Spatial independence in the observation of visual contours

J. P. MacCormick, A. Blake
Department of Engineering Science, University of Oxford
Parks Road, Oxford OX1 3PJ, UK
ph +44 1865 273919 fax +44 1865 273908
{jmac,ab}@robots.ox.ac.uk

Abstract

Two challenging problems in object recognition are: to output structures that can be interpreted statistically; and to degrade gracefully under occlusion. This paper proposes a new method for addressing both problems simultaneously. Specifically, a likelihood ratio termed the Markov discriminant is used to make statistical inferences about partially occluded objects. The Markov discriminant is based on a probabilistic model of occlusion. This model is a Markov random field, which acts as the prior for Bayesian estimation of the posterior using Markov chain Monte Carlo (MCMC) simulation.

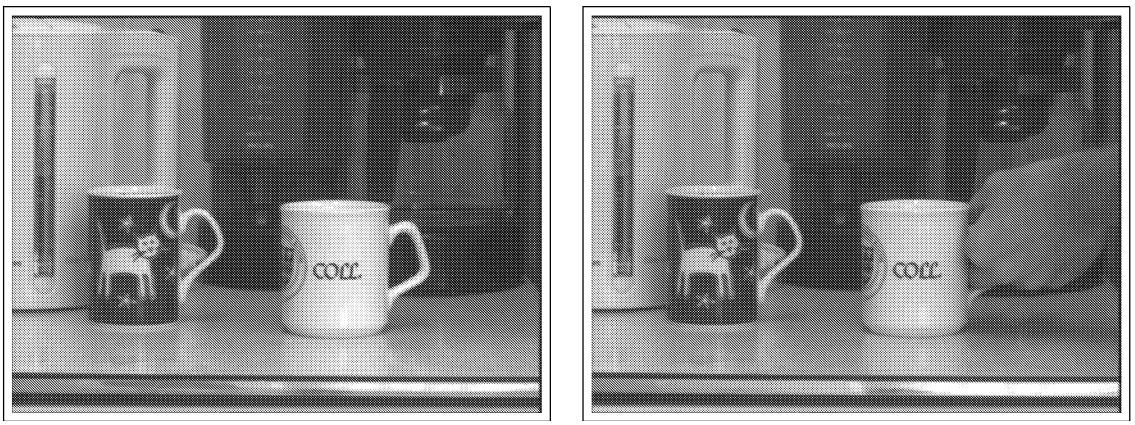
The method takes as its starting point a "contour discriminant" designed to differentiate between a target and random background clutter. We show that incorporating the prior on occlusions has two important advantages. First, partially occluded targets are assigned reasonable relative probability distributions even when they appear in scenes with unoccluded targets; this would allow a higher-level system to perform useful reasoning subsequently. Second, discrimination between partially occluded targets and background clutter is significantly improved. Both these advances can be explained by theoretical reasoning, and are demonstrated in experiments.

1 Introduction: detecting occluded objects

In some object recognition applications, it is sufficient for the output to consist of a single hypothesis. In other cases, however, the output must be statistically meaningful. Ideally the output should be a list of potential target configurations with their relative probabilities, or perhaps some other representation of the posterior distribution of target configurations. This permits data fusion with the outputs of other sensors, complex hypothesis tests, and the formulation of optimal strategies for performing high-level tasks. As a simple example, consider the task of spraying weeds while avoiding genuine plants. Given a cost function that specifies the penalties incurred by spraying a plant or missing a weed, it is easy to calculate the best strategy provided that the recognition system outputs probabilistic information. The initialisation of tracking systems is another example where statistical output could be used, since the resources of the tracking system can be distributed over the probable targets in a way that maximises some performance criteria.

This paper suggests a way of achieving statistically meaningful output for a certain subset of object recognition problems. It is assumed there is only one class of target objects to be localised (this can be thought of as recognition with a database of just one object). However, there may be more than one such target in the scene to be analysed, and some of the targets may be partially occluded. A typical example is shown in figure 1, where the problem is to localise the coffee mugs in the two images. Our objective in this paper is to design a system which reports the presence of the unoccluded mugs, and in addition detects the occluded mug *with an appropriate degree of confidence*. Note that a heuristically-based recognition system (relying,

for example, on the number of a certain type of feature matches) might have difficulty even with the left-hand image since the two targets might have very different scores. This problem is amplified in the right-hand image, where one mug is partially occluded: the heuristic scores of the two targets are very unlikely to reflect the actual relative probabilities that targets are present in those two configurations.



(a) cups with no occlusion (b) one cup partially occluded

Figure 1: ***Can a localisation system produce meaningful results from scenes like these?*** *A heuristically-based system may or may not be able to detect both mugs in both images after appropriate tuning. However, the real challenge is to report a realistic probability that the partially occluded mug in (b) is indeed a target. Heuristic scoring functions are of little use in answering this challenge.*

An essential component of systems which can output relative probabilities is a stochastic model of how the measured image features are generated. Several authors have suggested such models though most require a certain degree of heuristic input. Examples include [5, 7, 8, 14], but none of these address the specific problem of interest in this paper, which is to obtain realistic inferences despite occlusion. Amir and Lindenbaum [1] proposed a powerful method for assessing partially occluded targets, which used graph partitioning and "grouping cues" to draw inferences on whether missing edge information is due to occlusion. Although their model was designed entirely in terms of elementary probabilities, the output was chiefly useful for identifying a single best hypothesis. Indeed, the likelihoods for plausible configurations tended to differ by many orders of magnitude, possibly due to the assumption of independence between grouping cue measures.[1] Another effective approach was suggested by Rothwell [13]. This used image topology, and T-junctions in particular, to assess whether missing boundaries were genuine occlusions. It is not clear to what extent the "verification scores" of [13] can be used for statistical inferences, but in any case, the methodology presented here uses measurements based on a different set of image features. Hence the outputs of each system could in principle be fused to achieve even better performance.

The solution proposed here uses an approach involving a likelihood ratio termed a *contour discriminant*[9] to produce realistic probabilistic outputs. The next section reviews contour discriminants and explains why independence assumptions render the method inadequate for assessing partially occluded targets. Subsequent sections introduce the *Markov discriminant*, and describe experiments which demonstrate it has the desired properties.

[1] If this is indeed the reason for this effect, then it makes an interesting comparison with the contour discriminants discussed in this paper, since the Markov discriminant is designed to eliminate a certain independence assumption from the contour discriminant framework.

2 Contour discriminants

In [9] a new method of localising objects was introduced. It used a likelihood ratio termed a *contour discriminant* to assess whether hypothesised configurations of the target were likely to have been caused by the target itself or by random background clutter. Good results were achieved in experiments, in that the configuration with the highest discriminant was almost always a genuine target. In scenes containing more than one target, strong peaks in the contour discriminant were observed at every target, and in fact conditions were stated under which the values of the contour discriminant could be used to infer the posterior distribution for the presence of targets in the scene. Figure 2 explains how each hypothesised configuration is measured before its discriminant is calculated.

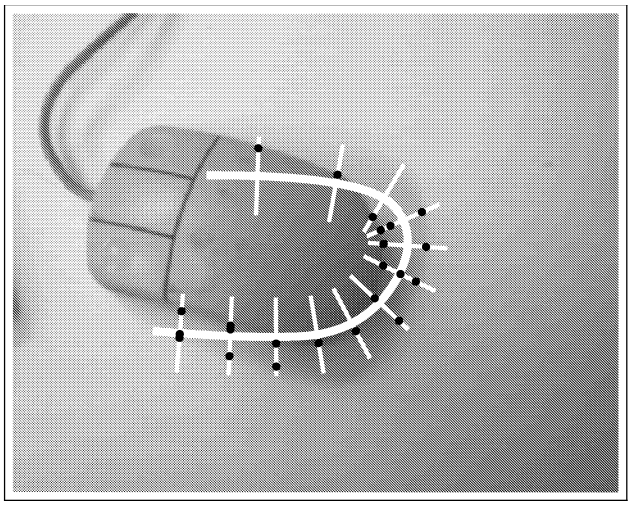


Figure 2: ***Measurement methodology and independence assumption.*** *The thick white line is a mouse-shaped contour in some hypothesised configuration. The thin lines are measurement lines, along which a one-dimensional feature detector is applied. Black dots show the output of the feature detector; it is these outputs which are modelled probabilistically by the method of [9]. The independence discriminant assumes outputs on different lines are independent, whereas the Markov discriminant introduced in this paper uses a more sophisticated approach.*

To understand the motivation behind the contour discriminant, suppose we are given a scene to analyse and are asked to consider two well-separated[2] configurations w_1 and w_2; each configuration is measured by the method of figure 2, obtaining $\mathbf{z}_1$ and $\mathbf{z}_2$ respectively. The objective is to infer from the measurements, and any prior knowledge, the probability that w_i is $\overline{w}$, the configuration of the true target. (For notational simplicity this section is phrased in terms of a two-hypothesis test, but everything generalises straightforwardly to the multi-hypothesis case.) It was shown in [9] that one way to approach this problem is to define a *contour discriminant* $D(w)$ by[3]

$$D(w) = \frac{F(\mathbf{z} \,|\, w = \overline{w})}{F(\mathbf{z} \,|\, w \neq \overline{w})}, \tag{1}$$

where F is the pdf of a probabilistic process producing the measurements $\mathbf{z}$ at configuration w. It then turns out that the probabilities for each w_i being the true configuration of the target

[2] A technical assumption which means the contours do not overlap significantly.

[3] If a prior on the configurations w is available, we can incorporate it by simply multiplying the discriminant by the value of the prior. The technical details of doing this were discussed in [9], but throughout this paper we assume for simplicity that the prior on configurations is uniform.

are in the ratio $D(w_1) : D(w_2)$. Note that here $\mathbf{z}$ is a vector whose number of elements is not known in advance, listing all measurements on all N measurement lines at the configuration w. The list of measurements on the nth line is denoted z_n, and because the contour discriminant assumes the outputs of different lines are independent, the discriminant can be written

$$D(w) = \prod_{n=1}^{N} d(z_n), \tag{2}$$

where $d(z_n)$ is a likelihood ratio analogous to (1) but for an individual measurement line. That is,

$$d(z_n) = f(z_n|w = \overline{w})/f(z_n|w \neq \overline{w}),$$

where f is the pdf of the probabilistic process generating features on a *single* measurement line. Note that this implies $F(\mathbf{z}) = \prod_{n=1}^{N} f(z_n)$. The precise model giving rise to f is not relevant to this paper, but for completeness it is briefly summarised. If $w \neq \overline{w}$, then because the contours are well-separated we can assume that all features at w were generated by random background clutter. These features are modelled as generated by a distribution termed "totally uniform", in which any number of features is equally likely and any position of features is equally likely too. This non-normalisable distribution can be made to work in this case by taking a suitable limit, and was chosen in order to assume as little as possible about the nature of random background clutter. If $w = \overline{w}$, there are two cases: either the target boundary is detected, or it is occluded. If detected, the boundary feature is drawn from a Gaussian distribution with small variance (typically 5 pixels); otherwise, it is integrated out as a hidden variable. In either case, the features on the interior portion of the measurement line are drawn from a Poisson distribution with parameter learnt from a template. The features on the exterior portion of the measurement line are due to clutter and therefore drawn from the totally uniform distribution described above.

Only one detail about f is important for this discussion: given that $w = \overline{w}$, it is assumed that there is a fixed probability q that each measurement line is occluded, and it is further assumed that these occlusion events occur *independently on each measurement line.* The importance of this fact is that it is a weakness of the contour discriminant, and this paper suggests a way to remove that weakness. To emphasise this, we will refer to the discriminant defined by (2) as the *independence discriminant.* A new expression which does not rely on the independence assumption will be introduced later and called the *Markov discriminant.*

To understand why this assumption can lead to unrealistic results, suppose a small but significant portion of the target outline is occluded — up to seven or eight measurement lines, for instance. Then the discriminant D is reduced dramatically, since according to our model of the feature formation process, seven or eight unlikely events have occurred independently. An example is shown in figure 3. In fact, only one unlikely event has occurred — a single interval of contour was occluded — but the values of the independence discriminant do not reflect this.

3 The Markov discriminant

To solve the problem seen in figure 3, we need a model reflecting the fact that occlusion events on nearby measurement lines are not independent. More specifically, the incorporation of such a model in a Bayesian framework will require a prior expressing the type and amount of occlusion expected. The approach taken here is to express the prior as a Markov random field (MRF), regarding the measurement lines round the contour as the sites of the MRF. The possible states of each site are "visible" and "occluded". Formally, suppose there are N measurement lines and denote the state of site n by s_n. We adopt the convention that $s_n = 1$ if site n is occluded, and $s_n = 0$ if site n is visible. An entire state vector $(s_1, \ldots s_N)$ will normally be denoted just by $\mathbf{s}$,

(a) output using independence discriminant on unoccluded mugs

(b) output using independence discriminant with one mug partially occluded

Figure 3: ***Relative values of the independence discriminant are not realistic for partial occlusions.*** *The independence discriminant performs adequately in (a), producing peaks in the posterior with similar magnitudes at each target. However the results in (b), in which one target is partially occluded, are not at all realistic. The displayed intensity of each contour in these figures is proportional to the log of its discriminant. The right-hand peak in (b) is actually* 10^{-8} *times weaker than the left-hand peak and would be invisible on a linear scale. Note carefully the sense in which the contour discriminant has failed. Both mugs are successfully detected, as there are strong peaks in the discriminant at each target. However, the magnitudes of these peaks are not realistic when interpreted as relative probabilities.*

and the set of all possible $\mathbf{s}$'s is written $\mathcal{S}$ — note that $\mathcal{S}$ has 2^N elements, and that typically N is between 10 and 100. The prior on $\mathcal{S}$ is denoted Θ, and the next section describes how the values of $\Theta(\mathbf{s})$ were determined in our examples.

Meanwhile, we continue the derivation of a new discriminant. The model for the formation of edge features on measurement lines is as follows:

1. A (generally small, possibly empty) subset of the measurement lines is selected as the occluded measurement lines, according to the prior Θ described in the next section. In other words, we draw a value of the occlusion state vector $\mathbf{s}$ from the prior $\Theta(\mathbf{s})$.

2. On each unoccluded measurement line, the feature generation process proceeds independently with the pdf $f(\cdot|w = \overline{w})$ described in the previous section, except that now the occlusion probability q is set to zero.

3. On each occluded measurement line, the feature generation process proceeds independently with the pdf $f(\cdot|w \neq \overline{w})$ described in the previous section.

Let $\tilde{F}$ be the new observation density arising from the model just described. For a given value of $\mathbf{s}$, we have

$$\tilde{F}(\mathbf{z} \,|\, w = \overline{w}, \mathbf{s}) = \left(\prod_{n \text{ s.t. } s_n=0} f(z_n \,|\, w = \overline{w}) \right) \left(\prod_{n \text{ s.t. } s_n=1} f(z_n \,|\, w \neq \overline{w}) \right).$$

Of course, to obtain an expression which can be used in calculating a new discriminant, we must

sum over all values of $\mathbf{s}$, weighting by the prior probabilities $\Theta(\mathbf{s})$. This gives

$$\tilde{F}(\mathbf{z} \,|\, w = \overline{w}) = \sum_{\mathbf{s} \in \mathcal{S}} \left(\prod_{n \text{ s.t. } s_n = 0} f(z_n \,|\, w = \overline{w}) \right) \left(\prod_{n \text{ s.t. } s_n = 1} f(z_n \,|\, w \neq \overline{w}) \right) \Theta(\mathbf{s}).$$

The denominator of the new discriminant is the same as the old one (1), so the second factor here cancels out and we get the following expression for the new discriminant:

$$\tilde{D}(w) = \sum_{\mathbf{s} \in \mathcal{S}} \left(\prod_{n \text{ s.t. } s_n = 0} \frac{f(z_n \,|\, w = \overline{w})}{f(z_n \,|\, w \neq \overline{w})} \right) \Theta(\mathbf{s}) = \sum_{\mathbf{s} \in \mathcal{S}} \left(\prod_{n \text{ s.t. } s_n = 0} d(z_n) \right) \Theta(\mathbf{s}).$$

To distinguish $\tilde{D}$ from the old (independence) discriminant D, we call $\tilde{D}$ the *Markov discriminant*. Although $\tilde{D}$ is a likelihood ratio and not a probability density, it will simplify our discussions to abuse notation slightly and write

$$\tilde{D}(w|\mathbf{s}) = \prod_{n \text{ s.t. } s_n = 0} d(z_n), \tag{3}$$

so that the expression for the Markov discriminant is just

$$\tilde{D}(w) = \sum_{\mathbf{s} \in \mathcal{S}} \tilde{D}(w|\mathbf{s}) \Theta(\mathbf{s}). \tag{4}$$

There is a crucial difficulty in calculating the Markov discriminant: the sum in (4) contains 2^N elements (recall N is the number of measurement lines) — far too many to be enumerated explicitly for typical values of N which are 10–100. Hence we must resort to a simulation technique, or equivalently, a Monte Carlo integration. The standard factored sampling method [4] would be to draw samples $\mathbf{s}^{(1)}, \ldots \mathbf{s}^{(K)}$ from Θ, and estimate (4) as

$$\frac{1}{K} \sum_{k=1}^{K} \tilde{D}(w|\mathbf{s}^{(k)}). \tag{5}$$

In fact, the convergence of this method is generally not rapid enough for practical estimation of the Markov discriminant. The results in this paper were instead calculated using an importance sampling technique described in section 7.

4 Prior for occlusions

As explained in the last section, the prior that models the types of occlusion expected will be expressed as a Markov random field whose sites are the measurement lines and whose state at each site is either "visible" or "occluded". Recall the notation for this: $\mathbf{s} = (s_1, \ldots s_N)$ is a state vector of the MRF, with $s_n = 1$ if the nth measurement line is occluded and 0 otherwise. The set of all possible $\mathbf{s}$'s is $\mathcal{S}$.

Our objective in this section is to define a prior Θ on $\mathcal{S}$. As explained in texts on Markov random fields (see [15], for example), there is a one-to-one correspondence between priors Θ and energy functions H. This correspondence is given by

$$\Theta(\mathbf{s}) = Z^{-1} \exp(-H(\mathbf{s})), \quad Z = \sum_{\mathbf{s}' \in \mathcal{S}} \exp(-H(\mathbf{s}')). \tag{6}$$

Configurations with higher energy have lower prior probability, as in thermodynamics. The method for designing such a prior has two steps. First, fix the functional form of a suitable energy function by incorporating intuitive notions of its desirable properties, and second, select or learn any parameters to achieve good behaviour for a given target object. The intuitive ideas to be incorporated by the first step are:

(a) Extensive occlusion is relatively unlikely.

(b) The occlusion is more likely to occur in a small number of contiguous intervals than in many separated intervals.

These are expressed by an energy function H of the form

$$H = \alpha \sum_{n=1}^{N} s_n - \beta \sum_{n=1}^{N} s_n s_{n+1}, \tag{7}$$

where α and β are positive real parameters to be determined later.[4] The first term in this expression penalises every occluded site, thus incorporating the intuitive idea (a) above. The second term encourages occlusion at adjacent sites, incorporating intuitive idea (b). It can be made even more explicit that idea (b) really has been captured here. Let $O = \sum s_i$ be the number of occluded sites and let I be the number of contiguous occluded intervals. Then penalising each of these quantities in the energy function would suggest adopting $H = \alpha' O + \beta' I$, for some α', β'. But observe that $I = O - P$, where $P = \sum s_i s_{i+1}$ is the number of adjacent pairs of occluded sites. Hence $H = (\alpha' + \beta')O - \beta' P$, exactly as in (7) if we take $\alpha = \alpha' + \beta'$ and $\beta = \beta'$.

The choice of precisely how to incorporate the two intuitive ideas is of course rather arbitrary. The above choice was guided by the desirability of simplicity. Note that the first term of (7) is the sum of single-site potentials, and the second term is the sum of pair potentials. This can be immediately recognised as the energy for an Ising model, and the graph of its neighbourhood system is the "necklace" shown in figure 4 — this is sometimes called a cyclic Markov random field [6]. Quantities of interest can now be calculated easily. For example, it turns out the probability of occlusion given that the two neighbours of a site are visible is given by

$$\text{Prob}(s_n = 1 \,|\, s_{n-1} = s_{n+1} = 0) = (1 + \exp(\alpha))^{-1},$$

and the probability of an occluded site between two other occluded sites is

$$\text{Prob}(s_n = 1 \,|\, s_{n-1} = s_{n+1} = 1) = (1 + \exp(\alpha - 2\beta))^{-1}.$$

In the examples shown here, the parameters were chosen so that the expected number of occluded sites is 5% of the total number of sites, and the expected number of contiguous intervals of occluded sites is 0.7; this corresponds to $\alpha = 5.33$ and $\beta = 5.0$. Alternatively, the values of α, β could, in principle, be learned from training data. A random sample from a simulation of this prior is shown in figure 4.

It is worth addressing one further question here: is it possible to specify α and β as a function of N, the total number of sites, in such a way that some desirable statistical properties are constant? If so, then this more general approach would be preferable to finding suitable α, β numerically for each new class of target. Unfortunately, this turns out to be a rather difficult problem. Statistical physicists are interested in the same question, and attempts to answer it have led to the deep and beautiful theory of renormalisation group transformations [2]. Vision researchers have also used renormalisation theory, though in a completely different context to this paper [3, 11]. Some unpublished work by Nicholls [10] suggests that numerical schemes for altering the parameters α, β may be of some use, but this is unnecessary for the limited range of examples addressed here.[5]

[4] To avoid messy notation, we have adopted the convention $s_{N+1} = s_1$.

[5] We are indebted to Geoff Nicholls who pointed out the connection to renormalisation group theory and suggested relevant literature.

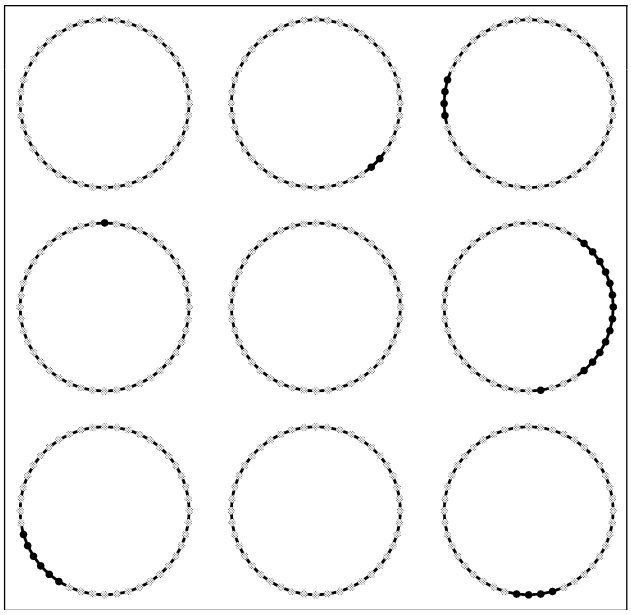


Figure 4: ***The prior for occlusions.*** *Each small dot is a site in the Markov random field, and the local characteristics of each site depend only on its two neighbours. In the application described here, each "site" is actually a measurement line on the contour (see figure 2), and the possible states of each site are "occluded" or "visible". Nine samples from the MRF described in the text are shown here; grey circles are visible and black circles are occluded. This MRF has 44 sites, the same as the number of measurement lines in the coffee mug template (see figure 7a).*

5 Realistic assessment of multiple targets

The first subsection below gives a broad outline and explanation of the result shown in figure 6. The second subsection gives precise details of how the experiment was carried out.

5.1 Explanation of results

Recall our objective in introducing the Markov discriminant: to obtain roughly equal peaks in the discriminant when evaluated at true target configurations, regardless of whether a small portion of the outline is occluded.

To test this we applied the method to two nearly identical scenes, figures 1(a) and (b). The first scene shows two coffee mugs with their handles visible; the second shows the same scene but this time the handle of the right-hand mug is occluded by a hand which is about to pick up the mug. Figures 3(a) and (b) show the posterior distribution as calculated by the independence discriminant. (The next subsection describes in detail exactly what these figures represent and how they were created.) When neither mug is occluded, the two peaks in the distribution differ by a factor of about 22 — not a particularly realistic result but at least the peaks have similar orders of magnitude. However, when the right-hand mug has its handle occluded (figure 3d), the independence discriminant evaluates the corresponding peak in the posterior as being approximately 10^{-8} times smaller in magnitude than the peak for the left-hand mug! This is essentially because 8 measurement lines were occluded, and since the non-detection probability q was set to 0.1, the independence discriminant considers that 8 independent events of probability 0.1 have occurred.

Next the Markov discriminant was applied to the same two scenes (figure 6). As before, when neither mug is occluded the two peaks in the posterior are of similar magnitude. However, figure 6(b) shows that the two peaks still have a similar magnitude even when the handle of

the right-hand mug is occluded. This is because, according to the Markov random field prior we used on the occlusion status vectors of the measurement lines, the event that 8 consecutive measurement lines are occluded is not considered particularly unlikely.

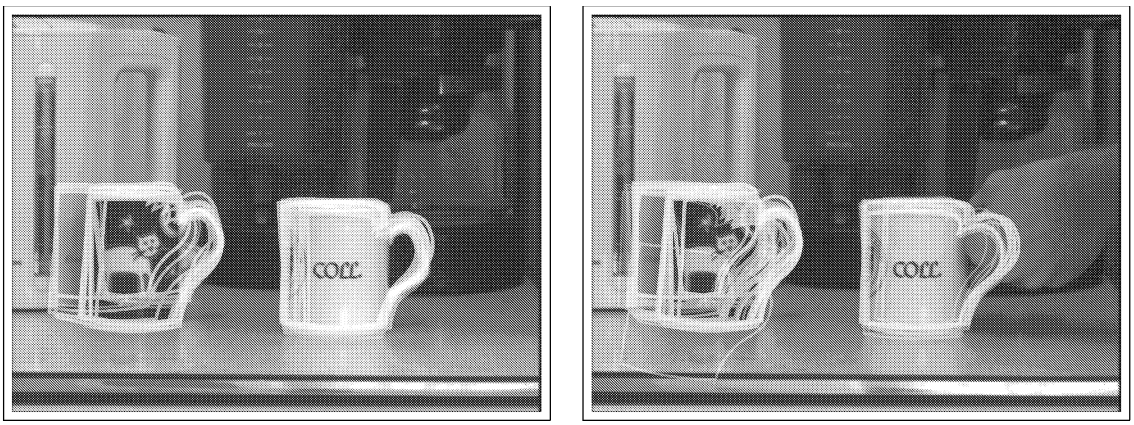
	ratio of independence discr	ratio of Markov discr
both mugs visible	22	0.31
one mug partially occluded	5.8×10^{-8}	0.24

Figure 5: ***Relative heights of peaks in posterior distributions shown in figures 3 and 5.*** *The figures shown are the peak value of the discriminant at the right-hand mug divided by the peak value at the left-hand mug. The ratios of peak values for the Markov discriminant are much more realistic than those given by the independence discriminant.*

More precise details of the relative values of the peaks in the posterior distributions are given in figure 5. Note that these numbers can actually be used for the type of statistical inferences we speculated about in the introduction. For example, suppose for simplicity we have prior knowledge that precisely one mug is present in the scene. Let w_1 be the configuration at the left-hand peak in the posterior and w_2 the configuration at the right-hand peak. Then

$$\begin{aligned}\text{Prob}(w_1 = \overline{w}) &= \frac{\tilde{D}(w_1)}{\tilde{D}(w_1) + \tilde{D}(w_2)} = \frac{\tilde{D}(w_1)/\tilde{D}(w_2)}{\tilde{D}(w_1)/\tilde{D}(w_2) + 1} \\ &= \frac{1/0.24}{1 + 1/0.24} = 0.81.\end{aligned}$$

Similar calculations can be made with more realistic priors on the number of mugs present in the scene.



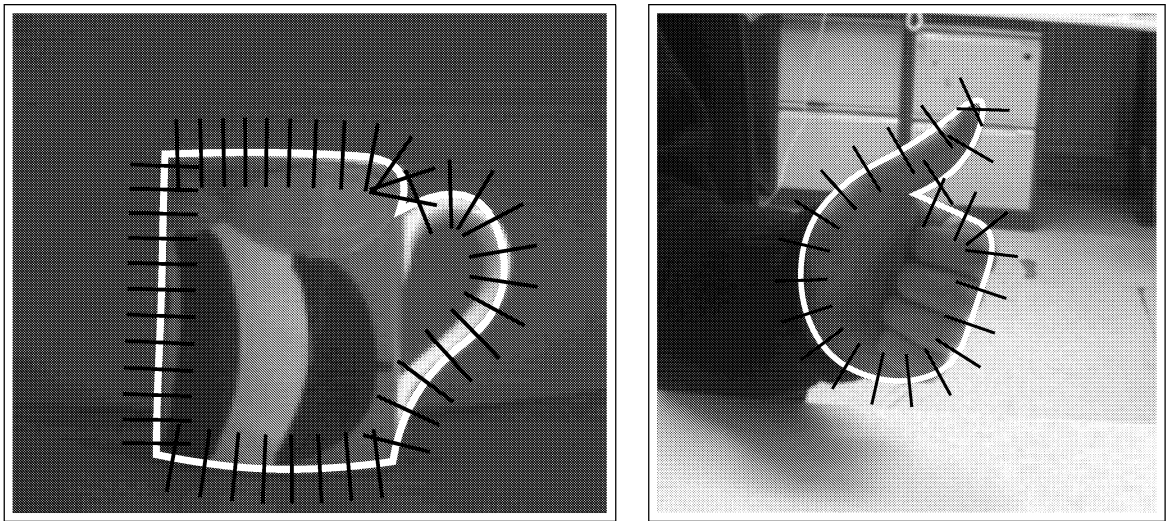
(a) output using Markov discriminant with on unoccluded mugs

(b) output using Markov discriminant with one mug partially occluded

Figure 6: ***The Markov discriminant produces more realistic posterior probabilities.*** *When both mugs are unoccluded (a), the peak posterior probability at the left-hand mug is about 3 times that at the right-hand mug. When the right-hand mug is occluded (b), the peaks differ by a factor of about 4. Both these results are more realistic than those calculated by the independence discriminant (figure 3), but the improvement is particularly marked in the case of partial occlusion.*

5.2 Experimental details

The coffee mug template was taken from the scene in figure 7(a); note that this is a different mug and background to those used for the experiment. The prior used for mug configurations had the following properties: uniform distribution over the two Euclidean translation parameters; rotation in the plane by an angle whose mean is zero and standard deviation 3°; scaling in the x-direction by a normally distributed factor whose mean is 1 and standard deviation 0.1; scaling in the y-direction by a normally distributed factor whose mean is 1 and standard deviation 0.05. For each scene, the same 10000 samples were drawn from this prior, and the independence contour discriminant evaluated for each one. The 100 configurations with the highest independence discriminants were recorded for further investigation, and the remainder discarded. This approach was taken mainly for a practical reason: it takes several seconds to estimate the Markov discriminant, whereas the independence discriminant has a closed-form formula which can be calculated in milliseconds.



(a) Coffee mug (b) Thumbs-up gesture

Figure 7: ***Templates for experiments.*** *(a) There are 44 measurement lines, which means 44 sites in the cyclic MRF used to specify the prior on which parts of the contour will be occluded. (b) Note that the signaller is wearing long-sleeved clothing so there are detectable edges on the wrist area.*

The Markov discriminants of the selected 100 configurations were then estimated by the importance sampling method described in section 7. In the plots of the posterior distributions shown in figures 1, 3, and 6, the total mass of each contour (intensity × width) is proportional to the log of the contour discriminant.

6 Improved discrimination with a single target

The main motivation for introducing the Markov discriminant was to obtain more realistic relative values in the peaks of the posterior distribution when two or more targets are present. However experiments showed that in some cases the performance of the method was significantly improved even when only a single target was present.

Figure 8 shows an example of this behaviour, where the discriminant approach is being used to search for a "thumbs-up" signal in a static grey-scale image. The template thumb-signal is shown in figure 7(b); note that there is good contrast on the wrist because the signaller is wearing long-sleeved clothing. The experiment involved searching for a thumb-signal when the

signaller might be wearing short sleeves, in which case no edges would be detected on the wrist. This absence of edge features is actually a generalised form of occlusion.

On around 90% of images tested, both the independence discriminant and the Markov discriminant correctly identified the single target as the strongest peak in the posterior. This type of behaviour is shown figures 8(a)–(d). However, occasionally some background clutter is scored higher than the true configuration by the independence discriminant, as in figure 8(e). Of course, the reason the independence discriminant fails is that it finds no edges on the wrist area and consequently gives the true configuration a low likelihood. By evaluating the contours using the Markov discriminant instead, this situation can be rectified: in figure 8(f), for example, the peak at the correct configuration is 5 times stronger than the one at the spurious hypothesis.

The reason for the improved discrimination is as follows. Recall that the independence discriminant is calculated by multiplying together the likelihood ratios $d(z_n)$ of the measurements on each individual measurement line. Therefore, the positioning of "unlikely" measurement lines (i.e. those which do not resemble the target) is irrelevant. For instance, a configuration with three very unlikely measurement lines will score the same regardless of whether these three lines are adjacent or separated from each other by intervening sites. The Markov discriminant, on the other hand, takes precisely this type of positioning into account: consecutive (or even nearby) unlikely measurement lines do not incur as great a penalty as separated ones. Consider figures 8(e) and (f) as an example: the two competing configurations have a similar number of poorly-scoring measurement lines, but on the true configuration these are all on the wrist, on consecutive lines. The Markov discriminant takes this into account and gives this configuration a higher likelihood.

7 Faster convergence using importance sampling

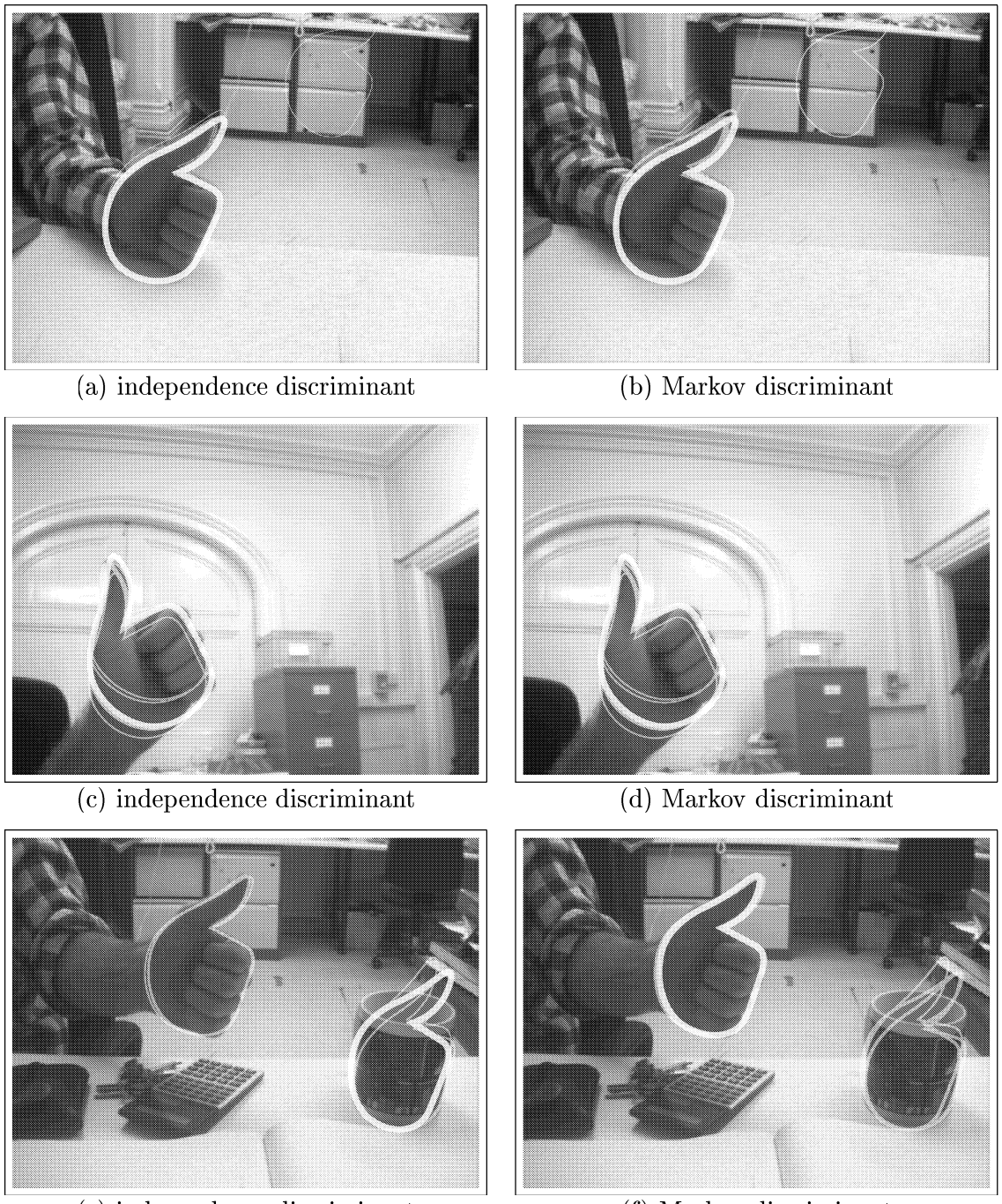
Because it is a product of up to N individual ratios, the likelihood ratio $\tilde{D}(w|\mathbf{s})$ defined by (3) is very sharply peaked when regarded as a function of $\mathbf{s}$. Hence the factored sampling estimate (5) is dominated by the few samples near a strong peak in $\tilde{D}(w|\mathbf{s})$, and the majority of samples contribute virtually nothing to the estimate. A standard method to reduce the variance of factored sampling estimates is called *importance sampling*[6] [12]. The basic idea is to spend more time sampling near the peaks of $\tilde{D}(w|\mathbf{s})$, and compensate for this by weighting the calculation appropriately. More specifically, suppose the samples $\mathbf{s}^{(1)}, \ldots \mathbf{s}^{(K)}$ are drawn from an importance distribution $\Theta'(\mathbf{s})$ on $\mathcal{S}$. Then as $K \to \infty$ the quantity (4) can be estimated by

$$\frac{1}{K} \sum_{k=1}^{K} \left(\tilde{D}(w|\mathbf{s}^{(k)}) \times \frac{\Theta(\mathbf{s}^{(k)})}{\Theta'(\mathbf{s}^{(k)})} \right). \tag{8}$$

This is true for essentially any choice of $\Theta'(\mathbf{s})$, but of course the idea is to obtain faster convergence and to this end Θ' should be chosen so that a higher proportion of the samples contribute significantly to the estimate.

In the particular case of this paper, our choice of Θ' is guided by the following observation: the likelihood ratios $d(z_n)$ of the individual measurement lines give very useful guidance on likely sites of occlusion. Consider a site i for which the value of $d(z_i)$ happens to be very low. Then it is very plausible that the target boundary was occluded at this site. Hence the importance function Θ' should be biased towards the possibility that site i is occluded. We will say in this case that site i is "encouraged" to be occluded. Of course this choice can be justified by a purely numerical argument. According to (3), the contributions $\tilde{D}(w|\mathbf{s})$ to the Markov discriminant are proportional to the product of the $d(z_n)$ *over only those lines which were not occluded.* Thus if

[6] In the specific case of statistical mechanical Gibbs samplers, importance sampling is sometimes called non-Boltzmann sampling. An accessible survey of the techniques involved is given in [2].



(a) independence discriminant (b) Markov discriminant

(c) independence discriminant (d) Markov discriminant

(e) independence discriminant (f) Markov discriminant

Figure 8: ***Improved discrimination with a single target.*** *In the first two examples (a and b, c and d), the independence discriminant and Markov discriminant have similar performance, but the Markov discriminant is significantly better than the independence discriminant at assessing partial matches, as in (e) and (f). Note the reason for the independence discriminant's failure here: the thumbs-up template had strong edges on the wrist, which are not present in (e). However, the Markov discriminant recognises that this generalised "occlusion" of three consecutive measurement lines on the wrist is not particularly unlikely, and therefore does not penalise the true configuration unduly.*

it happens that for some i the value of $d(z_i)$ is negligible, contributions to the discriminant will be non-negligible only for values of $\mathbf{s}$ that specify the ith site is occluded. Hence we are led to the following conclusion: if, for a fixed configuration w, there is a site i for which $d(z_i)$ is small, then the importance distribution $\Theta'(\mathbf{s})$ should strongly favour values of $\mathbf{s}$ with $s_i = 1$ (i.e. site i is occluded).

First we discuss the situation when only one site is "encouraged" in this way — the importance sampling method can easily be extended to encourage multiple sites but the notation becomes more complicated. Note that a different encouraged site i is selected for each configuration w. In our implementation this was done by selecting the central site of the longest contiguous low-scoring interval of contour. A measurement line was labelled as low-scoring if the likelihood ratio $d(z_n)$ was less than a parameter λ. In the rare event that no measurement line was low-scoring according to this definition, we selected the one with the lowest value of $d(z_n)$. For the examples shown here we took $\lambda = 0.2$, which corresponds to an occlusion probability of $5/6$.

The importance function for one encouraged site will be denoted Θ' and is a distribution of the Gibbs form (6) with energy denoted by H'. To define H', we must fix in advance a parameter $\gamma > 0$, which determines to what extent occlusion will be favoured at an encouraged site — this choice is discussed briefly below. Then, for a given configuration w, we select by the heuristic above a site i whose occlusion will be encouraged and define the energy of the importance function by

$$H'(\mathbf{s}) = H(\mathbf{s}) - \gamma s_i,$$

where $H(\mathbf{s})$ is the energy of the occlusion prior $\Theta(\mathbf{s})$ defined earlier. Observe that this choice does indeed "encourage" occlusion at the site i: when $s_i = 1$, the value of H' is reduced by γ, resulting in a more probable configuration.

Note that in order to perform importance sampling using (8), we must have an expression for $\Theta(\mathbf{s})/\Theta'(\mathbf{s})$. In this case, if we let Z' be the partition function — as defined by (6) — of the Gibbs distribution with energy H', we have

$$\begin{aligned} \frac{\Theta(\mathbf{s})}{\Theta'(\mathbf{s})} &= \frac{\exp(-H)}{Z} \Big/ \frac{\exp(-H')}{Z'} \\ &= \frac{Z'}{Z} \exp(-\gamma s_i). \end{aligned} \tag{9}$$

It remains to explain how to calculate Z'/Z. It turns out that for the simple 1-dimensional, 2-state MRFs used in this paper, all the partition functions can be calculated exactly from a recursive formula, but the ratio Z'/Z can also be estimated for much more general MRFs by a Monte Carlo integration. We omit the details, but some related techniques are described in [2]. Note that although we preselected a specific site i at which occlusion would be encouraged, the value of Z' does not depend on i. This is because there is an obvious isomorphism between the MRF at which site i is encouraged and the MRF at which site j is encouraged — just rotate the "necklace" (figure 4) by $i - j$ sites.

Now suppose we wished to improve this importance sampling approach by using an importance function which encouraged occlusion at two different sites i and j. The argument works as before: the importance function Θ'' is of Gibbs form with energy $H'' = H + \gamma(s_i + s_j)$, and a formula analogous to (9) holds:

$$\frac{\Theta(\mathbf{s})}{\Theta''(\mathbf{s})} = \frac{Z''}{Z} \exp(-\gamma(s_i + s_j)).$$

There is no longer an isomorphism between different choices of (i, j) — in fact the partition function of the importance distribution depends on $|i - j|$. However, for a given value of $|i - j|$,

the ratio Z''/Z can be calculated by either of the techniques mentioned above, so the method still works provided the $\lceil N/2 \rceil$ values needed are pre-computed. (Recall that N is the number of sites in the MRF, which is the number of measurement lines on the contour.) This approach can be extended to arbitrary numbers of "encouraged" sites, but the amount of pre-calculation necessary increases as $\binom{N}{E}$, where E is the number of encouraged sites.

The method is valid with any fixed value of γ, but once again the idea is to choose a value which causes (8) to converge quickly. If Monte Carlo integration is being used to estimate the partition function ratios, then an additional requirement is that these should also converge at an acceptable rate. We have not investigated the issue of how to choose an optimal γ, but empirical tests showed that $\gamma \approx \beta$ worked well. All results presented here took $\gamma = 4.5$. Similar comments apply to the parameter λ: the method is valid for any value, and empirical tests showed that $\lambda = 0.2$ is effective in speeding convergence.

The effectiveness of importance sampling with Θ' and Θ'' is shown in figure 9. Standard factored sampling and the importance sampling method were applied to the best 50 configurations of the unoccluded mug experiment. Each point on the graph is the standard error of 20 estimates of the Markov discriminant, and each of these 20 estimates was calculated from equation (8) with $K = 20000$. On average, the standard error of the importance sampling estimate is reduced by 60% if one site at a time is encouraged, and by 85% if pairs of sites are encouraged.

For completeness, here is the heuristic used to select which two measurement lines are encouraged for a given configuration: if there are two or more contiguous low-scoring intervals, select the central site in each of the two longest such intervals. If there is only one contiguous low-scoring interval, select the central site of that interval and the worst-scoring site excluding the interval. If there are no low-scoring lines, select the two with the lowest scores. The definition of "low-scoring" is the same as above.

8 Conclusion

Occlusion is a perennial problem for anyone working in object localisation. This paper introduced a new way of obtaining realistic inferences about partially occluded targets. Specifically, it addressed the problem of how to adjust the contour discriminant of [9] so that the posterior probabilities of occluded targets are not over-penalised. This was done by removing the assumption that occlusion events on different measurement lines are independent: instead, the measurement lines are treated as sites in a Markov random field whose behaviour is chosen to represent the types of occlusion expected in practice. Using this MRF as the prior for a Bayesian approach, a new discriminant termed the Markov discriminant was derived.

The paper demonstrated two situations where the Markov discriminant produces far more realistic output than the previously introduced independence discriminant. Firstly, it was shown that if the Markov discriminant is used to analyse a scene with one unoccluded target and one partially occluded target, the two peaks in the posterior distribution of target configurations have similar magnitudes (see table 5). This is in marked contrast to an identical analysis using the independence discriminant, where the posterior peaks differ by several orders of magnitude. Secondly, it was shown the Markov discriminant can significantly improve differentiation between partially occluded targets and background clutter (figure 8). This is because the Markov discriminant takes account of which types of occlusion have good support from the prior.

There is an important open problem associated with the method, related to the accuracy and speed of the estimates of the Markov discriminant. The improved performance was gained at the expense of introducing Monte Carlo simulation which takes seconds rather than milliseconds, and therefore precludes the use of the Markov discriminant in real time situations. The details given in section 7 show how importance sampling can be used to speed convergence, but even with these improvements it can take over 5 seconds on a desk-top workstation to obtain an

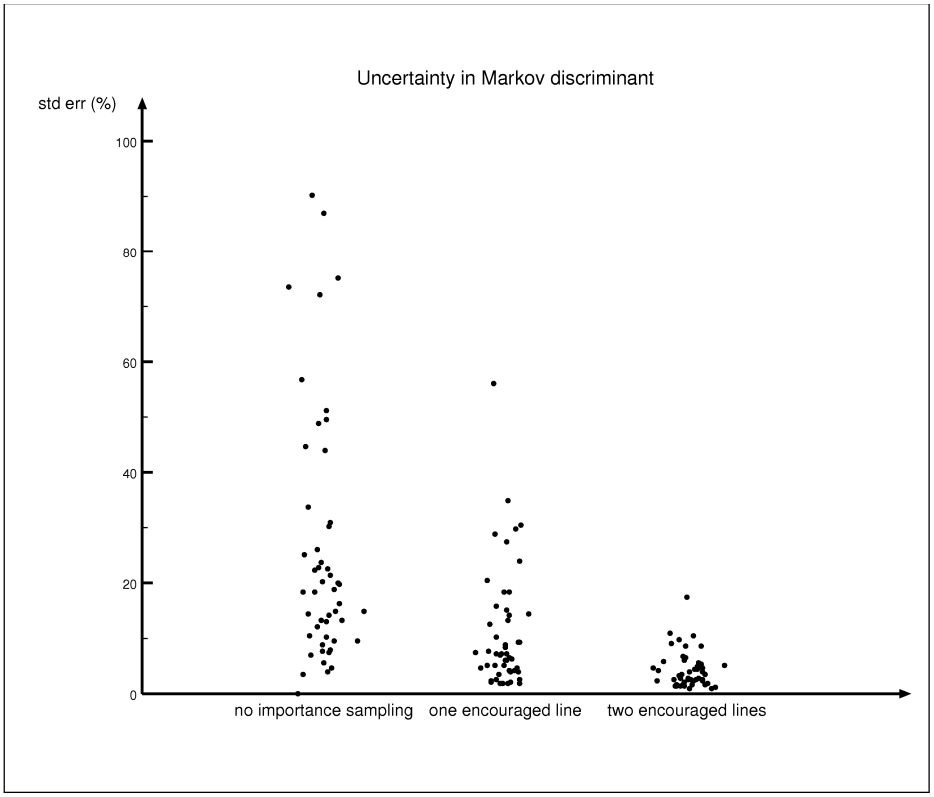


Figure 9: ***Importance sampling improves convergence of the MCMC estimates.*** *The Markov discriminant of the best 50 configurations from the unoccluded coffee mug experiment were estimated in three different ways: (1) standard factored sampling (i.e. no importance sampling) (2) importance sampling with one "encouraged" measurement line selected for each configuration as described in the text (3) importance sampling with two encouraged measurement lines. The uncertainty was found empirically by estimating the discriminant value 20 times for each configuration. On the y-axis is shown the standard error of the 20 estimates, expressed as a percentage of their mean.*

estimate with relative error less than 10%. Uncertainties of this magnitude are acceptable for the outputs described here, but better accuracy might be required for more precise statistical inferences. A related problem is therefore the choice of which configurations to analyse with the Markov discriminant, since it is essential that all significant peaks in the posterior are evaluated. Hence one avenue of future work on this topic will be to investigate ways of judiciously choosing when to use MCMC, and for how long.

The success of the Markov discriminant shows it is possible to apply rigorous probabilistic modelling and Bayesian methods to the classic problem of occlusion. Moreover it represents a significant advance in the potential of recognition systems to provide meaningful statistical information to higher-level systems. It remains to be seen, however, whether any of the current recognition paradigms can realise this potential.

References

[1] A. Amir and M. Lindenbaum. Grouping based non-additive verification. Technical Report 9518, Center for Intelligent Systems, Technion, 1996.

[2] D. Chandler. *Introduction to Statistical Mechanics.* Oxford University Press, 1987.

[3] B. Gidas. A renormalisation group approach to image processing problems. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 11(2):164–180, 1989.

[4] U. Grenander. *Lectures in Pattern Theory I, II and III.* Springer, 1976–1981.

[5] W.E.L. Grimson, D.P. Huttenlocher, and D.W. Jacobs. A study of affine matching with bounded sensor error. In *Proc. 2nd European Conf. Computer Vision*, pages 291–306, 1992.

[6] J.T. Kent, K.V. Mardia, and A.N. Walder. Conditional cyclic Markov random fields. *Adv. Appl. Prob.*, 28:1–12, 1996.

[7] T.K. Leung, M.C. Burl, and P. Perona. Finding faces in cluttered scenes using random graph matching. In *Proc. IEEE PAMI Conf.*, pages 637–644, Cambridge, June 1995.

[8] D.G. Lowe. Robust model-based motion tracking through the integration of search and estimation. *Int. J. Computer Vision*, 8(2):113–122, 1992.

[9] J.P. MacCormick and A. Blake. A probabilistic contour discriminant for object localisation. In *Proc. 8th Int. Conf. Computer Vision*, Jan 1998.

[10] G. Nicholls. Personal communication. 1997.

[11] P. Perez and F Heitz. Restriction of a Markov random field on a graph and multiresolution statistical image modelling. *IEEE Trans. Information Theory*, 42(1):180–190, 1996.

[12] B.D. Ripley. *Stochastic simulation.* New York: Wiley, 1987.

[13] C. Rothwell. Reasoning about occlusions during hypothesis verification. In *Proc. 4th European Conf. Computer Vision*, pages 599–609, April 1996.

[14] I. Shimshoni and J. Ponce. Probabilistic 3D object recognition. In *Proc. 5th Int. Conf. Computer Vision*, pages 488–493, 1995.

[15] G. Winkler. *Image analysis, random fields and dynamic Monte Carlo methods.* Springer, 1995.