INTERPRETING TERRESTRIAL IMAGES OF URBAN SCENES USING DISCRIMINATIVE RANDOM FIELDS

Filip Korč and Wolfgang Förstner

Department of Photogrammetry
Institute of Geodesy and Geoinformation
University of Bonn
Nussallee 15, Bonn 53115, Germany
filip.korc@uni-bonn.de, wf@ipb.uni-bonn.de
http://www.ipb.uni-bonn.de/

KEY WORDS: Markov random field, parameter learning, pseudo-likelihood, man-made structure detection

ABSTRACT:

We investigate Discriminative Random Fields (DRF) which provide a principled approach for combining local discriminative classifiers that allow the use of arbitrary overlapping features, with adaptive data-dependent smoothing over the label field. We discuss the differences between a traditional Markov Random Field (MRF) formulation and the DRF model, and compare the performance of the two models and an independent sitewise classifier. Further, we present results suggesting the potential for performance enhancement by improving state of the art parameter learning methods. Eventually, we demonstrate the application feasibility on both synthetic and natural images.

1 INTRODUCTION

This paper presents an investigation into the statistical modeling of spatial arrangements in the context of image understanding or semantic scene interpretation.

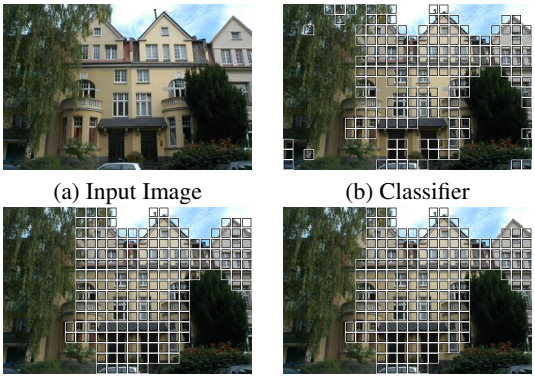
Our goal is the interpretation of the scene contained in an image as a collection of semantically meaningful regions. We are specifically interested in the interpretation of terrestrial images in build-up areas, showing man-made structures, vegetation, sky and other more specific objects. The result of such an interpretation could be a rich image description useful for 3D-city models of high level of detail. In this paper we focus on the problem of binary classification of image regions, especially on detecting man-made structures.

It has been argued that the incorporation of spatial dependencies in the image interpretation task is vital for improved performance. Markov Random Fields (MRF) allow modeling local contextual constraints in labeling problems in a probabilistic framework and since the early work on stochastic fields by (Besag, 1974), the pioneering on stochastic algorithms in MRF's by (Geman and Geman, 1984) and the work by (Modestino and Zhang, 1992) on MRF-based image interpretation, MRF is the most commonly used model for modeling spatial interactions in image analysis (Li, 2001).

MRFs are generally used in a *generative* framework that models the joint probability of the observed data and the corresponding labels. In other words, let $\mathbf{y}$ denote the observed data from an input image, where $\mathbf{y} = \{\mathbf{y}_i\}_{i \in S}$, $\mathbf{y}_i$ is the data from the ith site, and S is the set of sites. Let the corresponding labels at the image sites be given by $\mathbf{x} = \{x_i\}_{i \in S}$. In the considered generative MRF framework, the posterior over the labels given the data is expressed using the Bayes' rule as,

$$P(\mathbf{x}|\mathbf{y}) \propto P(\mathbf{x}, \mathbf{y}) = P(\mathbf{y}|\mathbf{x})P(\mathbf{x})$$

The prior over labels, $P(\mathbf{x})$ is modeled as a MRF. The observation model, $P(\mathbf{y}|\mathbf{x})$ is described in the following.



(a) Input Image (b) Classifier (c) MRF (d) DRF

Figure 1: Image interpretation. (a) Input image. (b - d) Man-made structure detection result for different methods. DRF improves the detection rate and reduces the false positive rate. Man-made structure is denoted by bounding squares superimposed on the input image.

Let us consider a common MRF formulation of binary classification problems. The labels are assumed to be $\mathbf{x_i} \in \{-\mathbf{1}, \mathbf{1}\}$ and the label interaction field $P(\mathbf{x})$ is assumed to be a homogeneous isotropic 2D lattice, thus an Ising model. Let us further assume, that the observation or likelihood model $P(\mathbf{y}|\mathbf{x})$ has a factorized form, i.e., $P(\mathbf{y}|\mathbf{x}) = \prod_{i \in S} P(\mathbf{y}_i|x_i)$ (Besag, 1986) (Li, 2001). Then the posterior distribution over labels can be written as,

$$P(\mathbf{x}|\mathbf{y}) = \frac{1}{Z} \exp\left(\sum_{i \in S} \log P(\mathbf{y}_i|x_i) + \sum_{i \in S} \sum_{j \in N_i} \beta x_i x_j \right) \quad (1)$$

where β is the interaction parameter of the MRF, $\mathbf{y}_i$ is data from a single image site i, N_i are the neighbors of site x_i and Z is the normalization constant, the partition function. We note that even though only the label prior $P(\mathbf{x})$ is assumed to be a MRF, captured in $\beta x_i x_j$, the assumption of conditional independence of data implies that the posterior given in Eq. (1) is also a MRF. If the conditional independence assumption is not used, the pos-

terior will generally not be a MRF making the inference difficult. Hence, for computational tractability data are generally assumed to be, conditioned on the labels, independent. In the following, we refer to the above described MRF model as the conventional MRF model.

Following the previous reasoning, let us now consider the task of detecting man-made structures in images of urban scenes. See Fig. 1 for an example. Looking at textures, lines, corners, etc., that are present in the image, we realize that given the man-made structure class label, the data is still highly dependent on its neighbors, e.g., edges and textures cover larger areas, than just a single image sites. Hence, the assumption of conditional independence indeed is restrictive.

In the context of classification, we are generally interested in estimating the posterior over labels given the observations, i.e., $P(\mathbf{x}|\mathbf{y})$. In the generative framework, we model the joint distribution $P(\mathbf{x}, \mathbf{y})$, which involves modeling additional complexity not relevant to the considered classification task. Simplifying assumptions are often needed to deal with the resulting complexity. In addition, learning full probabilistic model is hard if little training data is available. In a *discriminative* framework, on the contrary, the distribution $P(\mathbf{x}|\mathbf{y})$ is modeled directly. In this work, we describe a model called Discriminative Random Field (DRF), introduced by Kumar (Kumar and Hebert, 2003), that directly models the posterior distribution as a MRF.

The paper is organized as follows: After a review on recent related work, we describe the concept, parameter learning and the inference using DRF's. We investigate different learning strategies, namely pseudo-likelihood with and without prior information on the parameters. The central part of our investigation is the comparison of different inference methods, namely logistic classification, iterated conditional modes algorithm and max-flow/min-cut type of algorithm, which show to yield better results than shown in (Kumar and Hebert, 2004a). The application to real data shows the feasibility of the approach for detecting man-made structures in terrestrial images.

2 RELATED WORK

DRFs have been introduced in (Kumar and Hebert, 2003). The DRF models are based on the concept of Conditional Random Fields (CRF) (Lafferty et al., 2001) that have been proposed in the context of segmentation and labeling of 1D text sequences.

DRFs have later been modified in (Kumar and Hebert, 2004a) in a way that leads to model parameter learning as a convex optimization problem. DRFs model have been extended for multiclass problems for parts-based object detection (Kumar and Hebert, 2004b). Learning - inference coupling is studied in (Kumar et al., 2005). Further, formulation with hierarchical interactions can be found in (Kumar and Hebert, 2005). Learning in DRF can be accelerated using Stochastic Gradient Methods (Vishwanathan et al., 2006) and Piecewise Pseudo-likelihood (Sutton and McCallum, 2007).

In addition, CRF can formulated using hidden states (Quattoni et al., 2004, Wang et al., 2006, Quattoni et al., 2007). The concept of CRFs can be employed for incorporating a semantic object context (Rabinovich et al., 2007). Modeling temporal contextual dependencies in video sequences is described in (Sminchisescu et al., 2005). A semi-supervised learning approach to learning in CRF can be found in (Lee et al., 2007). Last, pseudo-likelihood based approximations are investigated in (Korč and Förstner, 2008).

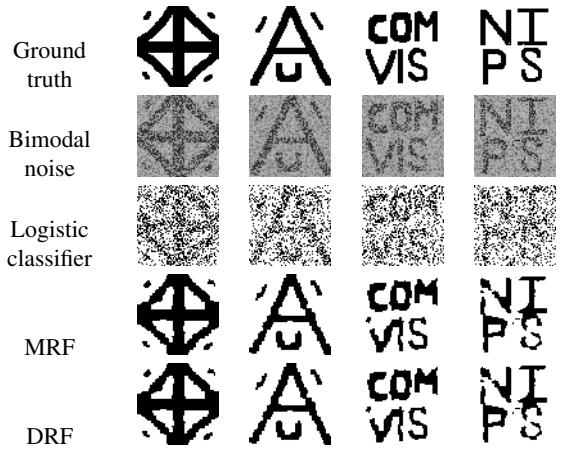


Figure 2: Image restoration. From top, row 1: ground truth images, row 2: input images (ground truth images corrupted with bimodal noise), row 3: logistic classifier result, row 4: MRF result, row 5: DRF result.

3 DISCRIMINATIVE RANDOM FIELD

Let us review the formulation of DRFs and discuss the model in the context of binary classification on 2D image lattices. A general formulation on arbitrary graphs with multiple class labels is described in (Kumar and Hebert, 2004b). Recalling notation introduced in the previous section, where $\mathbf{x} = \{x_i\}_{i \in S}$ denotes labels at image sites i, we now have $x_i \in \{-1, 1\}$ for a binary classification problem. The DRF model combines local discriminative models to capture the class association at individual sites with the interactions in the neighboring sites as:

$$P(\mathbf{x}|\mathbf{y}) = \frac{1}{Z} \exp\left(\sum_{i \in S} A_i(x_i, \mathbf{y}) + \sum_{i \in S} \sum_{j \in N_i} I_{ij}(x_i, x_j, \mathbf{y}) \right) \tag{2}$$

where Z is the partition function (normalizing constant) and N_i is the set of neighbors of the image site i. Let us note that, as opposed to the conventional MRFs, both the unary association potential, $A_i(x_i, \mathbf{y}) = \log P'(x_i|\mathbf{y})$, and the pairwise interaction potential, $I_{ij}(x_i, x_j, \mathbf{y}) = \log P''(x_i, x_j|\mathbf{y})$, depend explicitly on *all* the observations $\mathbf{y}$. The restrictive assumption of conditional independence of data, made in the conventional MRFs, is thus relaxed. Further, unlike conventional generative MRFs, where the pairwise potential is a data-independent prior over the labels, the pairwise potential in DRFs depend on the data $\mathbf{y}$ and allows thus *data-dependent* interaction among the labels. We encourage the reader to study the differences of the MRF model in Eq. (1) and the DRF model in Eq. (2) as they form the basis of our discussion.

In Eq.(2), $P'(x_i|\mathbf{y}) = \exp(A_i(x_i, \mathbf{y}))$ and $P''(x_i, x_j|\mathbf{y}) = \exp(I_{ij}(x_i, x_j, \mathbf{y}))$ are arbitrary unary and pairwise discriminative classifiers. This gives us freedom to choose any domain specific method to identify object classes (P') or neighborhood dependencies (P''), especially classifiers which use any type of features, especially image features which exploit possibly overlapping neighborhoods of the site in concern, or even depending on global image characteristics.

Let us denote the unknown DRF model parameters by $\theta = \{\mathbf{w}, \mathbf{v}\}$, the parameters $\mathbf{w}$ specifying the classifier for individual sites, the parameters $\mathbf{v}$ specifying the classifier for site neighborhoods. In this paper, as in (Kumar and Hebert, 2004a), we use a logistic function to specify the local class posterior, i. e.

$$P'(x_i|\mathbf{y}) = \sigma(x_i \mathbf{w}^T \boldsymbol{h}_i(\mathbf{y})) \tag{3}$$

where $\sigma(t) = 1/(1 + e^{-t})$. Here, $\boldsymbol{h}_i(\mathbf{y})$ is a sitewise feature vector, which has to be chosen such that a high positive weighted sum $\mathbf{w}^T\boldsymbol{h}_i(\mathbf{y})$ supports class $x_i = 1$. Similarly, to model $P''(x_i, x_j|\mathbf{y})$ we use a pairwise classifier of the following form: $P''(x_i, x_j|\mathbf{y}) = x_i x_j \mathbf{v}^T \boldsymbol{\mu}_{ij}(\mathbf{y})$. Here, $\boldsymbol{\mu}_{ij}(\mathbf{y})$ is a feature vector similarly being able to support or suppress the identity $x_i x_j = 1$ of neighbored classes.

It is helpful to note that by ignoring the dependence of the pairwise potential on the observed data $\mathbf{y}$, we obtain the conventional MRF smoothing potential, $\beta x_i x_j$ in Eq. (1), known as the Ising model.

In the following we assume the random field in Eq. (2) to be homogeneous and isotropic, i.e., the functional forms of A_i and I_{ij} are independent of the locations i and j. Henceforth we will leave the subscripts and use the notation A and I.

3.1 Parameter Learning

In the MRF framework the parameters of the class generative models, $p(\mathbf{y}_i|x_i)$ and the parameters of the prior label interaction field, $P(\mathbf{x})$ are generally assumed to be independent. Therefore they are learned, i. e. estimated from training data, separately (Li, 2001). In DRFs, on the contrary, all the model parameters have to be learned *simultaneously*.

We learn the parameters θ of the DRF model in a supervised manner. Hence, we use training images and the corresponding ground-truth labeling. As in (Kumar and Hebert, 2004a), we use the standard maximum likelihood approach and, in principle, maximize the conditional likelihood $P(\mathbf{x}|\mathbf{y}, \theta)$ of the DRF model parameters. However, this would involve the evaluation of the partition function Z which is in general NP-hard. To overcome the problem, we may either use sampling techniques or approximate the partition function. As in (Kumar and Hebert, 2004a), we use the pseudo-likelihood (PL) approximation $P(\mathbf{x}|\mathbf{y}, \theta) \approx \prod_{i \in S} P(x_i|\mathbf{x}_{N_i}, \mathbf{y}, \theta)$ (Besag, 1975),(Besag, 1977), which is characterized by its low computational complexity.

It has been observed (Greig et al., 1989), that in the case of the Ising MRF model, this approximation tends to overestimate the interaction parameter β, causing the MAP estimate of the field to be a poor solution. Same observation has been made for the interaction parameters in the DRFs, (Kumar and Hebert, 2004a). To overcome the difficulty, they propose to adopt the Bayesian viewpoint and find the maximum a posteriori estimates of the parameters by assuming a Gaussian prior over the parameters such that $P(\theta|\tau) = \mathcal{N}(\theta|\mathbf{0}, \tau^2\mathbf{I})$ where $\mathbf{I}$ is the identity matrix.

Thus, given M independent training images, we determine θ from

$$\hat{\theta}^{ML} \approx \underset{\theta}{\operatorname{argmax}} \prod_{m=1}^{M} \prod_{i \in S} P(x_i^m|\mathbf{x}_{N_i}^m, \mathbf{y}^m, \theta) P(\theta|\tau) \quad (4)$$

where

$$P(x_i|\mathbf{x}_{N_i}, \mathbf{y}, \theta) = \frac{1}{z_i} \exp\{A(x_i, \mathbf{y}) + \sum_{j \in N_i} I(x_i, x_j, \mathbf{y})\}$$

and

$$z_i = \sum_{x_i \in \{-1,1\}} \exp\{A(x_i, \mathbf{y}) + \sum_{j \in N_i} I(x_i, x_j, \mathbf{y})\}$$

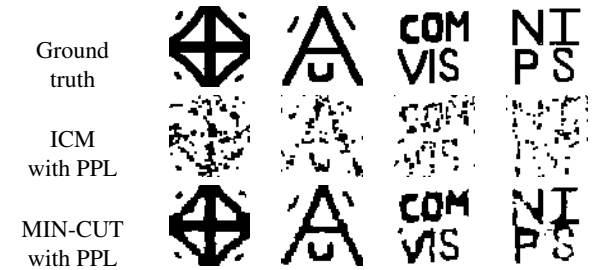


Figure 3: Image restoration with DRF. Different inference methods (ICM: Iterated Conditional Modes, MIN-CUT: min-cut/max-flow algorithm) in combination with single parameter learning method (PPL: Penalized Pseudo-likelihood).

As stated in (Kumar and Hebert, 2004a), if τ is given, the problem in Eq. (4) is convex with respect to the model parameters and can be maximized using gradient ascent. We implement a gradient ascent method variation with exact line search and maximize for different values of τ.

In our experiments, we adopt two methods of parameter learning. In the first set of experiments, we learn the parameters of the DRF using a uniform prior over the parameters in Eq. (4), i.e., $\tau = \infty$. This approach is referred to as the pseudo-likelihood (PL) learning method. Learning technique in the second set of experiments, where prior over the DRF model parameters is used, is denoted as the penalized pseudo-likelihood (PPL) learning method.

Discrete approximations of the partition function based on the Saddle Point Approximation (SPA) (Geiger and Girosi, 1991), Pseudo-Marginal Approximation (PMA) (McCallum et al., 2003) and the Maximum Marginal Approximation (MMA) are described in (Kumar et al., 2005). A Markov Chain Monte Carlo (MCMC) sampling inspired method proposed by Hinton (Hinton, 2002) and called the Contrastive Divergence (CD), is another way to deal with the combinatorial size of the label space. We leave that for future exploration.

3.2 Inference

When solving the inference problem we aim at finding the optimal label configuration $\mathbf{x}$ given an image $\mathbf{y}$. Currently, we use the MAP estimate as the criterion of optimality.

In case the probability distribution meets certain conditions (Kolmogorov and Zabih, 2004), MAP estimate can be computed exactly for undirected graphs and binary classification problems employing the max-flow/min-cut type of algorithms (Greig et al., 1989). We compute approximate MAP solution,

$$\mathbf{x} \leftarrow \underset{\mathbf{x}}{\operatorname{argmax}}\, P(\mathbf{x}|\mathbf{y})$$

as explained in (Kumar and Hebert, 2004a), using min-cut/max-flow algorithm and employ the algorithm described in (Boykov et al., 2001). The method is referred to as MIN-CUT.

For comparison, we use two other algorithms for inference. The first algorithm, referred to as ICM, is the Iterated Conditional Modes algorithm, (Besag, 1986), that given an initial labeling iteratively maximizes local conditional probabilities, cf. (4)

$$x_i \leftarrow \underset{x_i}{\operatorname{argmax}}\, P(x_i|\mathbf{x}_{N_i}, \mathbf{y})$$

ICM yields local estimate of the posterior. We implement a partially synchronous update scheme. In this scenario, we first divide image sites into coding sets, i.e., sets of non-neighboring pixels.

In each iteration, we compute a synchronous update of the image sites in a single coding set and then use the result to update the sites in the other coding set. This speeds up the usual sequential updating process.

For further illustration, we set the model interaction parameters to zero. This reduces MRF model to a maximum likelihood (ML) classifier. In the DRF case, the model is reduced to a logistic classifier in Eq. (3) yielding the second algorithm, referred to as LOGISTIC, that we use for comparison. In this case, given observation $\mathbf{y}$, optimal labeling $\mathbf{x}$ is found by maximizing the class posterior, i.e.,

$$x_i \leftarrow \underset{x_i}{\operatorname{argmax}}\, P(x_i|\mathbf{y})$$

Logistic inference is an example of a MAP solution where no label interaction is used.

In the future, we intend to explore an alternative to the MAP inference based on the Maximum Posterior Marginal (Marroquin, 1985). Such solution can be obtained using loopy Belief Propagation (Frey and MacKay, 1998).

4 EXPERIMENTS

To analyze the learning and inference techniques described in the previous section, we applied the DRF model to a binary image denoising task. The aim of these experiments is to recover correct labeling from corrupted binary images. We use the data that has been used in learning and inference experiments in (Kumar and Hebert, 2004a),(Kumar et al., 2005),(Kumar and Hebert, 2006) and compare our results with those published in the above mentioned works.

Four base images, see the top row in Fig. 2, 64×64 pixels each are used in the experiments. Two different noise models are employed: Gaussian noise and Bimodal (mixture of two Gaussians) noise. Details of the noise model parameters are given in (Kumar and Hebert, 2004a). For each noise model, 10 noisy images from the left most base image in Fig. 2 are used as the training set for parameter learning. Remaining 190 noisy images are used for testing.

The unary and pairwise features are defined as: $\mathbf{h}_i(\mathbf{y}) = [1, I_i]^T$ and $\boldsymbol{\mu}_{ij}(\mathbf{y}) = [1, |I_i - I_j|]^T$ respectively, where I_i and I_j are the pixel intensities at the site i and the site j. Hence, the parameter $\mathbf{w}$ and $\mathbf{v}$ are both two-element vectors, i.e., $\mathbf{w} = [w_0, w_1]^T$, and $\mathbf{v} = [v_0, v_1]^T$.

4.1 Logistic Classification, MRF and DRF

Let us first compare the DRF model with the conventional MRF model, described in Sec. 1, and the logistic classifier.

Each class generative model $P(\mathbf{y}_i|x_i)$ of the MRF in Eq. (1) is modeled as a Gaussian $P(I_i|x_i) = \mathcal{N}(I_i|\mu_1, \sigma^2)$. I_i is the pixel intensity at the site i. Standard deviation σ for both class generative models is fixed to the value of the standard deviation of the Gaussian noise model in use. We learn the MRF model parameters, i.e., Gaussian means μ_1 and μ_2 together with the interaction parameter β using pseudo-likelihood and gradient ascent.

We illustrate this comparison in Fig. 2. The original images or, in other words, the ground truth labeling, used in the experiments is shown in the first row. The second row depicts the input images, i.e., the ground truth images corrupted heavily by a bimodal noise. The third row illustrates classification result that we obtain by estimating a pixel label independent of its neighborhood.

	Inference Method	Learning Method PL	Learning Method PPL	Inference Time (sec)
Gaussian Noise	LOGISTIC	15.28	15.28	0.003
	ICM	4.33	4.33	0.081
	MIN-CUT	**2.54**	**2.55**	0.012
Learning Time (sec)		32	37	
Bimodal Noise	LOGISTIC	30.53	30.24	0.003
	ICM	22.51	22.43	0.103
	MIN-CUT	**5.69**	**5.65**	0.015
Learning Time (sec)		59	53	

Table 1: Image restoration with DRF. Pixelwise classification errors (%) on 190 test images. Rows show inference techniques and columns show parameter learning methods used for two different noise models. Means over 10 experiments are given.

The third and the fourth row in Fig. 2 illustrate two ways of incorporating label neighborhood dependencies in the classification process. The fourth row shows a typical classification result in case the contextual information is modeled using the conventional MRF model. Last row presents a typical result for classification with spatial dependencies captured by the DRF. We observe that the MRF and the DRF models yield comparable results in this case.

4.2 Parameter Learning

Finding optimal parameters of the DRF model means solving convex optimization problem in Eq. (4). For this purpose, we implement a variation of the gradient ascent algorithm that we describe in the following.

In all our experiments, we initialize the ascent method with the DRF model parameters $\theta^{(0)} = [\mathbf{w}^{(0)}; \mathbf{v}^{(0)}] = [0, 0, 0, 0]^T$. Then we repeat the gradient computation, line search and the update computation until a stopping criterion is satisfied. For the computation of the numerical gradient we use a certain small value of the spacing η' between points in each direction. This value is fixed during the whole computation. We specify the value in the following.

During line search computation, we make use of a variable spacing η. We start with some initial value of η and then anneal it according to a decrementing schedule. In our experiments, we start by choosing an initial value of $\eta = 0.2$ and anneal η by multiplying it with 0.4 until the value of η is smaller than 0.001. We choose the update step size via exact line search as being a multiple of the current spacing.

We iterate until a stopping criterion is satisfied. In our experiments, we run our optimization until the vector norm of the difference of the last two parameter vectors in the minimizing sequence is smaller than current spacing. Every time a convergence is reached current spacing is annealed in the way described above.

We note that it is the smallest value of the variable spacing η that is used in every iteration as the fixed spacing η' for the numerical gradient computation.

Our experimental observations motivate the use of exact optimization. An inexact approach, commonly used in practice, tends to stop the computation far from optimum.

4.3 Inference

We compare results of LOGISTIC, ICM and MIN-CUT inference for the case of parameters learned through PL, PPL and for both

	Gaussian noise MIN-CUT	Bimodal noise MIN-CUT	Learning time (Sec)
MMA, KH'05	34.34	26.53	636
PL, KH'05	3.82	17.69	300
CD, KH'05	3.78	8.88	207
PMA, KH'05	2.73	6.45	1183
SPA, KH'05	2.49	5.82	82
PL, ours	2.54 ± 0.04	5.69 ± 0.06	46 ± 8
PPL, ours	2.55 ± 0.04	$\mathbf{5.65} \pm 0.06$	$\mathbf{45} \pm 5$

Table 2: Image restoration with DRF. Pixelwise classification errors (%) on 190 test images. Rows show parameter learning methods and columns show inference technique used for two different noise models. KH'05 stands for the results published in (Kumar and Hebert, 2005). Mean $\pm$ standard deviation over 10 experiments is given for our results.

noise models. For PPL learning we used uniform prior over the association parameters, Gaussian prior over the interaction parameters and we optimize for different values of the parameter τ. Several typical results on synthetic and natural images can be respectively found in Fig. 3 and Fig. 4. Our experiments are further summarized in Tab. 1.

Gaussian prior over the interaction parameters $\{\mathbf{v}\}$ is used in our experiments, where uniform prior is used for the rest of the parameters. As in (Kumar and Hebert, 2004a), PPL learning with Gaussian prior over the interaction parameters $\mathbf{v}$ together with the MIN-CUT inference yields the lowest classification error for both noise models.

We run the PPL parameter learning for the following values of the prior parameter $\tau = \{1, 0.1, 0.01, 0.001\}$ and choose the parameter value based on the lowest resulting classification error.

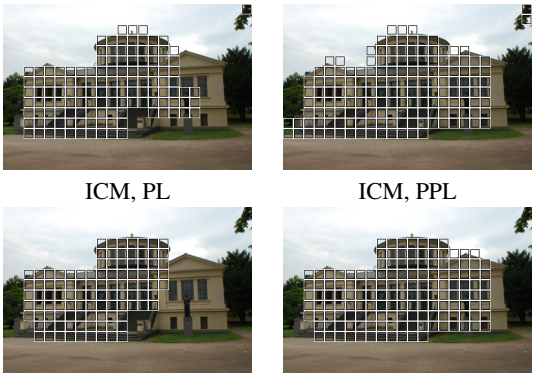
Further, we compare our results of MAP MIN-CUT inference used with PPL learning with the results of MAP MIN-CUT inference used in combination with other state of the art learning methods proposed in (Kumar et al., 2005) and mentioned in Sec. 3.1. We summarize the comparison in Tab. 2. For the bimodal noise model, our PPL learning with MAP MIN-CUT inference outperforms other learning methods in Tab. 2. Further, we note that enhanced performance is achieved while less time is needed for learning.

In (Kumar et al., 2005), the MAP inference with SPA learning is found to yield the lowest classification error for the Gaussian noise model. Both PL and PPL learning with the MIN-CUT inference yield comparable classification errors in this case. For the Bimodal noise model, the lowest error is obtained using the MPM inference with PMA learning. This combination is not mentioned in Tab. 2.

To conclude, we improve the PL and the PPL parameter learning methods and outperform other methods used in combination with MAP MIN-CUT inference proposed in (Kumar et al., 2005).

4.4 Natural Images

We demonstrate the application feasibility on natural images in the following. Our intention in this experiment is to label each site of a test image as *structured* or *non-structured*. We divide our test images, each of size 384×256 pixels, into non-overlapping blocks, each of size 16×16 pixels, that we call image sites. For each image site i, a 2-dim single-site gradient magnitude and orientation based feature vector is computed. We use linear discriminant and quadratic feature mapping to design the potential functions of the random field.



ICM, PL ICM, PPL

MIN-CUT, PL MIN-CUT, PPL

Figure 4: Image interpretation with DRF. Different inference algorithms (ICM: Iterated Conditional Modes, MIN-CUT: min-cut/max-flow algorithm) in combination with different parameter learning methods (PL: Pseudo-likelihood, PPL: Penalized Pseudo-likelihood).

	Sitewise Error	DR	FP	Learning Time (sec)
MRF, PL	29.57	61.11	11.42	10
DRF, PL	29.55	61.16	11.48	23
MRF, PPL	17.94	87.83	29.13	12
DRF, PPL	17.78	85.26	23.66	25

Table 3: Image interpretation. Rows show different models (MRF, DRF) trained using different learning methods (PL, PPL). Comparison is given in terms of sitewise error, detection rate (DR) and false positive rate (FP). 100 images were used to learn the model parameters and 47 images were used for testing. Means over 10 experiments are given.

We observe that, by imposing the smoothness label prior, the conventional MRF approach reduces the classification error of an independent sitewise classifier. The data-dependent smoothness label prior of the DRF model further reduces the false positive rate of the conventional MRF approach. In this experiment, 0.002 seconds was the average inference time. See Fig. 1 for illustration and Tab. 3 for more details.

5 CONCLUSIONS AND FUTURE WORK

We investigate discriminative random fields which provide a principled approach for combining local discriminative classifiers that allow the use of arbitrary overlapping features, with adaptive data-dependent smoothing over the label field. We show that state of the art parameter learning methods can be improved and that employing the approach for interpreting terrestrial images of urban scenes is feasible. Currently, we explore the ways of further improving the DRF model parameter learning.

ACKNOWLEDGEMENTS

The authors would like to thank V. Kolmogorov for the min-cut code and S. Kumar for the training and test data. The first author was supported by the EC Project FP6-IST-027113 eTRIMS.

REFERENCES

Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems. Journal of the Royal Statistical Society 36(2), pp. 192–236.

Besag, J., 1975. Statistical analysis of non-lattice data. The Statistician 24(3), pp. 179–195.

Besag, J., 1977. Efficiency of pseudo-likelihood estimation for simple gaussian fields. Biometrika 64, pp. 616–618.

Besag, J., 1986. On the statistical analysis of dirty pictures. Journal of the Royal Statistical Society 48(3), pp. 259–302.

Boykov, Y., Veksler, O. and Zabih, R., 2001. Fast approximate energy minimization via graph cuts. IEEE Trans. on Pattern Analysis and Machine Intelligence 23(11), pp. 1222–1239.

Frey, B. J. and MacKay, D. J. C., 1998. A revolution: belief propagation in graphs with cycles. In: Advances in neural information processing systems 10, MIT Press, pp. 479–485.

Geiger, D. and Girosi, F., 1991. Parallel and deterministic algorithms from mrfs: Surface reconstruction. IEEE Trans. on Pattern Analysis and Machine Intelligence 13(5), pp. 401–412.

Geman, S. and Geman, D., 1984. Stochastic relaxation, gibbs distribution, and the bayesian restoration of images. IEEE Trans. on Pattern Analysis and Machine Intelligence 6, pp. 721–741.

Greig, D. M., Porteous, B. T. and Seheult, A. H., 1989. Exact maximum a posteriori estimation for binary images. Journal of the Royal Statistical Society 51(2), pp. 271–279.

Hinton, G. E., 2002. Training products of experts by minimizing contrastive divergence. Neural Computation 14(8), pp. 1771–1800.

Kolmogorov, V. and Zabih, R., 2004. What energy functions can be minimized via graph cuts? IEEE Trans. Pattern Anal. Mach. Intell. 2, pp. 147–159.

Korč, F. and Förstner, W., 2008. Approximate parameter learning in Conditional Random Fields: An empirical investigation. In: Pattern Recognition, LNCS, Springer.

Kumar, S. and Hebert, M., 2003. Discriminative random fields: A discriminative framework for contextual interaction in classification. In: Proc. of the 9th IEEE International Conference on Computer Vision, Vol. 2, pp. 1150–1157.

Kumar, S. and Hebert, M., 2004a. Discriminative fields for modeling spatial dependencies in natural images. In: S. Thrun, L. K. Saul and B. Schölkopf (eds), Advances in Neural Information Processing Systems 16, MIT Press.

Kumar, S. and Hebert, M., 2004b. Multiclass discriminative fields for parts-based object detection. In: Snowbird Learning Workshop.

Kumar, S. and Hebert, M., 2005. A hierarchical field framework for unified context-based classification. In: 10th IEEE International Conference on Computer Vision, Vol. 2, IEEE Computer Society, pp. 1284–1291.

Kumar, S. and Hebert, M., 2006. Discriminative random fields. International Journal of Computer Vision 68(2), pp. 179–201.

Kumar, S., August, J. and Hebert, M., 2005. Exploiting inference for approximate parameter learning in discriminative fields: An empirical study. In: 5th International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition.

Lafferty, J., McCallum, A. and Pereira, F., 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In: Proc. 18th International Conf. on Machine Learning, pp. 282–289.

Lee, C.-H., Wang, S., Jiao, F., Schuurmans, D. and Greiner, R., 2007. Learning to model spatial dependency: Semi-supervised discriminative random fields. In: B. Schölkopf, J. Platt and T. Hoffman (eds), Advances in Neural Information Processing Systems 19, MIT Press, pp. 793–800.

Li, S. Z., 2001. Markov random field modeling in image analysis. Springer-Verlag New York, Inc.

Marroquin, J. L., 1985. Probabilistic solution of inverse problems. PhD thesis, Massachusetts Institute of Technology.

McCallum, A., Rohanimanesh, K. and Sutton, C., 2003. Dynamic conditional random fields for jointly labeling multiple sequences. In: NIPS Workshop on Syntax, Semantics, and Statistics.

Modestino, J. W. and Zhang, J., 1992. A markov random field model-based approach to image interpretation. IEEE Trans. on Pattern Analysis and Machine Intelligence 14(6), pp. 606–615.

Quattoni, A., Collins, M. and Darrell, T., 2004. Conditional random fields for object recognition. In: Advances in Neural Information Processing Systems 17, IEEE Computer Society, pp. 1521–1527.

Quattoni, A., Wang, S., Morency, L.-P., Collins, M. and Darrell, T., 2007. Hidden conditional random fields. IEEE Trans. on Pattern Analysis and Machine Intelligence 29(10), pp. 1848–1852.

Rabinovich, A., Vedaldi, A., Galleguillos, C., Wiewiora, E. and Belongie, S., 2007. Objects in context. In: Proc. of the 11th IEEE International Conference on Computer Vision.

Sminchisescu, C., Kanaujia, A., Li, Z. and Metaxas, D., 2005. Conditional random fields for contextual human motion recognition. In: Proceedings of the 10th IEEE International Conference on Computer Vision, Vol. 2, IEEE Computer Society, pp. 1808–1815.

Sutton, C. and McCallum, A., 2007. Piecewise pseudolikelihood for efficient training of conditional random fields. In: Z. Ghahramani (ed.), International Conference on Machine learning (ICML), Vol. 227, pp. 863–870.

Vishwanathan, S., Schraudolph, N. N., Schmidt, M. W. and Murphy, K., 2006. Accelerated training of conditional random fields with stochastic gradient methods. In: W. W. Cohen and A. Moore (eds), Proc. of the 24th International Conf. on Machine Learning, Vol. 148, ACM Press, pp. 969–976.

Wang, S. B., Quattoni, A., Morency, L.-P. and Demirdjian, D., 2006. Hidden conditional random fields for gesture recognition. In: Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition.