Constellation Models
Single Object, Multiple Feature Types, Incomplete Hypothesis

Markus Weber

April 23, 1999

1 Introduction

In this note we deal with the case of a single object which might be described through a set of features of different types. Several features of the same type might be used. We consider only the labeling problem, while we examine the case of incomplete labeling hypothesis.

2 Notation

The random variable O denotes the presence of the object. It can take on values O_1 (object there) or O_0 (object absent).

We assume F features, with $f \in \{1, \ldots, F\}$ and T features types. Let t be a vector of length F, with t_f indicating the type of feature f.

We denote by $d_f \in \{0, 1\}$ the fact that feature F has been detected and collect the d_f in a binary vector d.

N^i is the number of observed points of type i, $i \in \{1, \ldots, T\}$. The N^i shall be collected in a vector N.

x^i is a vector of candidate positions for type i, $i \in \{1, \ldots, T\}$. The x^i shall be collected in a "matrix" X.

Hypothesis are denoted by the vector h, where h_f is either zero (feature not detected, i.e. $d_f = 0$), or an index into the vector x^{t_f}, indicating the foreground feature.

A is the size of the image (length for 1D, area for 2D).

3 Labeling

The world (or at least one image), can be described through the following joint density: $p(X, h, N, d, O)$.

We assume that X and N are all that is extracted from one image, while d, O, and h are hidden. Note that once h is given, d is uniquely determined.

We are interested in the probability of a certain hypothesis, given the *observed* data, $p(h|X, N)$, which we can factor, using Bayes rule,

$$p(h|X, N) = \frac{1}{p(X, N)} p(X, N, h)$$

We probably don't know how to compute the first factor, but we won't need it anyway, since it is constant throughout the comparison.

We now have to compute p(X, N, h) which is

$$p(X, N, h) = \sum_{O,d} p(X, N, h, d, O)$$

Since d is completely determined by h, there is only one term in the sum over d which has non-zero probability, namely the one with $d = d(h)$. Thus,

$$p(X, N, h) = p(X, N, h, d(h), O_0) + p(X, N, h, d(h), O_1)$$

Here, we only obtain a contribution (namely $p(X, N, O_0)$) from the first term, if h is the hypothesis containing all zeros.

Now we can concentrate on the second term, which we factor in the following way [1]

$$p(X, N, h, d, O_1) = p(X|N, h, d, O_1)p(h|N, d, O_1)p(N|d, O_1)p(d|O_1)p(O_1)$$

We can insert the individual densities, using

$$p(X|N, h, d, O_1) = \frac{P_{fg}(x^{t_1}_{h_1}, \ldots, x^{t_F}_{h_F})}{\prod_{i=1}^{T} A^{(N^i - \tau_i(d))}}$$

which simply states that the points indexed through h are to be inserted into the foreground density, while all remaining points are to be inserted into the uniform background density, $p(x) = 1/A$. There are two intricacies. Firstly, $P_{fg}(x^{t_1}_{h_1}, \ldots, x^{t_F}_{h_F})$ has to be taken marginalized onto those dimensions, which correspond to *detected* features. Thus, if d_f (and therefore h_f) are zero, the corresponding $x^{t_f}_{h_f}$ needs to be eliminated (integrated out) from the density, which is easy e.g. in the case of a Gaussian fg density. Related to this is the need for $\tau(d)$ which is a vector with τ_i indicating the number of features of type i which are detected according to d.

For the second factor, we choose,

$$p(h|N, d, O_1) = \frac{1}{\prod_{f=1}^{F} (N^f)^{d_f}}$$

which accounts for the fact that, if h_f is not zero, then there are N^f ways to choose this feature out of its corresponding pool of candidates, while there is only "one way" not to choose any candidate for this feature.

For the number of background candidates, we use again the Poisson distribution

$$p(N|d, O_1) = \prod_{i=1}^{T} p_{Poiss}(N^i - \tau_i(d), \lambda_i)$$

Finally, we assume that $p(d|O_1)$ and $p(O_1)$ are learned explicitly.

Putting everything together, we obtain

$$p(h|X, N) = \frac{1}{p(X, N)} \left[p(X, N, O_0) + \frac{P_{fg}(x^{t_1}_{h_1}, \ldots, x^{t_F}_{h_F})}{\prod_{i=1}^{T} A^{(N^i - \tau_i(d))}} \frac{1}{\prod_{f=1}^{F} (N^f)^{d_f}} \prod_{i=1}^{T} p_{Poiss}(N^i - \tau_i(d), \lambda_i) p(d|O_1) p(O_1) \right]$$

---

[1] From here on, we will write d for $d(h)$ to avoid clutter. It is nevertheless assumed that d has the value consistent with h.

We can obtain a simpler expression by dividing this conditional probability by the corresponding one for the zero-hypothesis (the hypothesis that none of the candidates corresponds to the target object). Let us therefore compute the posterior probability of the zero-hypothesis, h_0. We obtain

$$\begin{aligned}
p(h_0|X,N) &= \frac{1}{p(X,N)}\left[p(X,N,h_0,d_0,O_0) + \frac{1}{\prod_{i=1}^{T} A^{N^i} \prod_{f=1}^{F} 1} \prod_{i=1}^{T} P_{\text{Poiss}}(N^i,\lambda_i) p(d|O_1) p(O_1)\right] \\
&= \frac{1}{p(X,N)} \frac{\prod_{i=1}^{T} P_{\text{Poiss}}(N^i,\lambda_i)}{\prod_{i=1}^{T} A^{N^i}} (p(d_0,O_0) + p(d_0,O_1)) \\
&= \frac{1}{p(X,N)} \frac{\prod_{i=1}^{T} P_{\text{Poiss}}(N^i,\lambda_i)}{\prod_{i=1}^{T} A^{N^i}} p(d_0)
\end{aligned}$$

Now we can normalize the expression for the posterior over the labels by this expression:

$$\frac{p(h|X,N)}{p(h_0|X,N)} = \frac{p(d,O_0)}{p(d_0)} + \frac{P_{fg}(x_{h_1}^{t_1},\ldots,x_{h_F}^{t_F}) \prod_{i=1}^{T} A^{\tau_i(d)}}{\prod_f^F (N^f)^{d_f}} \prod_{i=1}^{T} \frac{N^i!}{(N^i - \tau_i(d))!} \lambda^{-\tau_i(d)} \frac{p(d|O_1)}{p(d_0)} p(O_1)$$