A BAYESIAN APPROACH TO OBJECT RECOGNITION[1]

by M.N.M. van Lieshout [2]

ABSTRACT : In this paper we will indicate how Bayesian methods could be applied to feature detection and object recognition problems.

KEYWORDS : ICM algorithm, likelihood ratio, maximum a posteriori estimation, maximum likelihood estimation

1 Introduction

The Bayesian methods developed by Geman and Geman [4] and Besag [3] have become an important tool in a wide range of problems concerned with image reconstruction and segmentation. In this approach a prior probability distribution is selected for the true scene, typically assigning low probability to rough or irregular images. Usually the prior is chosen from the class of *Markov random fields*. These are pixel based models with nice local characteristics. Their global behaviour can be unrealistic, but the posterior distribution is chiefly sensitive only to the local properties of the prior.

In the context of object recognition the input is a digital image, but the desired output is a graphical object such as a line drawing or a list of filled circles. The most popular techniques to find the object are multi stage procedures, combining data smoothing with some form of template matching. The Bayesian approach handles these two phases simultaneously by effectively minimizing a linear combination of two error criteria, one associated with the degree of fit to the data and the other with the 'roughness' of the output image. As the Markov models from stochastic geometry [8] provide an adequate tool to model interaction between the objects, we will use them as our prior.

2 Notation

We assume that the objects to be recognized can be represented uniquely by a small number of parameters that determine size, shape and location. The space of all possible parameter values is denoted by U and assumed to be finite ($|U| = N$). Every $u \in U$ represents precisely one object $S(u)$ in the image space T. An image or object configuration is an unordered, finite list of objects :

$$x = \{x_1, \cdots, x_n\},\ x_i \in U,\ i = 1, \cdots, n,\ \ n \geq 0.$$

The region covered by the objects $S(x_i)$ will be denoted by

$$B(x) = \bigcup_{i=1}^{n} S(x_i).$$

[1] joint work with A.J. Baddeley

[2] Department of Mathematics and Computer Science,Free University
De Boelelaan 1081 a, 1081 HV Amsterdam, The Netherlands
Centre for Mathematics and Computer Science
Kruislaan 413, 1098 SJ Amsterdam, The Netherlands

The images encountered in practice are usually digitized. Moreover they are degraded by blurring and noise. These considerations lead to the following model. Let T be a finite pixel lattice, $|T| = M$, and denote the observed value at pixel t by y_t. Then

$$y_t = \phi(H(x)_t, n_t),$$

where H is a deterministic blurring function, n is noise and ϕ is an arbitrary transformation.

We will make the following assumptions:

1. $n_1, \cdots, n_M$ are independent

2. for every constant c, $\phi(c, n)$ has a probability density $f(c, \cdot)$ depending only on c.

Under these assumptions, the likelihood of observing y given the true pattern x is given by

$$l(y|x) = \prod_{t \in T} f(H(x)_t, y_t).$$

3 Likelihood ratio method

The aim of object recognition is to extract the objects from noisy data. In statistical terms, we want to estimate the true image x using the observations y. In general it is not possible to compute the maximum likelihood estimator of x directly. Therefore we need an iterative method, which in this case investigates the effect of adding and deleting objects. A new object $S(u)$ is added to x if the conditional likelihood $l(y|x \cup u)$ is sufficiently larger than $l(y|x)$, i.e. if

$$\frac{l(y|x \cup u)}{l(y|x)} > C,$$

where C is an arbitrary constant larger than or equal to 1. An existing object $S(u)$ is removed from the scene if

$$\frac{l(y|x \setminus u)}{l(y|x)} > C.$$

In most applications, the likelihood ratio depends only on a small number of pixels. Let $\sim$ be an equivalence relation representing the region over which the blur spreads. Set

$$\beta(I) = \bigcup_{t \in I} \{s : s \sim t\},$$

where I is a subset of pixels. So the observed intensity values in I are influenced only by the values at pixels in $\beta(I)$.
Then

$$\log l(y|x \cup u) - \log l(y|x) = \sum_{t \in \beta(S(u) \setminus B(x))} [\log f(H(x \cup u)_t, y_t) - \log f(H(x)_t, y_t)]$$

so that we have to evaluate the probability density f only for pixels in the region $\beta(S(u) \setminus B(x))$.

There are several possibilities to decide at what pixels objects should be added or deleted. The choice depends both on the hardware facilities and the kind of data on hand. For instance, if there are only a few objects and the noise variance is low, simple algorithms suffice, but if the image contains a lot of noise or many overlapping objects more sophisticated algorithms are required. Parallel algorithms can only be used if suitable processors are available.

A simple sequential algorithm is the following. Given a visitation schedule $\{u_1, \cdots, u_N\}$,

step 0 : $x^{(0)} = \emptyset$.

step 1 : for $i = 0, 1, 2, \cdots$:

for $j = 1, \cdots, N$ and $k = j + iN$,

$x^{(k)} = x^{(k-1)} \cup u_j$ if $u_j \notin x^{(k-1)}$ and $\frac{l(y|x^{(k-1)} \cup u_j)}{l(y|x^{(k-1)})} > C$

$x^{(k)} = x^{(k-1)} \setminus u_j$ if $u_j \in x^{(k-1)}$ and $\frac{l(y|x^{(k-1)} \setminus u_j)}{l(y|x^{(k-1)})} > C$

$x^{(k-1)}$ else.

To speed up the algorithm, one can ignore the background region by using a mask. The criterion is evaluated only in those areas where objects are likely to be added or deleted.

The parameters can be scanned in an arbitrary way. At every step the conditional likelihood of y given x is increased. As there are only a finite number of possible configurations, convergence is guaranteed and we end in a local maximum of the likelihood function. At worst there is cycling between images of equal conditional likelihood.

The method remains valid if an other initial pattern is chosen. For instance, one may initialize the likelihood ratio algorithm by means of the Hough transform [2], but according to our experience the final reconstruction will be less accurate.

Another possibility is *steepest ascent*. Objects are added or deleted in such a way that

$$\frac{l(y|x^{(k)})}{l(y|x^{(k-1)})}$$

is maximized at every step. The reconstruction will be more accurate, but requires more computation.

step 0 : $x^{(0)} = \emptyset$.

step 1 : for every pixel u compute $w(u)$, where

$w(u) = \frac{l(y|x^{(k-1)} \cup u)}{l(y|x^{(k-1)})}$ if $u \notin x^{(k-1)}$

$w(u) = \frac{l(y|x^{(k-1)} \setminus u)}{l(y|x^{(k-1)})}$ if $u \in x^{(k-1)}$.

step 2 : find the pixel u_k that maximizes $w(u)$

$x^{(k)} = x^{(k-1)} \cup u_k$ if $u_k \notin x^{(k-1)}$

$x^{(k)} = x^{(k-1)} \setminus u_k$ if $u_k \in x^{(k-1)}$.

step 3 : go to step 1.

It is important to stop at the right time. Iterating for too long yields poor results, because too many objects are added to the scene. If the data is not too noisy, usually the increase in likelihood will suddenly decrease when the 'best' reconstruction is reached. Again there is convergence to a local maximum of the likelihood function or cycling between images of equal conditional likelihood.

Each of these algorithms has its own merits. As mentioned before, it depends on the application which algorithm to use. For simple images with little noise, we recommend the iterative procedure, which gives quick convergence to a good solution. On the other hand, if there is a lot of noise but not too many objects, it is better to use the steepest ascent algorithm.

4 The Bayesian approach

The sequential likelihood ratio algorithm tends to form clusters of overlapping objects. In applications where we want to estimate the number of objects, this phenomenon is unwanted. Also the procedure is unstable in the sense that the results depend on the way the parameters are scanned. Furthermore, small changes in the data can imply large changes in the results.

Analogous to the introduction of prior models for the true image in segmentation problems [3], used as a penalty for rough or irregular images, we too introduce a prior model $p(\cdot)$ to solve this identification problem. The *Strauss model* [1] for instance can be used to model repusion between objects :

$$p(x) \propto \gamma^{r(x)}$$

where $r(x)$ denotes the number of overlapping objects in image x.

The MAP estimator $\hat{x}$ of the true image is given by

$$\hat{x} = \text{argmax } l(y|x)p(x).$$

It is sometimes called the *penalized maximum likelihood estimator* because $\hat{x}$ maximizes

$$\log l(y|x) + \log p(x).$$

Here the second term serves as a penalty on too tightly clustered patterns.

The situation described above strongly resembles the one in image segmentation. Here too, the MAP estimator is not directly computable. Therefore we first try to solve a more simple problem. As before, denote the current reconstruction by x and the data by y. Suppose we are currently visiting parameter u and want to decide whether object $S(u)$ should be part of the scene or not based on all available information. That is, we compare

$$P(u \text{ belongs to the image } |y, x_{(u)})$$

with

$$P(u \text{ does not belong to the image } |y, x_{(u)}).$$

Hence we include object u iff

$$\frac{l(y|u \cup x_{(u)})p(u \cup x_{(u)})}{l(y|x_{(u)})p(x_{(u)})} > 1.$$

Summarizing, the algorithm is

step 0 : $x^{(0)} = \emptyset$.

step 1 : for $i = 0, 1, \cdots$:

for $j = 1, \cdots, N$ and $k = j + iN$,

$x^{(k)} = x^{(k-1)} \cup u_j$ if $u_j \notin x^{(k-1)}$ and $\frac{l(y|x^{(k-1)} \cup u_j)p(x^{(k-1)} \cup u_j)}{l(y|x^{(k-1)})p(x^{(k-1)})} > 1$

$x^{(k)} = x^{(k-1)} \setminus u_j$ if $u_j \in x^{(k-1)}$ and $\frac{l(y|x^{(k-1)} \setminus u_j)p(x^{(k-1)} \setminus u_j)}{l(y|x^{(k-1)})p(x^{(k-1)})} > 1$

$x^{(k-1)}$ else.

Note that the static threshold value used in the likelihood ratio algorithm is replaced by one that depends on the current reconstruction and parameter value. As before, we can use different initial patterns. Also at every step

$$l(y|x)p(x)$$

does not decrease, so that we have convergence to a local maximum of the posterior distribution, or at worst cycling between states of equal posterior probability.

If we take the prior distribution from the class of *Markov models* [1, 5]

$$\frac{p(u \cup x_{(u)})}{p(x_{(u)})}$$

depends only on the neighbours of u. This property replaces the local properties of the Markov random fields used in segmentation problems.

There is a steepest ascent version of this algorithm as well.

step 0 : $x^{(0)} = \emptyset$.

step 1 : for every pixel u compute $w(u)$, where

$w(u) = \frac{l(y|x^{(k-1)} \cup u)p(y|x^{(k-1)} \cup u)}{l(y|x^{(k-1)})p(x^{(k-1)})}$ if $u \notin x^{(k-1)}$

$w(u) = \frac{l(y|x^{(k-1)} \setminus u)p(y|x^{(k-1)} \setminus u)}{l(y|x^{(k-1)})p(y|x^{(k-1)})}$ if $u \in x^{(k-1)}$.

step 2 : find the pixel u_k that maximizes $w(u)$

$x^{(k)} = x^{(k-1)} \cup u_k$ if $u_k \notin x^{(k-1)}$

$x^{(k)} = x^{(k-1)} \setminus u_k$ if $u_k \in x^{(k-1)}$.

step 3 : go to step 1.

5 Discussion

We have indicated how likelihood based procedures could be used to extract objects from noisy data. There are many interesting connections with existing techniques, including template matching methods [6], the Hough transform [2] and mathematical morphology [7]. Details and illustrations will be given in a forthcoming paper.

References

[1] Baddeley, A. and Møller, J.
Nearest-neighbour Markov point processes and random sets.
International Statistical Review **57** (1989) 89–121.

[2] Ballard, D.H. and Brown, C.M.
Computer vision.
Prentice-Hall (1982).

[3] Besag, J.
On the statistical analysis of dirty pictures (with discussion).
Journal of the Royal Statistical Society, Series B **48** (1986) 259 – 302.

[4] Geman, S. and Geman, D.
Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images.
IEEE Transactions on Pattern Analysis and Machine Intelligence **6** (1984) 721 – 741.

[5] Ripley, B.D. and Kelly, F.P.
Markov point processes.
Journal of the London Mathematical Society **15** (1977) 188 – 192.

[6] Rosenfeld, A. and Kak, A.C.
Digital picture processing (second edition).
Academic Press (1982).

[7] Serra, J.
Image analysis and mathematical morphology.
Academic Press (1982).

[8] Stoyan, D., Kendall, W.S. and Mecke J.
Stochastic geometry and its applications.
Wiley (1987).