

Lectures 6 and 8: Detecting and labelling constellation-objects in noise

P. Perona
California Institute of Technology

EE/CNS148 - Spring 2004

1 Introduction

Consider an object (object class) composed of F parts. Each part has a number of coordinates (position, velocity, size, contrast...) which we indicate with x_i . The typical appearance of the object (object class) is described by a joint 'object' probability density $p_{\text{fig}}(x_1, \dots, x_F)$. Each part is detected by an appropriate feature detector.

Apart from the object in the image there is a background which contains texture patches that may be mistaken for the features that are associated to the object parts.

There are three 'natural' questions:

Detection - Is there an object at all?

Counting - How many objects are there?

Labelling - Suppose that there is an object (or two, or three...), which one of the observed features correspond to the object parts?

We are going to address these questions in a probabilistic framework: we are going to address the detection problem by comparing the probability that the object is present vs. the probability that the object is absent, given the data that we observe. The counting and labeling problem may be similarly addressed by comparing the probability of different hypotheses. The aim of this lecture is to develop methods for calculating such quantities. In order to develop the intuition and the notation we will examine first the simple case of a 1-part object.

2 Toy problem: single-part object

Suppose that there is either one object or none, and that our object contains only one part. We detect in the image N candidate regions/points of coordinates $\bar{x} = (x_1, \dots, x_N)$.

Notation:

\mathcal{O} - The *object*. Random variable encoding for the presence of the object (i.e. $\mathcal{O} = 0$ means that no object is present, $\mathcal{O} = 1$ means that there is an object in the scene etc.).

\bar{x} - The *observations*, i.e. the detected points. We assume that the ordering of the vector \bar{x} is *random* i.e. that we are not listing detections in any special order (this fact is important - it will be used later).

h - The *hypothesis*: a variable telling us which observation is believed to correspond to the object feature. I.e. $h = i$ means that we believe that x_i corresponds to the object's feature and all other x_j are false alarms. We will indicate with $h = 0$ the hypothesis that all observations are false alarms and that the object's feature was not detected.

x^o - The *observed* coordinate of the object's feature.

\bar{x}_{bg} - The coordinates of the *background* features that have been detected. Naturally:
 $x^o \cup \bar{x}_{\text{bg}} = \bar{x}$.

n - The number of background features.

d - The index of the *detected* object feature, i.e. $d = \{0, 1\}$ where $d = 0$ means that the feature was not detected.

2.1 Detection

There are two possible events: $\mathcal{O} = 0$ and $\mathcal{O} = 1$. Then we may consider the ratio:

$$\begin{aligned} R(\mathcal{O}|\bar{x}) &\doteq \frac{P(\mathcal{O} = 1|\bar{x})}{P(\mathcal{O} = 0|\bar{x})} \\ &= \frac{P(\bar{x}|\mathcal{O} = 1) P(\mathcal{O} = 1)}{P(\bar{x}|\mathcal{O} = 0) P(\mathcal{O} = 0)} \\ &\doteq R(\bar{x}|\mathcal{O})R(\mathcal{O}) \end{aligned}$$

(the second equation is obtained using Bayes' rule). If the ratio is greater than one then we may conclude that the object is more likely to be there than not and viceversa. We assume that we know $R(\mathcal{O})$, i.e. the ratio of the *a priori* probability that the object is present and that the object is absent. If we have no idea whatsoever of the value of these probabilities then it is fair to assume that they are equal and that this ratio is equal to one.

In order to compute the ratio $R(\bar{x}|\mathcal{O})$ we need to compute the conditional probability density $P(\bar{x}|\mathcal{O})$. Observe first that there are a number of hypotheses that we may entertain to explain both $\mathcal{O} = 0$ and $\mathcal{O} = 1$, namely $h = 0, 1, \dots, n$; let's call \mathcal{H} the set of these hyp. Some of them have zero probability. There are two more random useful variables to be considered: n , the number of false alarms, and d , encoding whether the feature was detected. Observe that once we know the data \bar{x} and the hypothesis h then n and d are completely determined. Nevertheless it is useful to consider them as well as it will be clear later.

Observe that all hypotheses h are mutually exclusive and therefore we may write $P(\bar{x}|\mathcal{O}) = \sum_{h,n,d \in \mathcal{H}} P(\bar{x}, h, n, d|\mathcal{O})$.

Let's now separate the information contained in the variables:

$$P(\bar{x}, h, n, d|\mathcal{O}) = P(\bar{x}|h, n, d, \mathcal{O})P(h|n, d, \mathcal{O})P(n|d, \mathcal{O})P(d|\mathcal{O})$$

This is handy because we know how to estimate/compute each one of the terms in the product:

$P(\bar{x}|h, n, d, \mathcal{O})$ – It is the probability density of observing a certain ‘constellation’ of points where we know which point is given by the object and which are false alarms (h tells us that), and how many such false alarms there are (n tells us that). Once we know h and d it does not matter whether the object is there or not, and we know d therefore we may write:

$$\begin{aligned} P(\bar{x}|h, n, d, \mathcal{O}) &= P(\bar{x}|h, n) = P_{\text{fg}}(x^o)P_{\text{bg}}(\bar{x}_{\text{bg}}) \quad \text{if } h \neq 0 \\ &= P_{\text{bg}}(\bar{x}) \quad \text{if } h = 0 \\ P_{\text{bg}}(\bar{x}_{\text{bg}}) &= A^{-n} \end{aligned}$$

where we have assumed that the pdf of the position of the false alarms \bar{x}_{bg} is uniform of the area A of the image, and that the false alarms are independently distributed.

$P(h|n, d, \mathcal{O})$ – It is the probability of a certain hypothesis when we know the number of false alarms, whether the object's feature was detected or not, and whether the object is there at all. Observe that once we know n and d then \mathcal{O} is either determined or irrelevant and therefore we may write:

$$\begin{aligned} P(h|n, d, \mathcal{O}) &= P(h|n, d) \\ &= 1 \quad \text{if } h = 0 \text{ and } d = 0 \\ &= \frac{1}{n+1} \quad \text{if } h = i \neq 0 \text{ and } d = 1 \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

In the above we have assumed that all hypotheses are equally likely. This is true only if the order of the observations is random (as assume above). It would not be true if the observations were listed from left to right (i.e. if $i < j \Rightarrow x_i \leq x_j$) because in this case our prior knowledge of where we expect the object to be would be result in some hypotheses being more likely than others.

$P(n|d, \mathcal{O})$ – This is the probability of the number of false alarms. In first order approximation it is independent of wheter the object's feature was detected or not or whether the object is present or not. If one wants to be very precise one must say that the presence of the object will occlude part of the background and therefore limit the amount of area where false alarms will arise. If we suppose that false alarms arise independently from the background then we may model it with a Poisson probability density function $P_N(n, \lambda)$ whose mean frequency λ

is a function of A , the area of the image (the bigger the image the more false alarms we expect), and of whether the object is present or not (call $A_{\mathcal{O}}$ the area of the object):

$$\begin{aligned} P(n|d, \mathcal{O}) &= P(n|\mathcal{O}) = P_N(n; \lambda) \doteq \frac{\lambda^n}{n!} e^{-\lambda} \\ \lambda &= (A - \mathcal{O}A_{\mathcal{O}})\lambda' \\ \lambda &\approx A\lambda' \quad \text{if } A_{\mathcal{O}} \ll A \end{aligned}$$

where λ' is the expected number of false alarms per unit area of background, and $\mathcal{O}A_{\mathcal{O}}$ measures the total area taken by the object(s).

$P(d|\mathcal{O})$ – This is the probability that the feature of the object is / is not detected given the fact that the object is / is not there:

$$\begin{aligned} P(d = 0|\mathcal{O} = 0) &= 1 \\ P(d = 1|\mathcal{O} = 0) &= 0 \\ P(d = 0|\mathcal{O} = 1) &= 1 - p \\ P(d = 1|\mathcal{O} = 1) &= p \end{aligned}$$

We may now compute the desired ratio:

$$\begin{aligned} R(\bar{x}|\mathcal{O}) &= \frac{P(\bar{x}|\mathcal{O} = 1)}{P(\bar{x}|\mathcal{O} = 0)} \\ &= \frac{A^{-N}P_N(N; \lambda)(1 - p) + S_{\text{fg}}(\bar{x})A^{-(N-1)}N^{-1}P_N(N-1; \lambda)p}{A^{-N}P_N(N; \lambda)} \\ &= \left((1 - p) + p\frac{A}{\lambda}S_{\text{fg}}(\bar{x}) \right) \end{aligned}$$

where we define:

$$S_{\text{fg}}(\bar{x}) \doteq \sum_{i=1}^N P_{\text{fg}}(x_i)$$

and therefore:

$$R(\mathcal{O}|\bar{x}) = \left[(1 - p) + p\frac{A}{\lambda}S_{\text{fg}}(\bar{x}) \right] R(\mathcal{O})$$

A few observations:

1. In this expression we have made no assumption on which one of the detected features corresponds to the object. We are summing over all of them.
2. The expression that we derived is independent of the size of the image since $\lambda = \lambda'A$.
3. If p is small, then it is almost irrelevant to pay any attention to the data.

4. Suppose that we knew nothing about the position of the object, then the prior p_{fg} would be the uniform density $1/A$. Substituting above we get:

$$R(\mathcal{O}|\bar{x}) = \left((1-p) + p\frac{N}{\lambda} \right) R(\mathcal{O})$$

This says that our best bet consists in counting the number of detector responses and comparing it to the expected number λ . If they are the same, then we know nothing and we have to trust our priors $R(\mathcal{O}) = P(\mathcal{O} = 1)/P(\mathcal{O} = 0)$. If N is large, then we bias the prior towards $\mathcal{O} = 1$ etc.

How do we perform detection when we are not sure how many objects there may be? In this case we may pick a reasonable upper bound $N_{\mathcal{O}}$ for the total n. of objects that we expect, and sum over all possible number of objects:

$$R(\bar{x}) = \frac{P(\mathcal{O} > 0|\bar{x})}{P(\mathcal{O} = 0|\bar{x})} = \frac{\sum_{o=1}^{N_{\mathcal{O}}} P(\mathcal{O} = o|\bar{x})}{P(\mathcal{O} = 0|\bar{x})}$$

and proceed as before in computing $P(\mathcal{O}|\bar{x})$.

2.2 Counting

Once the object has been detected counting is easy: just find the maximum ration with respect to o :

$$\hat{N}_{\mathcal{O}} = \arg \max_o \frac{P(\mathcal{O} = o|\bar{x})}{P(\mathcal{O} = 0|\bar{x})}$$

Notice that none of these ratios may exceed one, and therefore these are *not* reliable decision criteria to decide whether the object is present.

2.3 Localization (labeling)

Localization consists in deciding which detected feature is most likely to be associated with the object. If the background probability density is constant, then one just has to find the feature \hat{i} that maximizes the foreground probability:

$$\hat{i} = \arg \max_i p_{\text{fg}}(x_i)$$

3 Multi-part object

Let's consider now an object containing multiple parts. We suppose that the parts look different, i.e. different feature detectors are 'tuned' to each. We are interested in detection, counting and labeling.

Notice that now the space of events is more complicated: even allowing for only $\mathcal{O} = \{0, 1\}$ we have numerous possibilities: either the object is not there, or the object is there and all features are detected, or some or all the features may be missed by the respective detectors.

In order to handle this more complex problem we need to upgrade our notation:

F - The *number of parts* that are present in the model.

\overline{N} - The *number of detected features*: it is an F -vector; N_f is the number of features that feature detector f found in the image.

N - The total number of detected features, i.e. $N = \sum_f N_f$.

X - The *observations*, i.e. coordinates of all the features that was found in the image. One may think of X as of a matrix, where each one of f rows contains N_f entries:

$$X = \begin{bmatrix} x_1^1 & x_1^2 & \dots & x_1^{N_1} \\ x_2^1 & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ x_f^1 & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ x_F^1 & \dots & \dots & x_F^{N_F} \end{bmatrix}$$

\overline{h} - The *hypothesis*: an F -vector of indices telling us which observations correspond to the object parts. I.e. $h_f = i$ means that we believe that the detected feature of coordinates x_f^i corresponds to the object's f -th part, and all other x_f^j are false alarms. We will indicate with $h_f = 0$ the hypothesis that all detected features of type f are false alarms and that the object's f -th part was not detected.

\overline{x}^o - The *observed* coordinates of the object's features.

\overline{x}^u - The *unobserved* coordinates of the object's features. Unlike the observed coordinates \overline{x}^o , which are constants, the unobserved coordinates \overline{x}^u are variables. Observe that, with some abuse of notation, $x_f^o = X_f^{h_f}$.

X_{bg} - The coordinates of the *background* features that have been detected. Naturally: $\overline{x}^o \cup X_{\text{bg}} = X$.

\overline{n} - The number of background (or false alarm) features. It is a vector containing the number n_f of background features detected by each detector f .

n - The total number of false alarms, i.e. $n = \sum_f n_f$.

\overline{d} - The vector of indices of the *detected* object parts. I.e. $d_f = \{0, 1\}$ where $d_f = 0$ means that the f -th part was not detected.

d - The total number of object parts that were detected, i.e. $d = \sum_f d_i$.

The main ideas of our decomposition remain valid:

$$P(\overline{x}, \overline{h}, \overline{n}, \overline{d} | \mathcal{O}) = P(\overline{x} | \overline{h}, \overline{n}, \overline{d}, \mathcal{O}) P(\overline{h} | \overline{n}, \overline{d}, \mathcal{O}) P(\overline{n} | \overline{d}, \mathcal{O}) P(\overline{d} | \mathcal{O})$$

However, each factor in the equation becomes more complicated:

$P(\bar{x}|\bar{h}, \bar{n}, \bar{d}, \mathcal{O})$ – Probability density of observing a certain set of features where we know which features are given by the object and which are false alarms (\bar{h} tells us that), and how many such false alarms there are (\bar{n} tells us that). Once we know \bar{h} and \bar{d} it does not matter whether the object is there or not, and we know d therefore we may write:

$$\begin{aligned} P(\bar{x}|\bar{h}, \bar{n}, \bar{d}, \mathcal{O}) &= P(\bar{x}|\bar{h}, \bar{n}) = P_{\text{fg}}(\bar{x}^o)P_{\text{bg}}(X_{\text{bg}}) \\ P_{\text{fg}}(\bar{x}^o) &\doteq \int d\bar{x}^u P_{\text{fg}}(\bar{x}^o \cup \bar{x}^u) \\ P_{\text{bg}}(X_{\text{bg}}) &= A^{-n} \end{aligned}$$

where we have assumed that the pdf of the position of the false alarms X_{bg} is uniform on the area A of the image, and that the false alarms are independently distributed, and we have indicated with $P_{\text{fg}}(\bar{x}^o)$ the probability density function of the position of the observed features, obtained by integrating the model's pdf $P_{\text{fg}}(\bar{x})$ over the unobserved variables \bar{x}^u .

$P(\bar{h}|\bar{n}, \bar{d}, \mathcal{O})$ – Probability of a certain hypothesis when we know the number of false alarms, which of the object's parts were detected or not, and whether the object is there at all. Observe that once we know \bar{n} and \bar{d} then \mathcal{O} is either determined or irrelevant and therefore we may write:

$$\begin{aligned} P(\bar{h}|\bar{n}, \bar{d}, \mathcal{O}) &= P(\bar{h}|\bar{n}, \bar{d}) \\ &= \prod_{f=1}^F n_f^{-d_f} \quad \text{if } \bar{h} \text{ compatible with } \bar{n} \text{ and } \bar{d}, \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

In the above we have assumed that all hypotheses that are compatible with \bar{d} and with \bar{n} are equally likely, which is true if the x_f^i are in random order.

$P(\bar{n}|\bar{d}, \mathcal{O})$ – The probability of the observed number of false alarms. If we suppose that false alarms arise independently, and that their likelihood is constant throughout the image, then we may model it with the product of Poisson probability density functions whose mean frequency λ_f is a function of A , the area of the image (the bigger the image the more false alarms we expect), and of whether the object is present or not (call $A_{\mathcal{O}}$ the area of the object):

$$\begin{aligned} P(\bar{n}|\bar{d}, \mathcal{O}) &= P(n|\mathcal{O}) = \prod_{f=1}^F P_N(n_f; \lambda_f) \\ &= \frac{\lambda^n}{\prod_f n_f!} e^{-\sum_f \lambda_f} \\ \lambda_f &= (A - \mathcal{O}A_{\mathcal{O}})\lambda'_f \\ \lambda_f &\approx A\lambda'_f \quad \text{if } A_{\mathcal{O}} \ll A \end{aligned}$$

where λ'_f is the expected number of false alarms of type f per unit area of background, and $\mathcal{O}A_{\mathcal{O}}$ measures the total area taken by the object(s).

$P(\bar{d}|\mathcal{O})$ – This is the probability that the features of the object are / are not detected given the fact that the object is / is not there. If one assumes that the probability of detection of the individual features is independent then:

$$P(\bar{d}|\mathcal{O}) = \prod_f P(d_f|\mathcal{O})$$

with:

$$\begin{aligned} P(d_f = 0|\mathcal{O} = 0) &= 1 \\ P(d_f = 1|\mathcal{O} = 0) &= 0 \\ P(d_f = 0|\mathcal{O} = 1) &= q_f = 1 - p_f \\ P(d_f = 1|\mathcal{O} = 1) &= p_f \end{aligned}$$

or, more succinctly:

$$P(d_f|\mathcal{O} = 1) = q_f^{1-d_f} p_f^{d_f}$$

In a more sophisticated model one may want to take into account the fact that neighboring features are likely to be occluded together and therefore one may want to take correlations into account.

Now we are ready to write the equation for the detection ratio:

$$\begin{aligned} P(X|\mathcal{O} = 1) &= \sum_{\bar{h} \in \mathcal{H}(\bar{N}, \mathcal{O}=1)} P_{\text{fg}}(\bar{x}^o) A^{-\sum_f (N_f - d_f)} \prod_{f=1}^F (N_f - d_f)^{-d_f} P_N(N_f - d_f; \lambda_f) q_f^{1-d_f} p_f^{d_f} \\ P(X|\mathcal{O} = 0) &= A^{-N} \prod_{f=1}^F P_N(n_f; \lambda_f) \end{aligned}$$

where $\mathcal{H}(\bar{N}, \mathcal{O} = 1)$ is the set of hypotheses that are consistent with the number of observed features and the fact that the object is present, and \bar{x}^o , \bar{d} and \bar{n} are functions of X and \bar{h} . Therefore:

$$\begin{aligned} R(\mathcal{O}|X) &= \sum_{\bar{h} \in \mathcal{H}(\bar{N}, \mathcal{O}=1)} P_{\text{fg}}(\bar{x}^o) \prod_{f=1}^F \frac{A^{d_f}}{\lambda_f^{d_f}} q_f^{1-d_f} p_f^{d_f} R(\mathcal{O}) \\ \mathcal{H}(\bar{N}, \mathcal{O} = 1) &= \{\bar{h} = [h_1, \dots, h_F] | 0 \leq h_f \leq N_f\} \\ d_f &= 0 \quad \text{if } h_f = 0 \quad \text{and 1 otherwise} \end{aligned}$$

Notice that the number of hypotheses that we need to sum over is rather large. It is possible to count the number of elements in \mathcal{H} by considering that each hypothesis \bar{h} is a vector containing F entries, the f -th one of which may take N_f values, hence:

$$|\mathcal{H}| = \prod_{f=1}^F N_f \approx \hat{N}^F$$

where \hat{N} is the average of the N_f . To give an example: suppose that our object is composed of 5 parts, each one of which has 10 detections, then we have 10^5 hypotheses!