two or three possible object types to distinguish, that might be quite sufficient to identify the object and its orientation. This is very easy and quick; you begin by setting a threshold illumination value which classifies every pixel as black or white. Suppose that this gives a value p_{xy} to the pixel at (x, y), with 1 for black and 0 for white. (The choice is obviously arbitrary.) Then it's easy to evaluate :

$\Sigma_x \Sigma_y\, p_{xy}$	zeroth moment	M	the size of the image
$\Sigma_x \Sigma_y\, y\, p_{xy}$	first moment about the x axis	M_x	The centroid is at
$\Sigma_x \Sigma_y\, x\, p_{xy}$	first moment about the y axis	M_y	(M_x/M, M_y/M)
$\Sigma_x \Sigma_y\, y^2\, p_{xy}$	second moment about the x axis	M_{xx}	These quantities give the
$\Sigma_x \Sigma_y\, x^2\, p_{xy}$	second moment about the y axis	M_{yy}	lengths and directions of the principal axes of an
$\Sigma_x \Sigma_y\, xy\, p_{xy}$	second cross moment	M_{xy}	equivalent ellipse.

A different approach is to control the light source, giving what is known as *structured light*. This can give a lot of information about the surfaces of objects which isn't available from views in ambient light.

Consider the example below, used in practice to measure the distance between the end of a vertical plate and a horizontal plate. The left-hand picture shows the view of the two plates as it might be in ambient light; in the right-hand picture, the plates are illuminated by a light source on the right which projects a single vertical plane of light (a "light stripe") (represented in the picture, for obvious reasons, as a vertical plane of dark). Compare the views in the circle, which represents the picture seen by a video camera.



Lidar (light detection and rangefinding, otherwise laser rangefinding) can be thought of as another sort of structured light. (It doesn't have to be a laser, but that's the easy way to control the light beam sufficiently precisely.) This depends on measuring the time of flight of a pulse of laser light as it travels to an object and the reflection returns. This gives a depth map of the field of view, from which it is comparatively easy to work out where there are objects, and what their shapes might be.

SOUND.

Sound is almost invariably used as sonar (as in bats – originally sound navigation and ranging, I think). There's no reason, so far as I know, why robots shouldn't just listen to the environment and interpret what they hear – at least as far as finding the direction of sounds – but they don't. Presumably that's because it isn't very useful.

With sonar, you can in principle detect almost anything, even if it isn't making a noise of its own. It's comparatively cheap, and unobtrusive if you use ultrasonic frequencies, which you do.

In practice, it isn't as simple as it sounds (no pun really intended), for several reasons. Soft things can absorb sound very effectively, so your robot might decide that an armchair was an open door. Hard things tend to give specular (mirror-like, as opposed to diffuse) reflections if they're anywhere near planar, so an oblique surface might be quite inaudible and look like a reflection from a wall further back. The directional quality is not too bad, but not spectacularly good either.

But it can be made to work in a rough and ready way, and is quite good at tasks like obstacle avoidance. You don't get anything like the amount of detail from sonar that you do from vision, so it isn't much good for identifying objects unless they have very distinctive shapes on a fairly large scale.

TOUCH.

A lot of work has been done to develop good touch sensors. It isn't as dramatic as vision, but can in principle give much more precise information on surface properties. If one could make touch-sensing grippers for robots, they would be very effective for working on manipulations in awkward places – consider how we can screw together nuts and bolts without being able to see them.

The ideal is to make something like skin, which will have, more or less, the properties of the more sensitive parts of the body, such as fingertips. The specifications are horrifying :

- resolution of 1 mm;
- area of a few square centimetres;
- sensitive to pressures over a range of at least 1 : 1000;
- hard-wearing.

(And that leaves out self-maintaining and self-repairing as too hard even to contemplate.) None of these things is insuperably difficult taken by itself; the really hard bit is to pack together several hundred sensitive devices in a sufficiently small area.

Three major approaches are used, depending on different sorts of sensor. Piezoelectric, piezoresistive, and optical sensors have been used. All of them seem to work not too badly, but (I think) none has yet managed to meet the skin specifications as yet.

Alan Creak,
April, 1998.