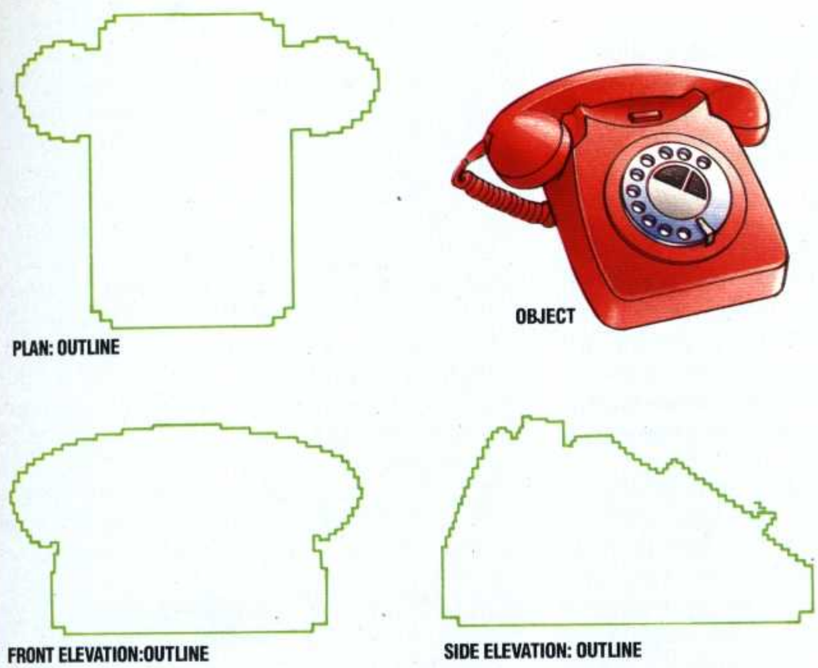

Object Recognition

To learn an object, a robot follows a process similar to human learning: it takes the image it sees and compares that image with a known object until it can find a match. Unlike humans, however, robots have limited visual abilities and can store only a very small range of object patterns. Without the depth of human experience, and the human's highly developed comparative abilities, the robot could not determine that the perceived object matched the stored outlines known as 'telephone', even by rotating its image into several positions

KEVIN JONES

objects that it is likely to see and then have it examine the image to see if any of these objects are present. This is known as the 'top-down' approach, so called because the robot begins with a very complex, high-level idea about what it might be seeing and then checks to see if its actual visual input corresponds.

OBJECT RECOGNITION

To illustrate the difference between the two methods, consider a robot that is looking at a table. The bottom-up approach consists of analysing the image and finding that it contains four vertical parts and, near the top of them, a large horizontal surface. This corresponds to the pre-programmed knowledge that a large surface could be resting on four legs and that this structure is called a table. The top-down approach would start with the robot looking at the table and asking itself 'is that a table?'. It can ask itself this question because it has an internal model of a table against which it checks its visual input.

In general, the bottom-up approach enables the robot to see things that it has never encountered, and to understand something about them — though to do this it requires a great deal of detailed programming to give it the necessary basic rules about the world that it will encounter. However, the top-down approach allows the robot to recognise only objects that it already has some internal knowledge of — so anything new to it will cause problems.

Both methods are used by robot designers, and sometimes a mixture of the two methods is utilised. It seems likely that humans use similar methods in their own visual perception, but we do it automatically and are unaware of these processes taking place.

But robot vision is, as yet, far from perfect. There are several reasons for this. One of the most important is the sheer amount of processing power that is needed to process an image. Remember that our example system had 125,000 pixels stored as one byte each — over 122 Kbytes of memory must thus be processed for each image. Although we have simplified our description, many of the processes that must be carried out on each pixel are mathematically fairly complicated. If the robot has to observe the world about it in 'real time' (i.e. it considers events as they happen), then 25 different images will be received every second (this is also true of television cameras). This means that the robot would need to process over 3,050 Kbytes of data each and every second — which is roughly equivalent to the contents of more than one dozen floppy disks!

To deal with the problem of processing, two approaches may be considered. One is to develop special-purpose hardware to perform the image-processing (such hardware is now becoming available). Alternatively, the image resolution and number of grey scale levels may be reduced to match existing hardware. This will result in the image being processed more quickly, but the picture quality will be poorer.

At present however, the subject of robot vision is still not completely understood — any more than the details of human sight are. Robots often make mistakes when using a vision system. It may well be that, ultimately, the only answer will be to develop systems in which the robot 'learns' to see things instead of being programmed in detail about what it can and cannot see. And it may eventually prove to be the case that a robot can never 'see' things properly until a way has been developed to give it a much greater knowledge of the surrounding world — knowledge that is comparable to our own, in fact.