Module Title: Introduction to Vision and Robotics
Exam Diet: SAMPLE PAPER 2004
Brief notes on answers:

1. (a) A technology domain is a particular form or level of technology that can be used to solve a problem. For example a robot behaviour could be controlled mechanically, electronically or computationally. Different domains have different characteristics, for example in size and complexity. Typically, robotics involves a hierarchy of domains interacting through transduction processes.

   (b) A wheeled robot will be statically stable with three or more wheels. Consider some possible designs with one wheel at the front and two at the back. The front wheel could be driven and steered, and the back two wheels passive. Or the back wheels could provide drive (on a single axle) while the front is used to steer. Or the front could be a passive castor and the back pair separately driven to enable turns. The latter robot can turn on the spot (unlike the first two examples) but to be holonomic it should be able to move in an arbitrary direction. This would require all three wheels to be steerable.

   (c) P = proportional, I= integral, D= derivative or differential. 'Proportional' means the control signal is proportional to the feedback error. 'Integral' integrates the error over time to allow the system to avoid steady state error. The 'derivative' term adjusts the control signal according to the rate of change of the system, acting like a friction force to prevent overshoot and oscillations.

   (d) 'Degrees of freedom' are the number of parameters needed to specify the position of an object (such as a robot endpoint) in space, or the number of independent movements with respect to a co-ordinate frame. The term is sometimes used for 'degrees of mobility' i.e. the number of separately controllable articulations of a robot, but some of these may be redundant.

   (e) Focus, shadow, saturation, specularities, non-uniform illumination, radial lens distortion.

   (f) A feature vector is a compact representation of an image, consisting of a set of descriptions or properties, such as compactness or number of corners. This can be used for recognition by comparing the feature vector of the current image to a stored vector (e.g. created by averaging over a number of examples). For recognition tasks it is useful to have the feature vector contain descriptors that are position, scale and rotation invariant.

   (g) Feature tracking - this looks for a change in position of a particular feature in the image from one frame to the next. Usually this involves searching for the best match in a small area around the previous position. It can fail if matching is not unique

   OR Correlation - this looks at the change in intensity over time at two positions. If the signal at one position correlates highly with the delayed signal at another position, this could indicate that the image is moving in the direction of the first position from the second.

   OR Gradient - image velocity can be extracted by taking the ratio of the spatial gradient to the temporal gradient at each point in the image.

(h) The process characteristics of a simple motor connected to a battery can be described as

$$V_{battery} = AV_{motor} + B\frac{dV_{motor}}{dt}$$

Adding gears with a ratio y will give an output velocity $V_{out} = V_{motor}/y$ (e.g. for $y > 1$ the velocity will decrease) and an output acceleration $\frac{dV_{out}}{dt} = y\frac{dV_{motor}}{dt}$ (e.g. for $y > 1$ the torque will increase). Thus the process equation in terms of V_{out} will be

$$V_{battery} = AyV_{out} + \frac{B}{y}\frac{dV_{out}}{dt}$$

This will alter the steady state speed and the half-life of the exponential decay towards that speed.

2. (a) Answer to first part of question 2

There are a number of ways a robot can avoid running into obstacles. These can be roughly grouped into two approaches: reactively sensing the obstacle and taking evasive action; or having prior knowledge of the location of obstacles and planning a route that avoids them.

In the first case there are many different sensory systems that can be used. The simplest is some kind of bump or whisker sensor that triggers a switch when an obstacle is encountered. If the robot is to avoid any contact, then a proximity sensor such as reflected IR could be used. Range sensors such as sonar or laser sensors allow detection of obstacles to occur over longer distances. These latter senses work by sending out a signal and measuring the time for the reflection to arrive. They can suffer from various errors if the reflection is not clean (could give examples).

The second case requires the robot to have more explicit knowledge, of the world around it, and its current position. The approach is often characterised as finding a representation of free-space, and searching for a connected route through this space to the goal.

(b) Answer to second part of question 2

Path integration uses basic geometry to work out the position of a robot, given an initial position and knowledge of the movement that occurred. A simple case is where the robot makes independent rotations and straight line movements. It could use an optical encoder on its wheels to estimate the heading angle θ and the distance travelled, d. Then $x_t = x_{t-1} + dsin\theta$ and $y_t = y_{t-1} + dcos\theta$.

Problems with path integration occur because small errors in the distance or heading estimate can accumulate into large errors in position. A popular solution is to have some external reference points in the world that can be used to recalibrate the system. These points are called landmarks. They are perceptual distinctive features in the environment that the robot can recognise, preferably from different viewpoints, and use to orient itself. A road sign is an example of an artificial landmark (i.e. one introduced for the purpose of navigation) whereas a pub sign might serve as a natural landmark, i.e. something already existing in the environment that can be adopted for navigation. Good properties for a landmark is that it doesn't move or change, can be detected over a wide area, and is unique.

(For extra marks could describe use of landmarks in visual homing or path planning or topological maps)

(c) Answer to third part of question 2

Assume the robot has detected a landmark at 45 degrees to the right and 2 metres away. Using open loop control, it would need to calculate the motor command(s) needed to turn 45 degrees and move 2 metres. It could do this using an inverse model: e.g. from the robot geometry an equation relating the motor signal to the end position of the robot could be derived, and this could be inverted to obtain an expression that, given a desired end position, specifies the motor commands.

One problem with open loop control is that it is unable to compensate for external disturbances, e.g. if one wheel slips while the robot is trying to execute the movement. Feed-forward control would deal with this by trying to measure such disturbances and adjust the motor output accordingly, e.g. adding an appropriate sum to the motor command to compensate for the detected wheel slip.

Feed-back control instead measures any error in the executed command, and uses this to correct the behaviour. For example, the goal is to turn till the landmark is straight ahead, and then move forward till the robot is at the landmark. The robot could turn at a rate proportional to the deviation of the landmark from zero degrees. In that way, it would make the appropriate correction and stop turning when zero was reached, without needing to know the exact relationship between turning commands and actual rotation, and automatically correcting for any disturbances.

3. (a) Answer to first part of question 3

An outline of the process would be: 1. Use camera to capture an image; 2. Process the image to extract important features; if necessary, subdivide image into regions that contain at most one landmark; 3. Determine through a matching algorithm the identity of the currently visible landmark; 4. Use the landmark as required for navigation.

For 1, a number of decisions need to be made, for example, how often to capture the image, whether to stop or keep moving, whether the image needs to be in colour, what resolution is sufficient, how the camera angle relates to the possible positions of landmarks and so on.

For 2, processing could include thresholding or other methods to separate possible landmarks from the background; ways to clean up the image; methods to isolate multiple shapes in the image or detect if no possible landmarks are currently visible; and the calculation of relevant feature properties, such as compactness or moments of the shape. It would be necessary to use scale-independent features because the landmark size will vary with distance, but it may not be necessary to use rotation invariant features. If the landmark image varies with viewing angle, the problem would be more difficult.

For 3, there would need to be some prior means of knowing what are the 'typical' features of the different shapes. This could be pre-defined (e.g. number of corners) or could be learnt using examples. Then for each shape in the current image the probability that it falls into a particular class of landmark could be calculated by one of several methods, such as using Bayes rule under the assumption of a gaussian distribution.

(b) Answer to second part of question 3

Active vision could be used in several ways. Detection of a particular shape could be confirmed by moving towards it and checking that the classification remains consistent. Two images taken a small distance apart could be used to separate landmarks from the background, as objects at different distances will have different relative movement between the two images. In cases where the landmark is partially obscured, either falling the edge of an image or behind some other object, an appropriate movement could be made to capture a new image that provides a better view.

(c) Answer to third part of question 3

Proprioception is the sensing of self-motion. Vision can be a powerful proprioceptive cue, because self motion will cause characteristic motion of the image on the retina. For example when moving forward, the image expands from the point towards which we are headed, at a rate proportional to our forward speed. 'Time to contact' can be extracted directly from a moving image without having to know the actual size or distance of the object we are moving towards. Image slip is a useful cue for maintaining balance.

Robots commonly use various other means to sense their position and movement. Optical encoders are often used to sense wheel rotation or joint movement. These work by detecting changes in light reflected from or passing through a disk with a regular pattern of segments that is positioned on the axle. Another type of position sensor is a potentiometer, which produces a variable resistance. Another class of sensors detect forces, e.g. strain gauges and piezoelectric sensors that respond to deformation. Robots might also be fitted with gyroscopes and tilt sensors.