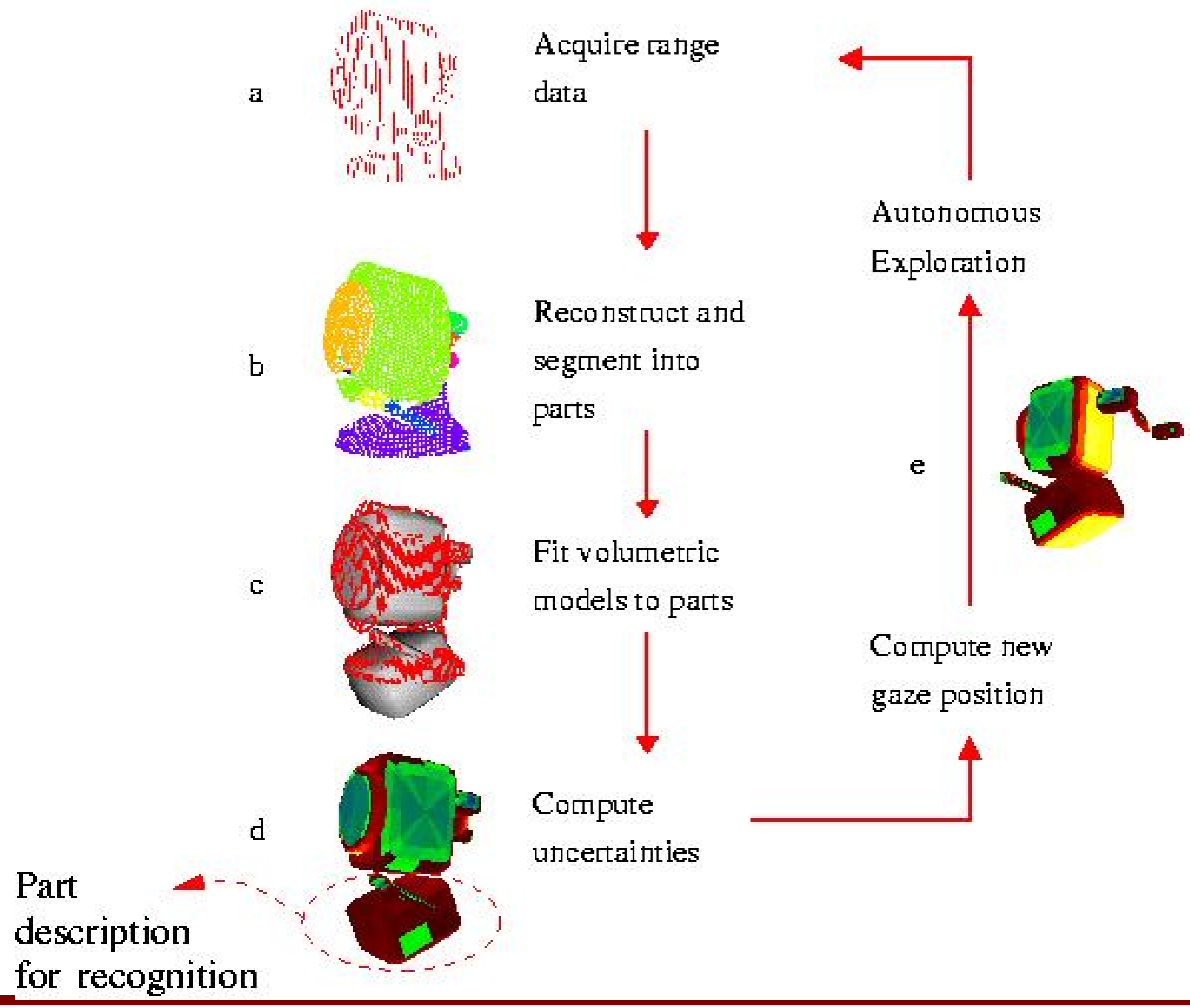
# ECSE-626

# Statistical Computer Vision

**Application of Bayesian Object Recognition**

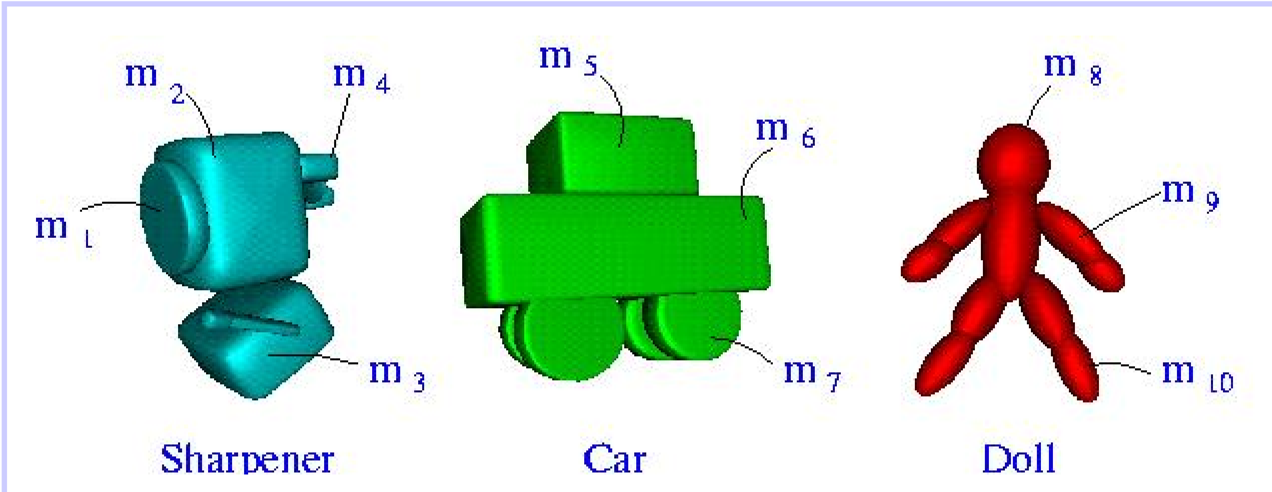
# Recognition of 3D Models

- We wish to recognize objects with complex shapes.
- Can find a single complex model to represent entire object:
  - Hard to find a general model
  - Sensitive to partial occlusion
- Alternately – *recognition-by-parts*: object represented by collection of parts. Recognition based on identification of parts.

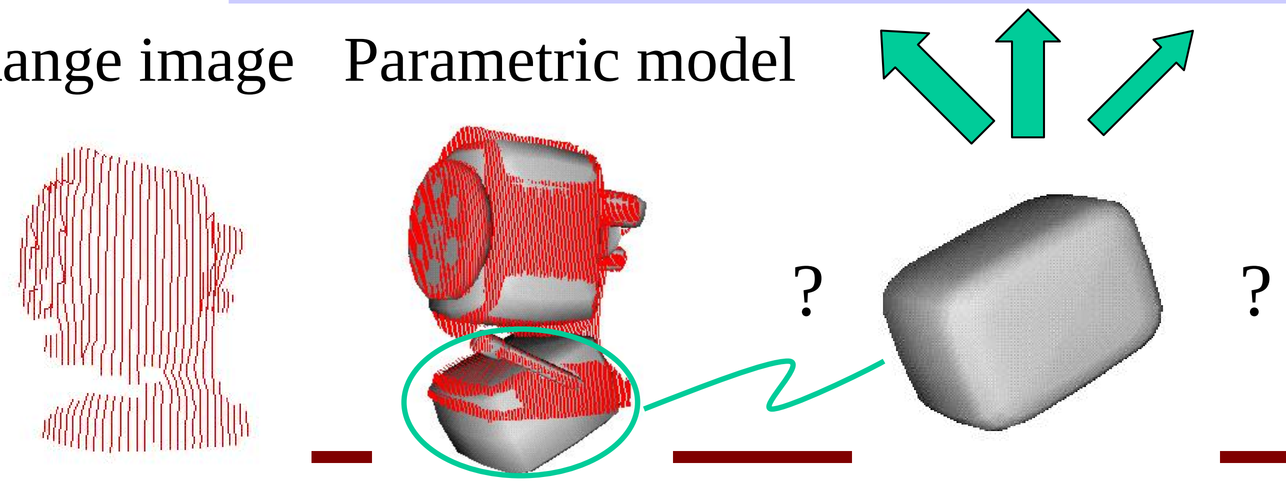

a
Acquire range
data
b
Reconstruct and
segment into
parts
c
Fit volumetric
models to parts
d
Compute
uncertainties
Part
description
for recognition
Autonomous
Exploration
e
Compute new
gaze position

# 3D Model Recognition

Database of object parts




Range image Parametric model




*Which part does the measurement correspond to?*

# Recognition of 3D Models

- Bottom-up system described used to gather sequence of data sets (i.e. from a laser rangefinder) from an object’s surface.
- Entire process, from range measurement to model fitting, referred to as *measurement* of an object.
- From a particular viewpoint, range data are gathered, segmented into parts.

# Part Recognition

- Define the data space, $D$.
- Segmented range data set for an observed part be denoted $\mathbf{d} \in D$
- Let $S$ be a discrete random variable representing the parts in the database which, in this case, take on a finite set of values, $\{S_i\}|_{\mathrm{i=1..K}}$.

# Part Recognition

- The goal of the recognition strategy is to compute a discrete conditional probability density function describing the likelihood of each of the $K$ parts in the database $\{S_i\}|_{\mathrm{i=1..K}}$ given the range data set **d**:

# Part Recognition

- The goal of the recognition strategy is to compute a discrete conditional probability density function describing the likelihood of each of the $K$ parts in the database $\{S_i\}|_{\mathrm{i=1..K}}$ given the range data set $\mathbf{d}$:

$$p(S_i \mid \mathbf{d}) \mid_{i=1..K}$$

# Information from Training

- We need to build an appropriate distribution to represent what is known about the physical theory that predicts estimates of the parameters given an object in the scene.
- No formal theory exists, so we build one empirically during *training* or *learning*.
- Monte Carlo experiments are run – $N$ measurements (in the defined sense) are gathered from different viewpoints about the object of interest.

# Information from Training

- Let $M$ denote the model space.
- Through a fitting process, each measurement leads to a parametric part descriptor, $\mathbf{m} \in M$
- The parameters of the $N$ measurements gathered about each part, $\{\mathbf{m}_j\}_{j=1..N}$ are used to build a sample mean, $\mu_i$ and a sample covariance matrix, $\mathbf{C}_i$, for each part class $S_j$.

# Information from Training

$$\mu_i = \frac{1}{N}\sum_{j=1}^{N}\mathbf{m}_j$$

$$\mathbb{C}_i = \frac{1}{N-1}\sum_{j=1}^{N}(\mathbf{m}_j - \mu_i)(\mathbf{m}_j - \mu_i)^T$$

# Information from Training

- A concise representative distribution for each class can be built by making the following assumption:

***Multivariate Normal Distribution****:*

*Each object part in the database can be represented by a single multivariate normal distribution in the M parameter space.*

# Information from Training

- The validity of the assumption should be verified empirically through experimentation.
- Should the assumption prove invalid, the shape can be modified without changing the theory.

# Information from Training

- The conditional probability density function representing the forward solution is estimated for each part in the database:

# Information from Training

- The conditional probability density function representing the forward solution is estimated for each part in the database:

$$p(\mathbf{m} \mid S_i) = N(\mu_i - \mathbf{m}, \mathbb{C}_i), \quad i = 1..K$$

where $N(x, \mathbb{C})$ denotes the multivariate normal distribution with mean $x$ and covariance $\mathbb{C}$.

# Information from Training

- The process of fitting the appropriate data set to a multivariate normal distribution is repeated for each part in the database.

# Information from Measurements

- In the context described, a fitting procedure takes an observed range image, **d**, and produces:
  - An estimate of the observed model parameters, **m,**
  - An estimate of their uncertainty represented by the covariance operator, $\boldsymbol{C_d}$.
- The probability density function representing the measurement information:

# Information from Measurements

- In the context described, a fitting procedure takes an observed range image, **d**, and produces:
  - An estimate of the observed model parameters, **m,**
  - An estimate of their uncertainty represented by the covariance operator, $\boldsymbol{C_d}$.
- The probability density function representing the measurement information: $p(\mathbf{m}|\mathbf{d})$.

# Information from Measurements

- $p(\mathbf{m}|\mathbf{d})$ measures the error in the model fitting process. It must take into account:
  - Errors in the approximation process,
  - Errors due to sensor noise.
- Direct solution difficult to obtain due to:
  - Complexity of physical process,
  - Shortage of statistically significant samples.
- Problem is ill-posed!

# Information from Measurements

- We showed that we can approximate solution locally by a Gaussian distribution, whose mean, $\hat{\mathbf{m}}$ is the maximum likelihood estimate obtained using an iterative least squares minimization based on the range data gathered.
- The result can be interpreted as a likelihood distribution, which leads to an estimate of the posterior distribution if we assume a uniform prior on model parameters, $p(\mathbf{m})$.

# Information from Measurements

- The result is that the conditional probability density function, $p(\mathbf{m}|\mathbf{d})$, is represented by a multivariate normal distribution:

$$p(\mathbf{m} \mid \mathbf{d}) = N(\hat{\mathbf{m}} - \mathbf{m}, \mathbb{C}_d).$$

# Prior Information

There are a discrete number of part hypotheses, $\{Si\}|_{i=1..K}$.

The probability density function used to represent this information, $p(S)$:

# Prior Information

There are a discrete number of part hypotheses, $\{Si\}|_{i=1..K}$.

The probability density function used to represent this information, $p(S)$:

$$p(S) = \sum_{i=1}^{K} p(S_i)\delta(S, S_i),$$

$p(S_i)$ is the subjective *a priori* probability that the i$^{th}$ part occurs and

$$\delta(A, B) = \begin{cases} 1 & \text{for A=B.} \\ 0 & \text{otherwise.} \end{cases}$$

# Bayesian Solution

One can formulate the Bayesian solution for the part recognition problem by application of the derived formulation.

We would like to compute the conditional posterior probability for each part in the database, $p(S_i \mid \mathbf{d})$:

# Bayesian Solution

One can formulate the Bayesian solution for the part recognition problem by application of the derived formulation.

We would like to compute the conditional posterior probability for each part in the database, $p(S_i \mid \mathbf{d})$:

$$p(S_i \mid \mathbf{d}) = p(S_i) \int_M \frac{p(\mathbf{m} \mid S_i)\, p(\mathbf{m} \mid \mathbf{d})}{p(\mathbf{m})} d\mathbf{m}, \qquad i = 1..K.$$

# Bayesian Solution

$p(\mathbf{m})$ is the prior distribution on model parameters, assumed to be uniformly distributed. Therefore:

$$p(S_i \mid \mathbf{d}) = \frac{p(S_i)}{C} \int_M p(\mathbf{m} \mid S_i)\, p(\mathbf{m} \mid \mathbf{d})\, d\mathbf{m}, \qquad i = 1..K,$$

where C is a constant.

# Bayesian Solution

Substituting the sources of information into the solution gives:

$$p(S_i \mid \mathbf{d}) \propto p(S_i) \int_M N(\mu_i - \mathbf{m}, \mathbb{C}_i) N(\hat{\mathbf{m}} - \mathbf{m}, \mathbb{C}_d)\, d\mathbf{m}, \qquad i = 1..K,$$

$$\propto p(S_i) N(\hat{\mathbf{m}} - \mu_i, \mathbb{C}_{Di}), \;\; i = 1..K,$$

where $\mathbb{C}_{Di} = \mathbb{C}_d + \mathbb{C}_i$.

# Bayesian Solution

Alternately, this can be expressed as:

$$p(Si \mid \mathbf{d}) \propto \sum_{i=1}^{K} p(S_i) N(\hat{\mathbf{m}} - \mu_i, \mathbb{C}_{Di}) \delta(S, S_i).$$

This results consists of a set of delta functions, one for each part in the database, where the delta is weighted by the belief in each part.

# Sequential Recognition

- The recognition problem described is difficult due to noise and uncertainties associated with each stage of the bottom-up system.
- Recognition from single viewpoints can lead to ambiguous results – similar levels of belief in more than one model.

# Sequential Recognition

Following the strategy outlined before,

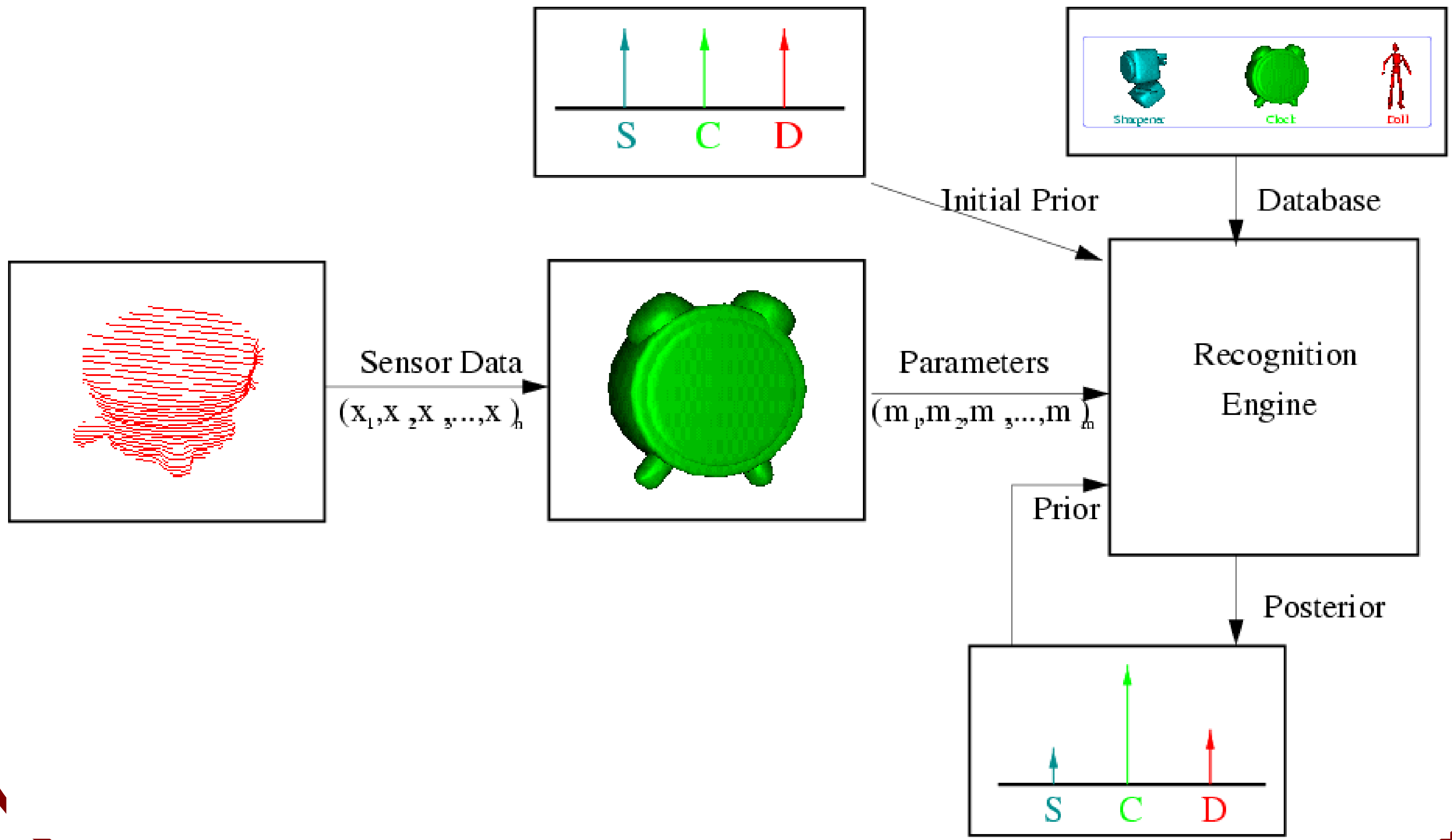
this leads to the following updating functions

for $p(S \mid \mathbf{d}_{1,\ldots,}\mathbf{d}_{t+1})$:

$$p(S \mid \mathbf{d}_{1,\ldots,}\mathbf{d}_{t+1}) \propto \sum_{i=1}^{K} p(S_i \mid \mathbf{d}_t) N(\hat{\mathbf{m}}_{t+1} - \mu_i, \mathbb{C}_{Di}) \delta(S, S_i).$$

The idea is that ambiguities should be resolved as the data sets are gathered, leading to convergence to the correct hypothesis in a short number of iterations.

# Sequential Recognition



# Part Detection

- Consider the case of a collection of object parts presented to the recognition module.
- Parts result from segmentation of 1 or more object data sets.
- Question:

  *Can we detect the presence of a particular part in the scene?*

# Part Detection

- Let $O$ denote an object in a database consisting of parts $\{S_i\}$.
- The scene consists of a set of $N$ measured parts resulting from the data set $\{\mathbf{d}\}$.
- Goal: Decide upon the likelihood of a particular database part $S$, being among the measured scene.
- This is denoted: $p(S|\{\mathbf{d}\})$.

# Part Detection

- One solution: perform recognition for each of the data sets in the scene, maintaining closed world assumption.
- Each of the data sets results in a posterior probability: $p(S|\mathbf{d}_i)|_{i=1..N}$.
- The likelihood of a particular database part, S, being among the measured scene can be defined as follows:

# Part Detection

- One solution: perform recognition for each of the data sets in the scene, maintaining closed world assumption.
- Each of the data sets results in a posterior probability: $p(S|\mathbf{d}_i)|_{i=1..N}$.
- The likelihood of a particular database part, S, being among the measured scene can be defined as follows:

$$p(S \mid \{\mathbf{d}\}) = 1 - \prod_{i=1}^{N} (1 - p(S \mid \mathbf{d}_i)).$$

# Part Detection

- Intuitively, this implies that the probability that the part is among the measured parts is equal to the negation of the probability that is *not* among the parts.

# Object Detection

- Let's examine the extension to multi-part object detection:

*What is the probability that object O is in the scene?*

- Let $R$ be the number of parts for a particular object, $O$.
- The collection of parts for O is then $\{S_i\}|_{i=1..R}$

# Object Detection

- The posterior probability that O is within the scene, given a collection of part measurements, is denoted:

# Object Detection

- The posterior probability that O is within the scene, given a collection of part measurements, is denoted:

$$p(O \mid \{\mathbf{d}\})$$

# Object Detection

$$p(O \mid \{\mathbf{d}\}) = 1 - \prod_{i=1}^{R} (1 - p(S_i \mid \{\mathbf{d}\}))$$

This implies that the probability that $O$ is in the scene is the negation of the probability that none of its parts are in the scene.