The input to the perception process is usually:

1. Digital data from a number of sensors
2. Partial model of the environment (a world model)

Sensor data can take a number of forms, such as a scalar, a vector $\mathbf{x}(\alpha, \beta)$ acquired over a time series $x(t)$, a scan $x_t(\theta_i)$, a vector field $\mathbf{x}$ or a 3-dimensional volume $x(\rho, \theta, \phi)$

The definition of state and the appropriate methods for estimation are closely related to the representation adopted for the application.

The rigid-pose pose of a robot in the world is characterized by position and orientation with respect to a reference frame:

$$\text{pose is defined by a 2-tuple: } (\mathbf{R}, \mathbf{H})$$

$\mathbf{R}$ is the orientation of the object represented by a rotation matrix with respect to a reference frame. $\mathbf{H}$ represents the translation of the object with respect to the reference frame

Grid-based presentations

In grid-based representation the world is tessellated into a number of cells. The tessellation can either be uniform or tree based using quad-tree or oct-trees. Tree-based methods are well suited for handling inhomogeneous and large-scale data sets.

In a grid model each cell contains a probability over the parameter set.

In a grid model for representation of the physical environment, cells specify occupied (O) or free (F) and the cell encode the probability $P(\text{occupied})$. Initially, when we have no information the grid is initialised to $P(O) = 0.5$ to indicate unknown.

Assuming we have sensor models available that specify $P(R \mid S_{ij})$, probability of detection objects for a given sensor and location. So, using Bayes Theorem, we can update the grid model according to:

$$p_{ij}(t+1) = \frac{P(R \mid S_{ij} = O)\, p_{ij}(t)}{P(R \mid S_{ij} = O)\, p_{ij}(t) + P(R \mid S_{ij} = F)(1 - p_{ij}(t))}$$

where p_{ij} is computed across the grid model whenever new data is acquired.