J Sign Process Syst
DOI 10.1007/s11265-009-0349-0

Markov Random Fields with Asymmetric Interactions for Modelling Spatial Context in Structured Scene Labelling

Daniel Heesch · Maria Petrou

Received: 6 February 2008 / Revised: 29 January 2009 / Accepted: 2 February 2009
© 2009 Springer Science + Business Media, LLC. Manufactured in The United States

Abstract In this paper we propose a Markov random field with asymmetric Markov parameters to model the spatial and topological relationships between objects in structured scenes. The field is formulated in terms of conditional probabilities learnt from a set of training images. A locally consistent labelling of new scenes is achieved by relaxing the Markov random field directly using these conditional probabilities. We evaluate our model on a varied collection of several hundred hand-segmented images of buildings. The incorporation of spatial information is shown to improve greatly the performance of some trivial classifiers.

Keywords MRF · Scene labelling · Random field relaxation · Machine learning · Contextual vision · Spatial context

1 Introduction

Recent years have seen notable improvements in the performance of object classifiers. Greater robustness against occlusion and intraclass variability has been achieved by describing objects by a large number of local and largely view-invariant features (e.g. [6, 19, 20, 23]). For a specific class (e.g. face detection), efficient classification methods, such as boosting, allow recognition to be in real-time (e.g. [22]). Some of these models have the additional benefit of biological plausibility. The hierarchical feed-forward architecture of [18], for example, aims to mimic the ventral stream of visual information processing and is able to predict with great accuracy whether or not an object is present in a scene.

It seems, however, that, in order to be able to scale to the several thousands of categories humans discriminate without effort, appearance based object classification needs to be complemented by techniques that utilise contextual information. Context may be described as any dependency between the object to be recognised and everything else in the scene, be this other objects or the scene as a whole. Experimental evidence suggests that humans do exploit both types of dependency during object recognition. It is well established, for example, that the nature of a scene can be recognised based on low spatial frequency information [13]. Recent neuro-imaging studies support the view that low spatial frequencies are processed in the cortex at a very early stage during visual recognition [2]. Spatial relationships between objects are one example of contextual information that is captured by the low frequency content of an image. In fact, Bar and Aminoff in [1] establish "early activation of cortical context networks" that appear to store spatial relationships. Spatial context may thus play a key role as an early facilitator during object recognition.

The purpose of this work is to demonstrate the importance of contextual vision over vision based on the use of individual characteristics, using a scaled down version of a visual recognition problem. Our goal is to learn these spatial and topological relationships from a set of hand-labelled data and to utilise this information in a Markov random field (MRF) model to achieve a consistent labelling of new scenes. The MRF is defined

D. Heesch · M. Petrou (✉)
Communications and Signal Processing Group,
Electrical and Electronic Engineering, Imperial College
London, Exhibition Road, London SW7 2AZ, UK
e-mail: maria.petrou@imperial.ac.uk

Published online: 24 February 2009

not over a pixel array, but over the set of regions that correspond to objects. From training data we learn the probability distribution over labels for a region, given the objects in its local neighbourhood. These supply the conditional probabilities that define the MRF and are used during an iterative relaxation scheme to find a probable realisation given the structural relationships observed in a new scene.

Unlike the MRFs hitherto used in computer vision, the MRFs we use here have asymmetric interactions (A influences B differently from how B influences A), which makes it difficult, if not impossible, to express them in terms of cliques and a global cost function. Such asymmetric MRFs are characteristic of natural complex systems and they may be used to model, for example, the interaction between neurons in the human brain [9, 10], population dynamics or company interactions in computational economics [15]. For this reason, we do not define a global cost function, but we use instead a relaxation method [14] appropriate for producing the states of such an MRF and a criterion that allows us to select the right state.

We validate our approach on a set of more than 250 photographs of buildings that were manually segmented and labelled. This domain is particularly interesting as it exhibits sufficiently tight structural constraints to benefit from our approach, and a fair amount of structural variability to challenge it.

This paper is structured as follows. Section 2 presents related work. Section 3 introduces the asymmetric Markov Random Field model. Section 4 details how it is used to label new scenes. Section 5 describes a series of experiments to validate our approach. Section 6 concludes the paper.

2 Related Work

We consider here related works that are concerned with modelling peer-to-peer, rather than hierarchical, dependencies. A natural choice for probabilistic modelling of local dependencies are Markov random fields [11, 14], defined either on a segmentation of the image as in [4, 12] or on a rectangular grid as in [7, 8, 19]. The authors in [7] and [19] define a conditional random field over individual pixels. In [19], contextual information is incorporated by using the joint boosting algorithm [21] for learning potential functions and by employing a novel feature that captures local dependencies in appearance. Neither work considers explicitly spatial relationships, although in [7] the absolute position of a site is included in the potential function.

In [4], it is assumed that training images are associated with a bag of words with no explicit mapping between regions and terms. This renders the learning task more difficult but makes it easier to get hold of large amounts of training data. The MRF is specified through single and pair-wise clique potential functions learnt from the data. To make the estimation problem tractable, potential functions are symmetric with respect to their arguments (labels of adjacent image regions). The model does not capture asymmetric dependencies, nor does it take into account spatial relationships.

In [12], an MRF is defined over image regions by specifying the clique functions for all types of single and pair-wise cliques. The potential functions are taken to be a weighted sum of m basis functions, the parameters of which are set manually.

Our objectives are similar to those in [4] and [12]. Unlike those two, however, we allow neighbouring blobs to influence each other differently depending on their relative spatial position. Our model consists of conditional probabilities that are learnt directly from the data using structural information as can be obtained from the low spatial frequency content of an image.

3 The Model

3.1 Asymmetric MRF

Let $S = \{1, \ldots, N\}$ index a set of regions in an image. We assume that each region is associated with a random variable f_i which takes its value from a discrete set of class labels. The field $F = \{f_i : i \in S\}$ is assumed to be Markovian in the sense that the probabilistic dependencies among f_i are restricted to spatial neighbourhoods $\mathcal{N}_i$, that is,

$$P(f_i | f_{S-i}, R) = P(f_i | f_{\mathcal{N}_i}, R_i) \tag{1}$$

where R denotes the matrix of pair-wise spatial relationships between regions, and R_i the row pertaining to region i. We assume, therefore, that the conditional dependencies depend not only on the identity of the neighbouring regions but also on their relative spatial relationships with the ith region. This is an important component of our model as it allows us to capture the non-isotropic nature of many scenes. For convenience, we refer to a particular observation pair $(f_{\mathcal{N}_i}, R_i)$ as the *neighbourhood configuration* or simply *configuration*, and to the ith region associated with it as the *focal region*. These concepts are demonstrated in Fig. 1.

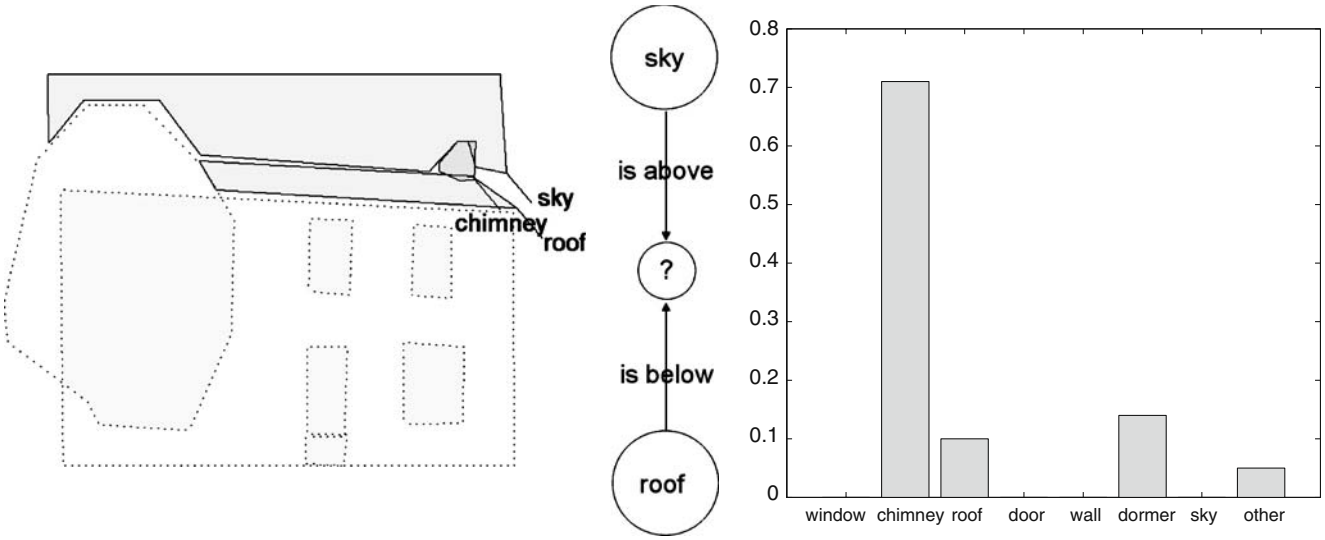

Figure 1 A particular configuration associated with a chimney. On the *left*, a schematic representation of the configuration $(f_{\mathcal{N}_i}, R_i)$ shown in the *middle* and the conditional probability distribution over all labels associated with that configuration, $P(f_i|f_{\mathcal{N}_i}, R_i)$, as obtained from training images, shown on the *right*. The distribution tells us that a region below sky and above a roof is a chimney with 71% probability, but it may also be a dormer with 14% probability, or another roof with 10% probability.

3.2 Neighbourhoods

Since we need to learn the conditional distributions from a relatively small training set, we limit the neighbourhood to at most six regions: the neighbour above, below, to the left and to the right of region i, as well as the regions containing and being contained by region i. The neighbourhood relation is reciprocal and two regions are neighbours if they are separated by no more than a certain distance threshold. The distance between two regions $A, B \subset \mathcal{R}^2$ is computed as

$$d(A, B) = \sum_{i\in\{x,y\}} \min_{a\in A, b\in B} |a_i - b_i| \tag{2}$$

where a_x represents the x coordinate of point a. Other choices of a distance function are of course conceivable. This particular one has the effect that two regions need not be the same to have a zero distance but may be overlapping, exactly adjacent or contained in one another. For example, a wall that surrounds a number of windows has a zero distance from each of them. If regions are non-overlapping, the distance along each direction is given by the smallest Euclidean distance between any two points of the two regions. This has the advantage that the distance between two regions is not affected by their respective sizes (as would have been the case under many other measures such as the Hausdorff metric). For a distance cutoff of 0, the neighbourhood consists of all regions the bounding boxes of which overlap with or touch the focal region. Were the regions regularly arranged like pixels, the resulting neighbourhood would have been the familiar 8-pixel neighbourhood. The optimal distance cutoff is learnt through cross-validation. The left panel of Fig. 2 depicts the distribution over configuration sizes for a distance cutoff of zero. The right panel illustrates how the configurations become larger as the distance cutoff increases.

Given a distance cutoff, the conditional probability distributions (Eq. 1) are learnt by noting for each region i, observed in a set of training images, its corresponding configuration $(f_{\mathcal{N}_i}, R_i)$. In other words, each region i contributes its observed neighbourhood configuration. We then count how many configurations of a certain combination of neighbourhood-relations/labels we observe and compute frequencies of observation from them. The results can conveniently be stored in the form of a hash table with the key being a particular configuration and the value being the conditional probabilities over labels for the focal region. Given a region with known neighbourhood configuration, we can thus rapidly obtain a probability distribution over labels at the focal region. To ensure that the joint distribution of the MRF is nowhere zero, we add a small positive value to each zero-valued conditional probability and subsequently normalise.

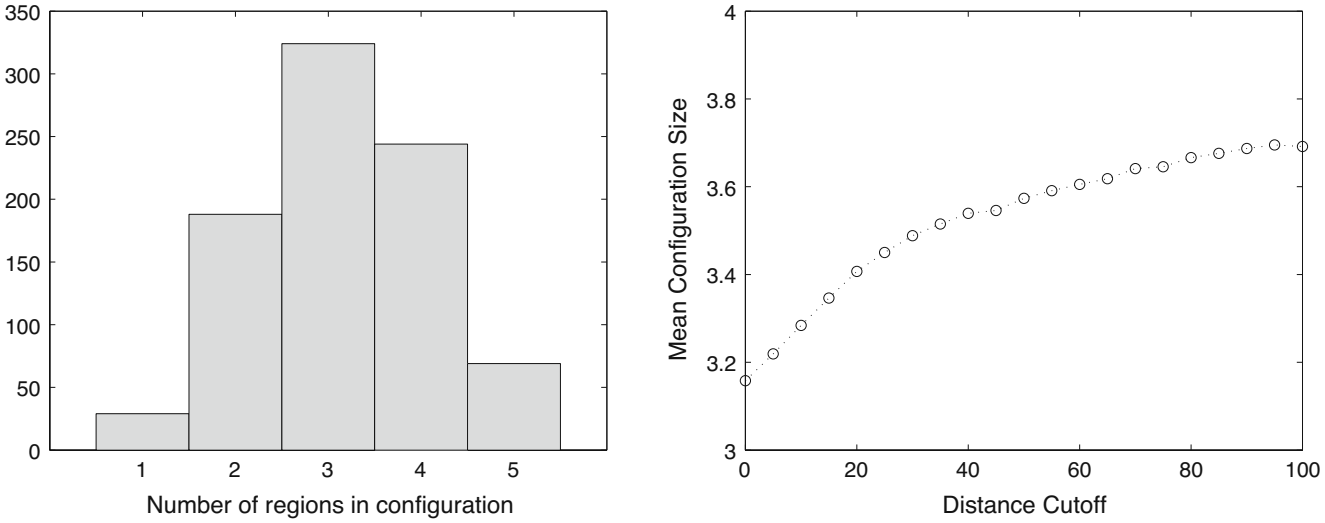

Figure 2 Frequency distribution of different configuration sizes for a distance cutoff of zero (*left*). As we increase the distance cutoff, the configurations become larger (*right*). Note that the maximum allowed neighbourhood size is 6.

4 Labelling of New Scenes

This section details how to obtain probable realisations of the MRF given a new scene. We make the assumption that scenes have been segmented into regions such that each region corresponds to an object to be recognised. How these regions are obtained in the first place is a problem in its own right and outside the focus of this work. We shall simply take it for granted that an appropriate segmentation has been achieved.

4.1 Global Versus Local Relaxation

In the definition of an MRF *accepted by the computer vision community*, an MRF is always equivalent to a Gibbs distribution, and therefore, a global cost function may be written, which, when minimised, produces a realisation of the MRF[1]. When, however, the MRF corresponds to a directed graph, i.e. when the interactions between units are asymmetric, this equivalence is no longer valid. In that case, the MRF may be allowed to emerge only through local updates, and the realisation that results is no longer unique [14].

A standard technique to find a probable realisation of an MRF is simulated annealing [14] which allows a stochastic label update at a site to be retained with a certain probability P_r even if the new realisation of the field is less probable. By letting P_r converge to zero, the field eventually settles at a maximum of the joint probability distribution. In other words, simulated annealing strives to find solutions that are globally maximally consistent.

Because of the difficulty of defining cliques, our asymmetric field is formulated purely in terms of local, conditional probability distributions (Eq. 1). We aim to find labellings that are locally consistent by repeatedly sampling from these conditional distributions, in an iterative way:

$$f_i^{(n+1)} \sim P(f_i | f_{\mathcal{N}_i}^{(n)}, R_i). \tag{3}$$

4.2 Graph Colouring

In order to iteratively update regions based on the current labelling of their neighbourhood, we partition the set of regions into a set of codings. The idea of a coding was first introduced by Besag [3] in the context of the iterated conditional mode algorithm for MRF parameter estimation. A coding is equivalent to the concept of

[1]Note that although a directed graph *for the computer vision community* cannot be called an MRF, and as a consequence, according to the same community, asymmetric MRFs do not exist, from the mathematical point of view the concept is not only possible, but, in our opinion, also prevalent in nature: if the value of a node is assumed to be the outcome of a random experiment, we are clearly dealing with a random field; if we model this outcome by considering only the neighbours (defined in some pre-specified sense) of the node, we are clearly dealing with a Markovian random field. Nowhere in these two definitions enters the concept of directionality. We consider the assertion that a directed graph cannot be a Markov random field rather idiosyncratic of the computer vision community and not valid if strict mathematical definitions are applied. That is why in this paper we use the mathematically correct terms "asymmetric MRFs" or "non-Gibbsian MRFs".

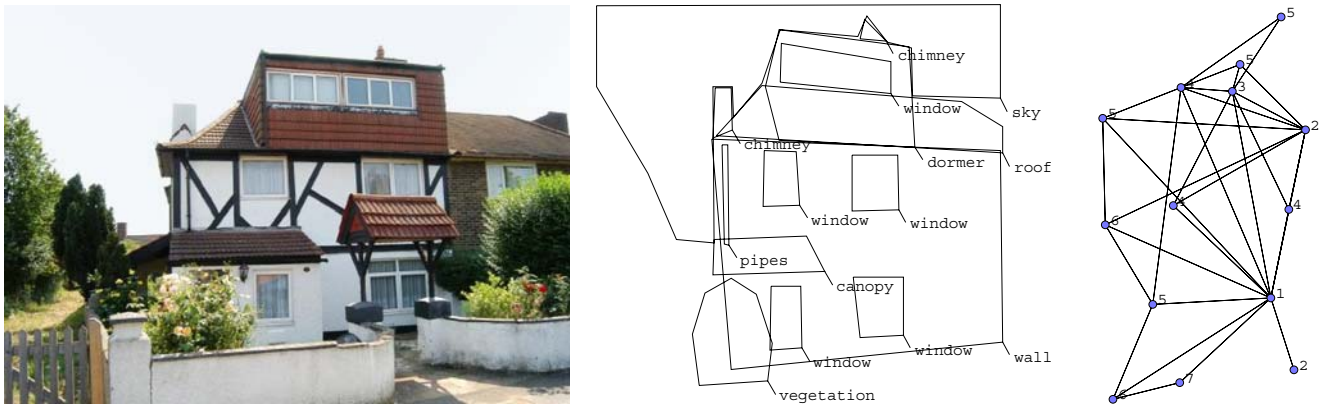

Figure 3 Original image (*left*). Hand-segmented and hand-labelled training image (*middle*). Vertex colouring of the neighbourhood graph (*right*): vertices with the same number have non-overlapping neighbourhoods.

a vertex colouring of a graph, that is, it constitutes a partitioning of the set of vertices (= regions) so that no two adjacent vertices (= neighbouring regions) belong to the same partition. Because of the assumption of Markovianity, the likelihood over vertices of the same colour reduces to a simple product of the respective conditional probabilities. Note that coding or colouring is necessary for the relaxation of the field, so that the updating of the label of a region is not negated by the updating of the label of a neighbouring region. We employ a greedy colouring strategy, in which vertices are visited in order of decreasing vertex degree (i.e. number of neighbours). Each vertex is assigned the first possible colour from a list of colours. One example of a colouring is given in Fig. 3. The wall has the largest number of neighbours and is correspondingly assigned the first colour ("1").

4.3 Choosing a Solution

Regions are updated within each coding by retrieving and sampling from the probability distribution corresponding to that region's current neighbourhood configuration. For this simple updating scheme, convergence cannot be guaranteed and indeed the field tends to oscillate between different solutions. To let the field settle on a particular solution, we compute the joint probability for all vertices of the same colour $\mathcal{C}_j$

$$P(f_{\mathcal{C}_j}|R) = \prod_{i\in\mathcal{C}_j} P(f_i|f_{\mathcal{N}_i}, R). \tag{4}$$

An estimate of the overall probability of the data may be obtained by averaging the value of $P(f_{\mathcal{C}_j}|R)$ over all codings. To account for the fact that codings are generally of different size, we do not use the arithmetic mean, but:

$$P(f_1, \ldots, f_N) \approx \frac{1}{N}\sum_j |\mathcal{C}_j| \left[\prod_{i\in\mathcal{C}_j} P(f_i|f_{\mathcal{N}_i}, R)\right]^{\frac{1}{|\mathcal{C}_j|}}. \tag{5}$$

Some explanation for the choice of this formula is necessary here. As we cannot write a global cost (or energy) function for the asymmetric MRF, we have to devise a pseudo-likelihood function (corresponding to a pseudo-energy function), which will guide us when performing our relaxation. The product that appears on the right-hand side of Eq. 5 is over all nodes of a single colouring of the graph. As these nodes have non-overlapping neighbourhoods, their conditional probabilities are independent and when multiplied they produce the joint probability density function of the combination of labels assigned to this colouring. For each colouring of the graph we have a different such joint probability density function of its combination of labels. We have then somehow to find a way to combine all of them, in order to define our objective function, which, as explained above, does not exactly correspond to the true energy function of the system, since our system is not integrable. However, the objective function we shall define this way will guide us into relaxing the MRF into one of its possible states. As each colouring j of the graph may contain a different number of nodes, $|\mathcal{C}_j|$, each $\prod_{i\in\mathcal{C}_j} P(f_i|f_{\mathcal{N}_i}, R)$ is the product of $|\mathcal{C}_j|$ factors and in order to remove that dependence we take the $|\mathcal{C}_j|$ root of it. Then we sum all these partial likelihood terms weighted by the fraction of nodes $|\mathcal{C}_j|/N$ each represents. The updating now is performed as follows. Let p be the ratio between the estimated joint probability after and before the update. We accept the change with probability 1 if $p > 1$ and

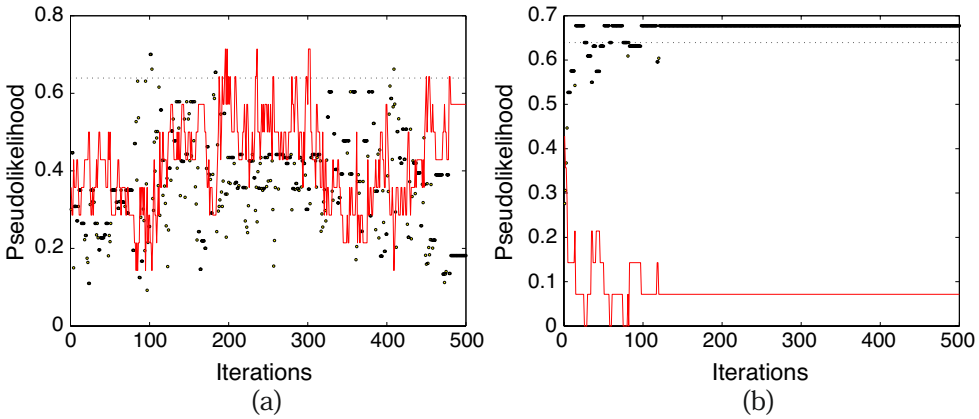

Figure 4 Dynamics of the stochastic updating process with and without maximisation of the pseudo-likelihood $P(f_i, \ldots, f_N)$ (Eq. 5).

with probability $p^{\frac{1}{T}}$ otherwise. T is the temperature parameter the value of which decreases exponentially with time. Figure 4 shows an example of how the value given by Eq. 5 increases over successive iterations. One iteration here involves the update of the labels of all regions. In both graphs the dotted line marks the value of the pseudo-likelihood associated with the true labelling. The continuous line shows the proportion of misclassified regions. During the relaxation process, the labels of the regions are updated based on the conditional probabilities. In both graphs, the points are the values of the pseudo-likelihood we wish to maximise by this process. In (a), a new labelling is always accepted. In (b), a labelling is only accepted when it improves the current value of the pseudo-likelihood or when it is worse by no more than a value that decreases with time.

5 Experiments

For our experiments, we collected 253 images of buildings from the World Wide Web. Each image was manually segmented into regions that corresponded to parts of the building or parts of the environment, such as sky or vegetation. Each region was labelled by hand using an annotation tool similar to LabelMe. The complete dataset contains nearly 6,000 regions covering a dozen classes.[2]

We allow for the following seven labels (with respective frequencies): "window" (0.507), "chimney" (0.054),"roof" (0.053), "door" (0.087), "wall" (0.089), "dormer" (0.015), "sky" (0.055), "other" (0.14). The "other" label aggregates all remaining structures that were annotated (e.g. "pipes" and "balcony"). We report the performance of the various algorithms we tried in terms of classification accuracy, i.e. the proportion of regions that have been labelled correctly. To estimate how an algorithm is able to deal with data that it was not trained on, we use the leave-one-out method of cross-validation, i.e. we remove one image from the set at a time and use it as our test image while training on the remaining 252 images.

5.1 Comparison With Other Methods

We compare our asymmetric MRF model with four other classification models, the trivial maximum prior classifier, a non-contextual Bayes classifier, an alternative contextual model, that uses probabilistic relaxation to find a locally consistent labelling, and an isotropic MRF classifier, i.e. a classifier that ignores spatial neighbourhood information.

5.1.1 Maximum Prior

Our simplest benchmark is a maximum prior classifier that assigns to each region the most frequently occurring label (i.e. "window").

5.1.2 Non-contextual Bayes Classifier

As a non-contextual benchmark we implemented a Parzen classifier that classifies regions based on the posterior probabilities given measurements of a number of low-level features from the region. We use a set of three features that can easily be obtained from the low-frequency content of a scene: the mean intensity, the normalised area of the region and its vertical position. For each feature, the posterior probability over classes is given by Bayes' rule with the class-conditional densities being approximated using a Parzen window with

[2]The images alongside their annotation and segmentation information are available at http://www.commsp.ee.ic.ac.uk/~dheesch/ngmrf/data/.

a Gaussian kernel function centred on a set of class exemplars E_c

$$p(x|c) \propto \sum_{x_i : i \in E_c} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{|x - x_i|}{2\sigma^2}\right), \quad (6)$$

where σ is learnt through cross-validation. We assume that each feature is conditionally independent given the class, and thus we compute the overall class probability density as the product of feature-specific posteriors.

5.1.3 Probabilistic Relaxation

Probabilistic relaxation is an alternative contextual labelling technique [5, 17]. The contextual information consists of the conditional probabilities of a label, given that another label is found in a particular relative position to the first. In each iteration of the relaxation process, the label probabilities are updated based on the probabilities at the previous time step, modulated by the support a particular label f_i receives from neighbouring labels,

$$P^{(n+1)}(f_i = c) = \frac{P^{(n)}(f_i = c) Q_i(c)}{\sum_{\mu \in \mathcal{L}} P^{(n)}(f_i = \mu) Q_i(\mu)} \quad (7)$$

with support function

$$Q_i(c) = \sum_{j : f_j \in \mathcal{N}_i} \sum_{\nu \in \mathcal{L}} P(f_i = c | f_j = \nu, r_{ij}) P(f_j = \nu). \quad (8)$$

Here $\mathcal{L}$ denotes the label set. The compatibilities are learnt from the data in a similar way as are the conditional distributions for neighbourhood configurations in the MRF model. Note that unlike the MRF model, which allows configurations to comprise up to six regions, this particular formulation of probabilistic relaxation is limited to binary dependencies. This makes statistical learning of dependencies easier but comes at the expense of limited modelling power.

5.1.4 Isotropic MRF (ISO-MRF)

To assess the added value one gains by considering explicitly the spatial and topological relationships between neighbouring regions, we implemented a simpler Markov random field in which the Markovian dependencies depend only on the labels of neighbouring regions but not on their spatial relationships. The isotropic MRF model thus assumes that a neighbour of a certain focal region has the same effect on the latter regardless of its relative position, i.e. in the notation of Section 3,

$$P(f_i | f_{S-i}, R) = P(f_i | f_{\mathcal{N}_i}). \quad (9)$$

Table 1 Comparison of different methods using 5,682 blobs for training.

Model	Accuracy	Std dev	Known configurations
Max prior	0.521	0.0006	–
Parzen	0.690	0.125	–
Prob rel	0.568	0.134	–
ISO-MRF	0.434	0.145	0.9769
Asym-MRF	0.729	0.124	0.8949

Max prior: each region is given the same, most frequently occurring label; Parzen: non-contextual classification; Prob rel: binary relations only used; ISO-MRF: isotropic Markov random field; Asym-MRF: asymmetric Markov random field. Performance is measured in terms of the fraction of regions classified correctly. The last column gives the fraction of configurations in the test data observed in the training set.

5.1.5 Results

Table 1 shows the results for probabilistic relaxation, the Parzen classifier, the isotropic MRF and the asymmetric MRF models. Note that we use the output of the Parzen classifier to initialise the labelling for the MRF models. In order to assess the variability in performance, we opted for a leave-one-out strategy. The results are the average over 253 images with 5,682 regions.

The best results are obtained by the asymmetric MRF, followed closely by the non-contextual Parzen classifier. The performance of the isotropic MRF is notably worse even in comparison with the simple maximum prior method.

The confusion matrix (Table 2) reveals that the greatest accuracy is achieved for windows. That many other classes are misclassified as windows may be attributed to the strong "window" prior that influences the result through the non-contextual Parzen initialisation. Doors, in particular, are frequently mistaken for

Table 2 Confusion matrix for Asym-MRF labelling.

	wi	ch	ro	do	wa	do	sk	ot
Window	2848	50	5	81	0	0	25	131
Chimney	20	151	50	5	0	5	10	15
Roof	25	20	101	0	30	10	25	76
Door	348	5	0	20	5	0	0	96
Wall	40	0	25	5	292	10	10	91
Dormer	30	15	20	5	5	15	5	0
Sky	15	10	10	0	5	5	192	30
Other	217	15	15	40	30	5	25	343

The top row entries are indexed by the first two letters of the respective label and they correspond to the labels assigned by the classifier. Thus, matrix element a_{ij} gives the number of regions of the ith class that have been classified as belonging to the jth class. The codes along the top row correspond to the first two letters of the labels in the first column.

Table 3 Dependence of contextual classification on initial conditions.

Initialisation scheme	Initial	Asym-MRF
Parzen	0.690	0.729 (0.124)
Max prior	0.521	0.654 (0.127)
Random	0.315	0.621 (0.135)

The second column shows the accuracy after initialisation with three different initialisation schemes. The initial accuracy of the random assignment is $1 - \sum_c p_c(1 - p_c)$ where p_c is the prior of the cth class.

windows as these two classes exhibit very similar spatial relationships with other building parts whilst having markedly different priors.

5.2 Robustness to Initialisation

Apart from the non-contextual initialisation scheme used in the results reported above, we also investigate two other initialisation schemes to assess the robustness of the contextual inference to the initial conditions. The first scheme assigns each region the most frequently occurring label (in this case "window"), the second draws labels randomly from the prior distribution, i.e. it will result in a similar initial distribution of classes within the image but with random assignment of classes to regions. The results are shown in Table 3. While we notice a performance degradation compared with the Parzen initialisation, the contextual model improves over the new baselines of 0.52 and 0.32, when the alternative initialisation schemes are used. So, although the result depends on the initialisation, contextual information always offers an added value to the classifier.

6 Conclusions

We presented a Markov random field model for contextual labelling of objects in structured scenes. In our model the context of a region consists not only of the identity of neighbouring regions but also, crucially, on their relative spatial and topological relationships. By incorporating what are typically asymmetric relationships, the Markov random field is capable of modelling the non-isotropic nature of typical scenes. Because of the asymmetry, it is difficult to define clique potentials. The model is, therefore, formulated directly in terms of conditional distributions that are learnt from training data, rather than in terms of a global Gibbs distribution.

Given a new scene, the Markov random field is relaxed by iteratively sampling from conditional probability distributions. We proposed an objective function to help us identify good labelling solutions. The objective function is based on the vertex colouring of the region neighbourhood graph and is not the global cost function usually associated with Gibbsian MRFs. Comparison with a non-contextual classifier and a contextual classifier that ignores spatial relations demonstrated the advantages of using asymmetric MRFs, with spatial and topological relationships incorporated, for modelling visual context.

This paper was concerned with high level vision, so we did not consider here the problem of actually segmenting the regions that have to be labelled. This is a problem of image processing and actually a significant problem in its own right. Ideally, segmentation and labelling should be integrated in a system with feedback loops, like for example the system proposed in [16].

Acknowledgement This work was supported by the FP6 European project eTRIMS.

References

1. Bar, M., & Aminoff, E. (2003). Cortical analysis of visual context. *Neuron, 38*, 347–358.
2. Bar, M., Kassam, K., Ghuman, A., Boshyan, J., Schmidt, A., Dale, A., et al. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences, 103*(2), 449–454.
3. Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society, B, 36*, 192–236.
4. Carbonetto, P., de Freitas, N., & Barnard, K. (2004). A statistical model for general contextual object recognition. In *Proc European conf computer vision* (pp. 350–362).
5. Christmas, W. J., Kittler, J., & Petrou, M. (1995). Structural matching in computer vision using probabilistic relaxation. *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI, 17*, 749–764.
6. Csurka, G., Bray, C., Dance, C., & Fan, L. (2004). Visual categorisation with bags of keypoints. In *Proc European conf computer vision*, Prague, Czech Republic, May 11–14.
7. He, X., Zemel, R., & Ray, D. (2006). Learning and incorporating top-down cues in image segmentation. In *Proc European conf computer vision*, Graz, Austria, May 7–13.
8. Kumar, S., & Hebert, H. (2003). Discriminative random fields: A discriminative framework for contextual interaction in classification. In *Proc int'l conf computer vision*, Nice, France, October 14–17.
9. Li, Z. (1999). Visual segmentation by contextual influences via intra-cortical interactions in the primary visual cortex. *Networks: Computation in Neural Systems, 10*, 187–212.
10. Li, Z. (2001). Computational design and nonlinear dynamics of a recurrent network model of the primary visual cortex. *Neural Computation, 13*(8), 1749–1780.
11. Li, S. (1995). *Markov random field modeling in computer vision*. New York: Springer.
12. Modestino, J., & Zhang, J. (1992). A Markov random field model-based approach to image interpretation. *IEEE Trans Pattern Analysis and Machine Intelligence, 14*(6), 606–615.

13. Oliva, A., & Torralba, A. (2001). Modelling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision, 42*(3), 145–175.
14. Petrou, M., & Garcia Sevilla, P. (2006). *Image processing: Dealing with texture*. New York: Wiley, ISBN-13: 978-0-470-02628-1.
15. Petrou, M., Gautam, S., & Giannoutakis, K. N. (2006). Simulating a digital business ecosystem. In M. Costantino & C. A. Brebbia (Eds.), *Computational finance and its applications II* (pp. 277–287). Southampton: WIT, ISBN 1-84564-1744.
16. Petrou, M. (2007). Learning in computer vision: Some thoughts. In L. Rueda, D. Mery and J. Kittler (Eds.), *Progress in pattern recognition, image analysis and applications. The 12th Iberoamerican congress on pattern recognition, CIARP 2007, Vina del Mar-Valparaiso, November, LNCS 4756, pp 1–12*. New York: Springer.
17. Rosenfeld, A., Hummel, A., & Zucker, S. (1976). Scene labelling by relaxation operations. *Trans SMC, 6*(6), 420–433.
18. Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorisation. *Proceedings of the National Academy of Sciences of the United States of America, 104*(15), 6424–6429.
19. Shotton, J., Winn, J., Rother, C., & Criminisi, A. (2006). Textonboost: Joint appearance, shape and context modelling for multi-class object recognition and segmentation. In *Proc European conf computer vision*, Graz, Austria, May 7–13.
20. Sivic, J., & Zisserman, A. (2003). Video google: A text retrieval approach to object matching in videos. In *Proc int'l conf computer vision*, Nice, France, October 14–17, pp. 1–8.
21. Torralba, A., Murphy, K., & Freeman, W. (2004). Sharing features: Efficient boosting procedures for multiclass object detection. In *Proc int'l conf computer vision and pattern recognition* (pp. 762–769).
22. Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Proc int'l conf computer vision and pattern recognition*, Kauai, Hawaii, December 9–14.
23. Winn, J., Criminisi, A., & Minka, T. (2005). Object categorisation by learnt universal visual dictionary. In *Proc int'l conf computer vision* (pp. 1800–1807).

Daniel Heesch has until 2006 worked as a computer vision researcher in the Communications and Signal Processing Group at Imperial College London. His first degrees are in Biology (Oxford University, 2000) and Mathematics (Open University, 2005). He was awarded his PhD in Computer Science from Imperial College London for his study of network structures for interactive image retrieval. His main interests lie in information retrieval, computer vision and mathematical biology. Daniel is cofounder and research director of Pixsta Ltd, a London-based startup company he had set up to foster and commercialise image retrieval research.

Maria Petrou studied Physics at the Aristotle University of Thessaloniki, Greece, Applied Mathematics in Cambridge and she did her PhD in the Institute of Astronomy in Cambridge, UK. She obtained her DSc also from Cambridge in 2009. She is currently the Professor of Signal Processing and the Head of the Communications and Signal Processing Group at Imperial College London. She has published more than 350 scientific papers, on Astronomy, Remote Sensing, Computer Vision, Machine Learning, Colour analysis, Industrial Inspection, and Medical Signal and Image Processing. She has co-authored two books "Image Processing: the fundamentals" and "Image Processing: Dealing with texture" both published by John Wiley in 1999 and 2006, respectively. She has supervised to successful completion 39 PhD theses, and examined more than one hundred. She is a Fellow of the Royal Academy of Engineering, Fellow of IEE, Fellow of IAPR, Fellow of the Institute of Physics, Senior member of IEEE and a Distinguished Fellow of the British Machine Vision Association.