Interactive Image Segmentation Using an Adaptive GMMRF Model

A. Blake, C. Rother, M. Brown, P. Perez, and P. Torr

Microsoft Research Cambridge UK,
7 JJ Thomson Avenue, Cambridge CB3 0FB, UK.
http://www.research.microsoft.com/vision/cambridge

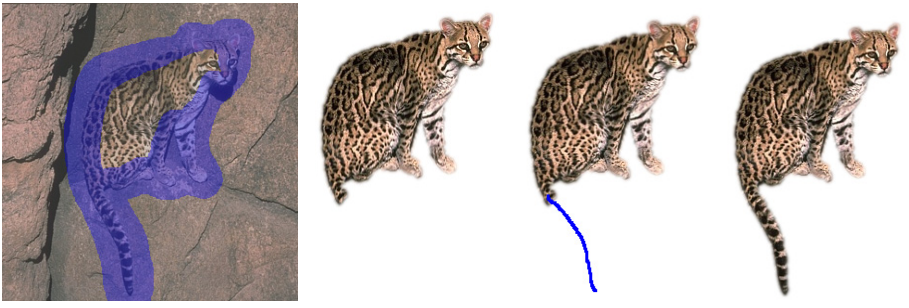
Abstract. The problem of interactive foreground/background segmentation in still images is of great practical importance in image editing. The state of the art in interactive segmentation is probably represented by the graph cut algorithm of Boykov and Jolly (ICCV 2001). Its underlying model uses both colour and contrast information, together with a strong prior for region coherence. Estimation is performed by solving a graph cut problem for which very efficient algorithms have recently been developed. However the model depends on parameters which must be set by hand and the aim of this work is for those constants to be learned from image data.

First, a generative, probabilistic formulation of the model is set out in terms of a "Gaussian Mixture Markov Random Field" (GMMRF). Secondly, a pseudolikelihood algorithm is derived which jointly learns the colour mixture and coherence parameters for foreground and background respectively. Error rates for GMMRF segmentation are calculated throughout using a new image database, available on the web, with ground truth provided by a human segmenter. The graph cut algorithm, using the learned parameters, generates good object-segmentations with little interaction. However, pseudolikelihood learning proves to be frail, which limits the complexity of usable models, and hence also the achievable error rate.

1 Introduction

The problem of interactive image segmentation is studied here in the framework used recently by others [1,2,3] in which the aim is to separate, with minimal user interaction, a foreground object from its background so that, for practical purposes, it is available for pasting into a new context. Some studies [1,2] focus on inference of transparency in order to deal with mixed pixels and transparent textures such as hair. Other studies [4, 3] concentrate on capturing the tendency for images of solid objects to be coherent, via Markov Random Field priors such as the Ising model. In this setting, the segmentation is "hard" — transparency is disallowed. Foreground estimates under such models can be obtained in a precise way by graph cut, and this can now be performed very efficiently [5]. This has application to extracting the foreground object intact, even in camouflage — when background and foreground colour distributions overlap at least in part. We have not come across studies claiming to deal with transparency and camouflage simultaneously, and in our experience this is a very difficult combination. This paper therefore addresses the problem of hard segmentation problem in camouflage, and does not deal with the transparency issue.

T. Pajdla and J. Matas (Eds.): ECCV 2004, LNCS 3021, pp. 428–441, 2004.



(a) (b) (c) (d)

Fig. 1. Illustrating the GMMRF **algorithm for interactive segmentation.** The user draws a fat pen trail enclosing the object boundary (a), marked in blue. This defines the "trimap" with foreground/background/unclassified labels. The GMMRF algorithm produces (b). Missing parts of the object can be added efficiently: the user roughly applies a foreground brush (c), marked in blue, and the GMMRF method adds the whole region (d).

The interactive segmentation problem. The operation of adaptive segmentation by our model, termed a "Gaussian Mixture Markov Random Field" (GMMRF), is illustrated by the following example. The desired object has been cleanly separated from its background with a modest amount of user interaction (figure 1). Stripping away details of user interaction, the basic problem input consists of an image and its "trimap" as in figure 1a). The trimap defines training regions for foreground and background, and the segmentation algorithm is then applied to the "unclassified" region as shown, in which all pixels must be classified foreground or background as in fig. 1b). The classification procedure needs to apply:

- jointly across those pixels;
- matching the labels of adjoining labelled pixels;
- using models for colour and texture properties of foreground and background pixels learned from the respective training regions;
- benefiting from any general notions of region coherence that can be encoded in prior distributions.

2 Generative Models for Interactive Segmentation

This section reviews generative models for two-layer (foreground/background) colour images in order to arrive at the simplest capable, probabilistic model for interactive segmentation — the *contrast-sensitive GMMRF*. A generative, probabilistic model for image pixel data can be expressed in terms of colour pixel data $\mathbf{z}$, opacity variables $\boldsymbol{\alpha}$, opacity prior and data likelihood as follows.

Image. $\mathbf{z}$ is an array of colour (RGB) indexed by the (single) index n:

$$\mathbf{z} = (z_1, \ldots, z_n, \ldots, z_N),$$

with corresponding hidden variables for transparency $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N)$, and hidden variables for mixture index $\mathbf{k} = (k_1, \ldots, k_N)$. Each pixel is considered, in general,

to have been generated as an additive combination of a proportion α_n ($0 \leq \alpha_n \leq 1$) of foreground colour with a proportion $1 - \alpha_n$ of background colour [2,1]. Here we concentrate attention on the hard segmentation problem in which $\alpha_n = 0, 1$ — binary valued.

Gibbs energy formulation. Now the posterior for $\boldsymbol{\alpha}$ is given by

$$p(\boldsymbol{\alpha} \mid \mathbf{z}) = \sum_{\mathbf{k}} p(\boldsymbol{\alpha}, \mathbf{k} \mid \mathbf{z}) \tag{1}$$

and

$$p(\boldsymbol{\alpha}, \mathbf{k} \mid \mathbf{z}) = \frac{1}{p(\mathbf{z})} p(\mathbf{z} \mid \boldsymbol{\alpha}, \mathbf{k}) p(\boldsymbol{\alpha}, \mathbf{k}). \tag{2}$$

This can be written as a Gibbs distribution, omitting the normalising constant $1/p(\mathbf{z})$ which is anyway constant with respect to $\boldsymbol{\alpha}$:

$$p(\boldsymbol{\alpha}, \mathbf{k} \mid \mathbf{z}) \propto \frac{1}{Z_L} \exp -\mathcal{E} \;\text{ with }\; \mathcal{E} = L + U + V. \tag{3}$$

The intrinsic energies U and V encode the prior distributions:

$$p(\boldsymbol{\alpha}) \propto \exp -U \;\text{ and }\; p(\mathbf{k} \mid \boldsymbol{\alpha}) \propto \exp -V \tag{4}$$

and the extrinsic energy L defines the likelihood term

$$p(\mathbf{z} \mid \boldsymbol{\alpha}, \mathbf{k}) = \prod_n p(z_n | \alpha_n, k_n) = \frac{1}{Z_L} \exp -L. \tag{5}$$

Simple colour mixture observation likelihood. A simple approach to modelling colour observations is as follows. At each pixel, foreground colour is considered to be generated randomly from one of K Gaussian mixture components with mean $\mu(\ldots)$ and covariance $P(\ldots)$ from the foreground, and likewise for the background:

$$p(z_n | \alpha_n, k_n) = \mathcal{N}\left(z; \mu(k_n, \alpha_n), P(k, \alpha_n)\right), \;\; \alpha_n = 0, 1, \; k_n = 1, \ldots, K \tag{6}$$

and the components have prior probabilities

$$p(\mathbf{k} | \boldsymbol{\alpha}) = \prod_n p(k_n | \alpha_n) \;\text{ with }\; p(k_n | \alpha_n) = \pi(k_n, \alpha_n). \tag{7}$$

The exponential coefficient for each component, referred to here as the "extrinsic energy coefficient" is denoted $S(k, \alpha) = \frac{1}{2}(P(k, \alpha))^{-1}$. The corresponding extrinsic term in the Gibbs energy is given by

$$L = \sum_n D_n \tag{8}$$

where

$$D_n = \left[z_n - \mu(k_n, \alpha_n)\right]^\top S(k_n, \alpha_n) \left[z_n - \mu(k_n, \alpha_n)\right]. \tag{9}$$

A useful special case is the *isotropic* mixture in which $S(k_n, \alpha_n) = \lambda(k_n, \alpha_n) I$.

Note that, for a pixelwise-independent likelihood model as in (6), the *partition function* Z_L for the likelihood decomposes multiplicatively across sites n. Since also the partition function Z_0 for the prior is always independent of $\boldsymbol{\alpha}$, the MAP estimate of $\boldsymbol{\alpha}$ can be obtained (3) by minimising $\mathcal{E} - \log Z_L$ with respect to $\boldsymbol{\alpha}$ and this can be achieved exactly, given that α_n is binary valued here, using a "minimum cut" algorithm [4] which has recently been developed [3] to achieve good segmentation in interactive time (around 1 second for a 500^2 image).

Simple opacity prior (No spatial interaction). The simplest choice of a joint prior $p(\boldsymbol{\alpha})$ is the spatially trivial one, decomposing into a product of marginals

$$p(\boldsymbol{\alpha}) = \prod_n p(\alpha_n)$$

with, for example, $p(\alpha_n)$ uniform over $\alpha_n \in \{0, 1\}$ (hard opacity) or $\alpha_n \in [0, 1]$ (variable transparency). This is well known [3] to give poor results and this will be quantified in section 5.

Ising Prior. In order to convey a prior on object coherence, $p(\boldsymbol{\alpha})$ can be spatially correlated, for example via a first order Gauss-Markov interaction [6,7] as, for example, in the *Ising* prior:

$$p(\boldsymbol{\alpha}) \propto \exp -U \text{ with } U = \gamma \sum_{m,n \in \mathcal{C}} [\alpha_n \neq \alpha_m], \tag{10}$$

where $[\phi]$ denotes the indicator function taking values $0, 1$ for a predicate ϕ, and where $\mathcal{C}$ is the set of all cliques in the Markov network, assumed to be two-pixel cliques here. The constant γ is the *Ising parameter*, determining the strength of spatial interaction. The MAP estimate of $\boldsymbol{\alpha}$ can be obtained by minimising with respect to $\boldsymbol{\alpha}$ the Gibbs energy (3) with L as before, and the Ising U (10). This can be achieved exactly, given that α_n is binary valued here, using a "minimum cut" algorithm [4]. In practice [3] the homogeneous MRF succeeds in enforcing some coherence, as intended, but introduces "Manhattan" artefacts that point to the need for a more subtle form of prior and/or data likelihood, and again this is quantified in section 5.

3 Incorporating Contrast Dependence

The Ising prior on opacity, being homogeneous, is a rather blunt instrument, and it is convincingly argued [3] that a "prior" that encourages object coherence *only* where contrast is low, is far more effective. However, a "prior" that is dependent on data in this way (dependent on data in that image gradients are computed from intensities $\mathbf{z}$) is not strictly a prior. Here we encapsulate the spirit of a contrast-sensitive "prior" more precisely as a gradient dependent likelihood term of the form

$$f(\nabla \mathbf{z} | \boldsymbol{\alpha}, \beta, \gamma) \tag{11}$$

where β is a further coherence parameter in a Gauss-Markov process over z.

It turns out that the contrast term introduces a new technical issue in the data-likelihood model: long-range interactions are induced that fall outside the Markov framework, and therefore strictly fall outside the scope of graph cut optimisation. The long-range interaction is manifested in the partition function Z_L for the data likelihood. This is an inconvenient but inescapable consequence of the probabilistic view of the contrast-sensitive model. While it imperils the graph cut computation of the MAP estimate, and this will be addressed in due course, correct treatment of the partition function turns out to be critical for *adaptivity*. This is because of the well-known role of partition functions [7] in *parameter learning* for Gibbs distributions. Note also that recent advances in *Discriminative Random Fields* [8] which can often circumvent such issues, turn out not to be applicable to using the GMMRF model with trimap labelling.

3.1 Gibbs Energy for Contrast-Sensitive GMMRF

A modified Gibbs energy that takes contrast into account is obtained by replacing the term U in (10) by

$$U^+ = \gamma \sum_{m,n\in\mathcal{C}} [\alpha_n \neq \alpha_m] \exp -\frac{\beta}{\gamma}\|z_m - z_n\|^2, \tag{12}$$

which relaxes the tendency to coherence where image contrast is strong. The constant β is supposed to relate to γ via observation noise and we set

$$\frac{\gamma}{\beta} = C \left\langle \|z_m - z_n\|^2 \right\rangle, \text{ with } C = 4 \tag{13}$$

where $\langle \ldots \rangle$ denotes expectation over the test image sample, and taking the constant $C = 4$ is justified later.

The results of minimising $\mathcal{E} = L + U^+$ (neglecting $\log Z_L$ — see later) gives considerably improved segmentation performance [3]. In our experience "Manhattan" artefacts are suppressed by the reduced tendency of segmentation boundaries to follow Manhattan geodesics, in favour of following lines of high contrast. Results in section 5 will confirm quantitatively the performance gain.

3.2 Probabilistic Model

In the contrast-sensitive version of the Gibbs energy, the term U^+ no longer corresponds to a prior for $\boldsymbol{\alpha}$, as it did in the homogeneous case (10). The entire Gibbs energy is now

$$\mathcal{E} = \sum_n D_n + \gamma \sum_{m,n\in\mathcal{C}} [\alpha_n \neq \alpha_m] \exp -\frac{\beta}{\gamma}\|z_m - z_n\|^2. \tag{14}$$

Adding a "constant" term

$$\gamma \sum_{m,n\in\mathcal{C}} (1 - \exp -\frac{\beta}{\gamma}\|z_m - z_n\|^2) \tag{15}$$

to $\mathcal{E}$ has no effect on the minimum of $\mathcal{E}(\alpha)$ w.r.t. $\boldsymbol{\alpha}$, and transforms the problem in a helpful way as we will see. The addition of (15) gives $\mathcal{E} = U + L$ where U is the earlier Ising prior (10) and now L is given by

$$L = \sum_n D_n + \gamma \sum_{m,n \in \mathcal{C}} [\alpha_n = \alpha_m](1 - \exp -\frac{\beta_{\alpha_n}}{\gamma}\|z_m - z_n\|^2), \tag{16}$$

with separate texture constants β_0, β_1 for foreground and background respectively. This is a fully generative, probabilistic account of the contrast-sensitive model, cleanly separating prior and likelihood terms. Transforming $\mathcal{E}$ by the addition of the constant term (15) has had several beneficial effects. First it separates the energy into a component U which is active only at foreground/background region boundaries, and a component L whose contrast term acts only within region. It is thus clear that the parameter β is a textural parameter — and that is why it makes sense to learn separate parameters $\beta_0,\ \beta_1$. Secondly when, for tractability in learning, the Gibbs energy is approximated by a quadratic energy in the next section 4, the transformation is in fact essential for the resulting data-likelihood MRF to be *proper* (ie capable of normalisation).

Inference of foreground and background labels from posterior. Given that L is now dependent on $\boldsymbol{\alpha}$ and $\mathbf{z}$ simultaneously, the partition function Z_L for the likelihood, which was a locally factorised function for the simple likelihood model of section 2, now contains non-local interactions over $\boldsymbol{\alpha}$. It is no longer strictly correct that the posterior can be maximised simply by minimising $\mathcal{E} = L + U$. Neglecting the partition function Z_L in this way can be justified, it turns out, by a combination of experiment and theory — see section 6.

MAP inference of $\boldsymbol{\alpha}$ is therefore done by applying min cut as in [3] to the Gibbs energy $\mathcal{E}$. For this step we can either marginalise over $\mathbf{k}$, or maximize with respect to $\mathbf{k}$, the latter being computationally cheaper and tending to produce very similar results in practice.

4 Learning Parameters

This section addresses the critical issue of how mixture parameters $\mu(k, \alpha)$, $P(k, \alpha)$ and $\pi(k, \alpha)$ can be learned from data simultaneously with coherence parameters $\{\beta_\alpha\}$. In the simple uncoupled model with $\beta_\alpha = 0$ for $\alpha = 0, 1$ mixture parameters can be learned by conventional methods, but when coherence parameters are switched on, learning of all parameters is coupled non-trivially.

4.1 Quadratic Approximation

It would appear that the exponential form of L in (16) is an obstacle to tractability of parameter learning, and so the question arises whether it is an essential component of the model. Intuitively it seems well chosen because of the switching behaviour built into the exponential, that switches the model in and out of its "coherent" mode. Nonetheless,

in the interests of tractability we use a quadratic approximation, solely for the parameter learning procedure. The approximated form of the extrinsic energy (16) becomes:

$$L^* = \sum_n D_n + \sum_{m,n \in \mathcal{C}} \beta_{\alpha_n} [\alpha_n = \alpha_m] \|z_m - z_n\|^2. \tag{17}$$

and the approximation is good provided $\beta_{\alpha_n} \|z_m - z_n\|^2 < \gamma$.

Learning $\boldsymbol{\gamma}$. Note that since the labelled data consists typically of separate sets of foreground and background pixels respectively (fig 1a,b) there is no training data containing the boundary between foreground and background. There is therefore no data over which the Ising term (10) can be observed, since γ no longer appears in the approximated L^*. Therefore γ cannot simply be learned from training data in this version of the interactive segmentation problem. However for the switching of the exponential term in (16) to act correctly it is clear that we must have

$$\exp -\frac{\beta_{\alpha_n}}{\gamma} \|z_m - z_n\|^2 \approx 1 \tag{18}$$

throughout regions of homogeneous texture so, over background for instance, we must have $\gamma \geq \beta_{\alpha_n} \|z_m - z_n\|^2$, which is also the condition for good quadratic approximation above. Given that the statistics of $z_m - z_n$ are found to be consistently Gaussian in practice, this is secured by (13).

4.2 Pseudolikelihood

A well established technique for parameter estimation in formally intractable MRFs like this one is to introduce a "pseudolikelihood function" [6] and maximise it with respect to its parameters. The pseudolikelihood function has the form

$$\mathcal{P} = \exp -\mathcal{E}^* \tag{19}$$

where $\mathcal{E}^*$ will be called the "pseudo-energy", and note that $\mathcal{P}$ is free of any partition function. There is no claim that $\mathcal{P}$ itself approximates the true likelihood, but that, under certain circumstances, its maximum coincides asymptotically with that of the likelihood [7] — asymptotically, that is, as the size of the data $\mathbf{z}$ tends to infinity. Strictly the results apply for integer-valued MRFs, so formally we should consider $\mathbf{z}$ to be integer-valued, and after all it does represent a set of quantised colour values.

Following [7], the pseudo-energy is defined to be

$$\mathcal{E}^* = \sum_n (-\log p(z_n \mid \mathbf{z}_n,\ \boldsymbol{\alpha}, \mathbf{k})) \tag{20}$$

where $\mathbf{z}_n = \mathbf{z} \backslash \{z_n\}$ — the entire data array omitting z_n. Now

$$p(z_n \mid \mathbf{z}_n\ \boldsymbol{\alpha}, \mathbf{k}) = p(z_n \mid \{z_m,\ m \in \mathcal{B}_n\},\ \boldsymbol{\alpha}, \mathbf{k}) \tag{21}$$

by the Markov property, where $\mathcal{B}_n$ is the "Markov blanket" at grid point n — the set of its neighbours in the Markov model. For the earlier example of a first-order MRF, $\mathcal{B}_n$

simply contains the N, S, E, W neighbours of n. Taking into account the earlier details of the probability distribution over the Markov structure, it is straightforward to be show that, over the training set

$$p(z_n \mid \{z_m,\ m \in \mathcal{B}_n\},\ \boldsymbol{\alpha}, \mathbf{k}) \propto \exp - \left[D_n + \sum_{m \in \mathcal{B}_n} V_{nm} \right] \tag{22}$$

where

$$V_{nm} = \beta_{\alpha_n}[\alpha_m = \alpha_n] \|z_n - z_m\|^2. \tag{23}$$

Terms $\gamma[\alpha_m \neq \alpha_n]$ from U have been omitted since they do not occur in the training sets of the type used here (fig 1), as mentioned before. Finally, the pseudo-likelihood energy function

$$\mathcal{E}^* = \sum_n \mathcal{E}_n^* \ \text{ with } \ \mathcal{E}_n^* = Z_n^* + D_n + \sum_{m \in \mathcal{B}_n} V_{nm} \tag{24}$$

and $\exp - Z_n^*$ is the (local) partition function.

4.3 Parameter Estimation by Autoregression over the Pseudolikelihood

It is well known that the parameters of a Gaussian MRF can be obtained from pseudolikelihood as an auto-regression estimate [9]. For tractability, we split the estimation problem up into $2K$ problems, one for each foreground and background mixture component, treated independently except for sharing common constants β_0 and β_1. For this purpose, the mixture index k_n at each pixel is determined simply by maximisation of the local likelihood:

$$k_n = \arg\max_k p(z_n|\alpha_n, k)\pi_k^{\alpha_n}. \tag{25}$$

Further, for tractability, we restrict the data $\{z_n\}$ to those pixels (much the majority in practice) whose foreground/background label α agrees with all its neighbours — ie the n for which

$$\alpha_m = \alpha_n \text{ for all } m \in \mathcal{B}_n.$$

Observables z_n with a given class label α_n and component index k_n are then dealt with together, in accordance with the pseudolikelihood model above, as variables from the regression

$$z_n - \overline{z} \mid \mathbf{z}_n \sim \mathcal{N}(A(x_n - \overline{z}), P) \tag{26}$$

where

$$x_n = \frac{1}{M} \sum_{m \in \mathcal{B}_n} z_m, \tag{27}$$

and $M = |\mathcal{B}_n|$. The mean colour is estimated simply as

$$\overline{z} = \langle z \rangle \tag{28}$$

where $\langle \ldots \rangle$ denotes the sample mean over pixels from class α and with component index k. We can solve for A and P using standard linear regression as follows:

$$A = \left\langle (z - \overline{z})(x - \overline{z})^\top \right\rangle \left(\left\langle (x - \overline{z})(x - \overline{z})^\top \right\rangle \right)^{-1} \tag{29}$$

and

$$P = \left\langle \epsilon \epsilon^\top \right\rangle \text{ where } \epsilon = z - \overline{z} - A(x - \overline{z}). \tag{30}$$

Lastly, model parameters for each colour component, for instance of the background, should be obtained to satisfy

$$M\beta_0 I = P^{-1} A \tag{31}$$

$$S(k_n, \alpha_n) = P^{-1} - M\beta_0 I, \tag{32}$$

$$\mu(k_n, \alpha_n) = \overline{z}, \tag{33}$$

and similarly for the foreground.

Note there are some technical problems here. First is that (31) represents a constraint that is not necessarily satisfied by P and A, and so the regression needs to be solved under this constraint. Second is that in (32) $S(k_n, \alpha_n)$ should be positive definite but this constraint will not automatically be obeyed by the solution of the autoregression. Thirdly that one value of β_0 needs to satisfy the set of equations above for all background components. The first problem is dealt with by restricting S to be isotropic so that P and A must also be isotropic and the entire set of equations are solved straightforwardly under isotropy constraints. (In other words, each colour component is regressed independently of the others.) It turns out that this also solves the second problem. An unconstrained autoregression on typical natural image data, with general symmetric matrices for the $S(k_n, \alpha_n)$, will, in our experience, often lead to a non-positive definite solution for $S(k_n, \alpha_n)$ and this is unusable in the model. Curiously this problem with pseudolikelihood and autoregeression seems not to be generally acknowledged in standard texts [7, 9]. Empirically however we have found that the problem ceases with isotropic $S(k_n, \alpha_n)$, and so we have used this throughout our experiments. The use of isotropic components seems not to have much effect on quality provided that, of course, a larger number of mixture components must be used than for equivalent performance with general symmetric component-matrices. Lastly, the tying of β_0 across components can be achieved simply by averaging *post-hoc*, or by applying the tying constraint explicitly as part of the regression which is, it turns out, quite tractable (details omitted).

5 Results

Testing of the GMMRF segmentation algorithms uses a database of 50 images. We compare the performance of: i) Gaussian colour models, with no spatial interaction; ii) the simple Ising model; iii) the full contrast-sensitive GMMRF model with fixed interaction parameter γ; iv) the full GMMRF with learned parameters. Results are obtained using 4-connectivity, and isotropic Gaussian mixtures with $K = 30$ components as the data potentials D_n.

Test Database. The database contains 15 training and 35 test images, available online[1]. The database contrasts with the form of ground truth supplied with the Berkeley

[1] http://www.research.microsoft.com/vision/cambridge/segmentation/

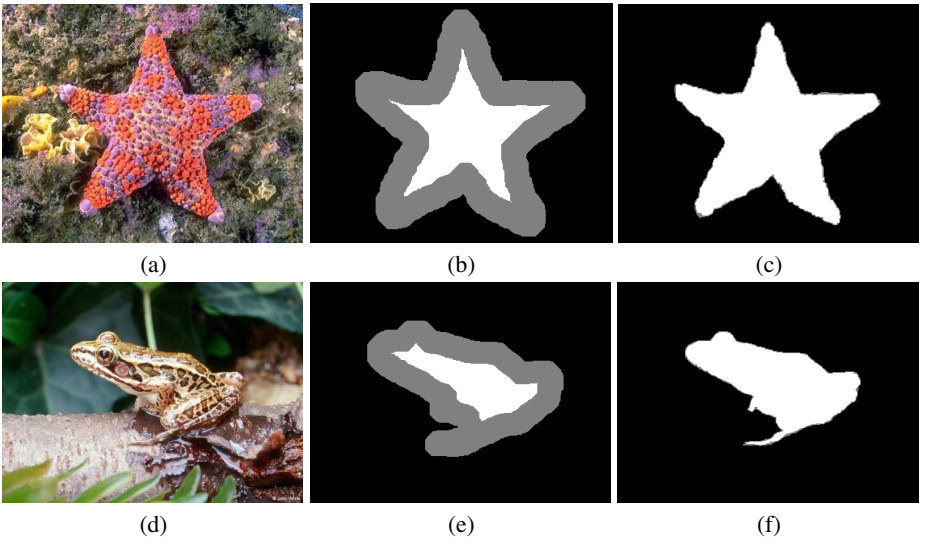


Fig. 2. (a,d) Two images from the test database. (b,e) User defined trimap with foreground (white), background (black) and unclassified (grey). (c,f) Expert trimap which classifies pixels into foreground (white), background (black) and unknown (grey); unknown here refers to pixels too close to the object boundary for the expert to classify, including mixed pixels.

database[2] which is designed to test exhaustive, for bottom up segmentation [10]. Each image in our database contains a foreground object in a natural background environment (see fig. 2). Since the purpose of the dataset is to evaluate various algorithms for *hard* image segmentation, objects with no or little transparency are used. Consequently, partly transparent objects like trees, hair or glass are not included. Two kinds of labeled trimaps are assigned to each image. The first is the user trimap as in fig. 2(b,e). The second is an "expert trimap" obtained from painstaking tracing of object outlines with a fine pen (fig. 2(c,f)). The fine pen-trail covers possibly mixed pixels on the object boundary. These pixels are excluded from the error rate measures reported below, since there is no definitive ground truth as to whether they are foreground or background.

Evaluation. Segmentation error rate is defined as

$$\epsilon = \frac{\text{no. misclassified pixels}}{\text{no. pixels in unclassified region}} , \tag{34}$$

where "misclassified pixels" excludes those from the unclassified region of the expert trimap. This simple measurement is sufficient for a basic evaluation of segemntation performance. It might be desirable at some later date to devise a second measure that quantifies the degree of user effort that would be required to correct errors, for example by penalising scattered error pixels.

[2] http://www.cs.berkeley.edu/projects/vision/grouping/segbench/

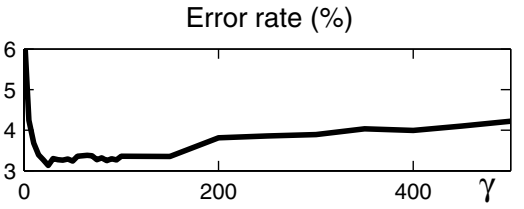


Fig. 3. The error rate on the training set of 15 images, using different values for γ. The minimum error is achieved for $\gamma = 25$. The GMMRF model uses isotropic Gaussians and 4-connectivity.

Segmentation model	Error rate
GMMRF; discriminatively learned $\gamma = 20$ ($K = 10$ full Gaussian)	7.9%
Learned GMMRF parameters ($K = 30$ isotropic Gauss.)	8.6%
GMMRF; discriminatively learned $\gamma = 25$ ($K = 30$ isotropic Gaussian)	9.4%
Strong interaction model ($\gamma = 1000$; $K = 30$ isotropic Gaussian)	11.0%
Ising model ($\gamma = 25$; $K = 30$ isotropic Gaussian)	11.2%
Simple mixture model – no interaction ($K = 30$ isotropic Gaussian)	16.3%

Fig. 4. Error rates on the test data set for different models and parameter determination regimes. For isotropic Gaussians, the full GMMRF model with learned parameters outperforms both the full model with discriminatively learned parameters, and simpler alternative models. However, exploiting a full Gaussian mixture model improves results further.

Test database scores. In order to compare the GMMRF method with alternative methods, a fixed value of the Ising parameter γ is learned discriminatively by optimising performance over the training set (fig. 3), giving a value of $\gamma = 25$. Then the accompanying β constant is fixed as in (13). For full Gaussians and 8-connectivity the learned value was $\gamma = 20$. The performance on the test set is summarized in figure 4 for the various different models and learning procedures. Results for the two images of figure 2 from the test database, are shown in figure 5. As might be expected, models with very strong spatial interaction, simple Ising interaction, or without any spatial interaction at all, all perform poorly. The model with no spatial interaction has a tendency to generate many isolated, segmented regions (see fig. 5). A strong interaction model ($\gamma = 1000$) has the effect of shrinking the the object with respect to the true segmentation. The Ising model, with $\gamma = 25$ set by hand, gives slightly better results, but introduces "Manhattan" artefacts — the border of the segmentation often fails to correspond to image edges. The inferiority of the Ising model and the "no interaction model" has been demonstrated previously [3] but here is quantified for the first time.

In contrast, the GMMRF model with learned γ is clearly superior. For isotropic Gaussians and 4-connectivity the GMMRF model with parameters *learned* by the new pseudolikelihood algorithm leads to slightly better results than using the discriminatively learned γ.

The lowest error rate, however, was achieved using full covariance Gaussians in the GMMRF with discriminatively learned γ. We were unable to compare with pseudolikelihood learning; the potential instability of pseudolikelihood learning (section 4) turns out to be an overwhelming obstacle when using full covariance Gaussians.

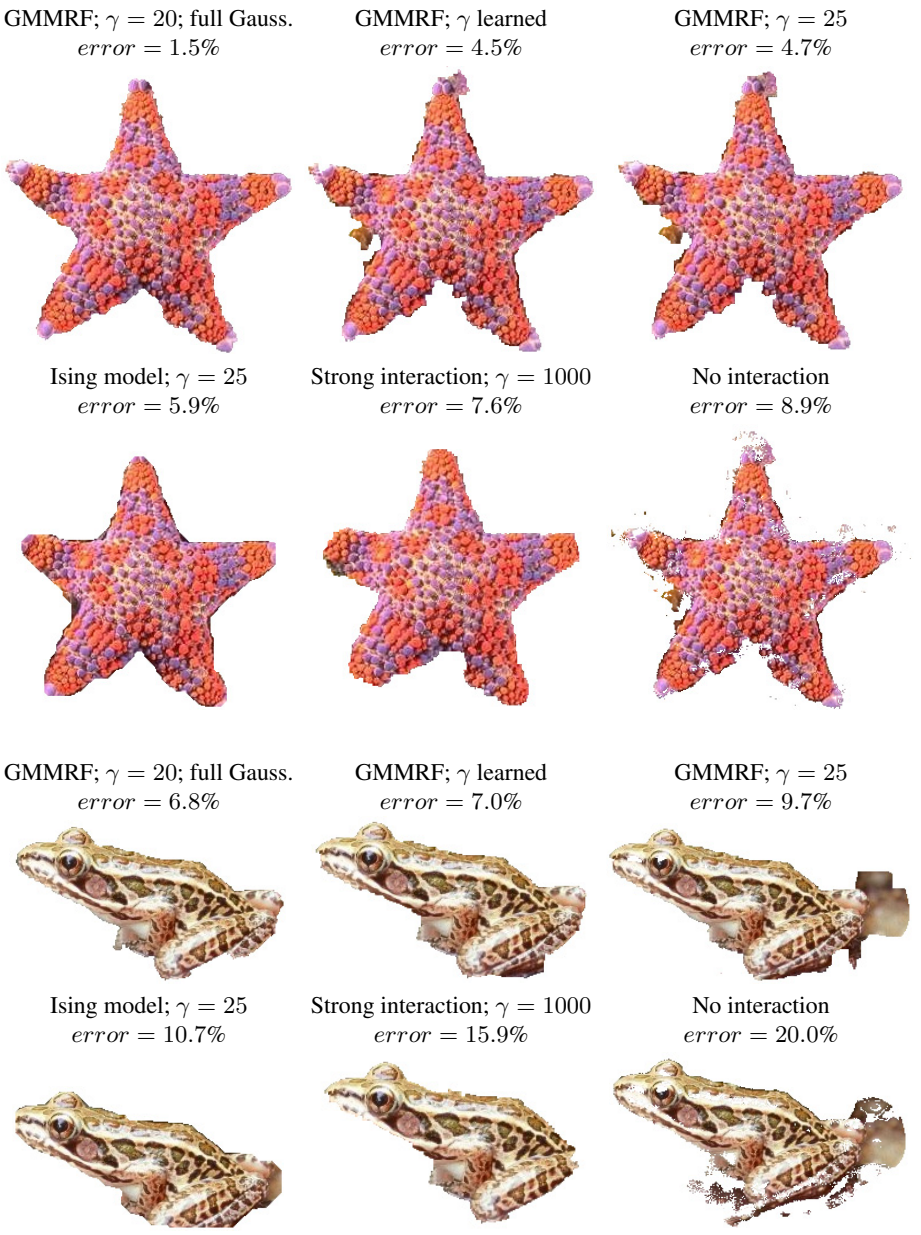


Fig. 5. Results for various segmentation algorithms (see fig. 4) for the two images shown in fig. 2. For both examples, the error rate increases from top left to bottom right. GMMRF with pseudolikelihood learning outperforms the GMMRF with discriminatively learned γ parameters, and the various simpler alternative models. The best result is achieved however using full covariance Gaussians and discriminatively learned γ.

6 Discussion

We have formalised the energy minimization model of Boykov and Jolly [3] for foreground/background segmentation as a probabilistic GMMRF and developed a pseudolikelihood algorithm for parameter learning. A labelled database has been constructed for this task and evaluations have corroborated and quantified the value of spatial interaction models — the Ising prior and the contrast-sensitive GMMRF. Further, evaluation has shown that parameter learning for the GMMRF by pseudolikelihood is effective. Indeed, it is a little more effective than simple discriminative learning for a comparable model (isotropic GMMRF); but the frailty of pseudolikelihood learning limits the complexity of model that can be used (eg full covariance GMMRF is impractical) and that in turn limits achievable performance. A number of issues remain for discussion, as follows.

DRF. The Discriminative Random Field model [8] has recently been shown to be very effective for image classification tasks. It has the great virtue of banishing the issues concerning the likelihood partition function Z_L that affect the GMMRF. However it can be shown that the DRF formulation cannot be used with the form of GMMRF developed here, and trimap labelled data, because the parameter learning algorithm breaks down (details omitted for lack of space).

Line process. The contrast-sensitive GMMRF has some similarity to the well known line process model [11]. In fact it has an important additional feature, that the observation model is a non-trivial MRF with spatial interaction (the contrast term), and this is a crucial ingredient in the success of contrast-sensitive GMMRF segmentation.

Likelihood partition function. As mentioned in section 3.2, the partition function Z_L depends on $\boldsymbol{\alpha}$ and this dependency should be taken into account when searching the MAP estimate of $\boldsymbol{\alpha}$. However, for the quadratically approximated extrinsic energy (17), this partition function is proportional to the determinant of a sparse precision matrix, which can be numerically computed for given parameters and $\boldsymbol{\alpha}$. Within the range of values used in practice for the different parameters, we found experimentally that

$$\log Z_L(\boldsymbol{\alpha}) = \text{const} + \kappa \sum_{m,n \in \mathcal{C}} [\alpha_n \neq \alpha_m], \tag{35}$$

with κ varying within a range $(0, 0.5)$. Hence, by ignoring $\log Z_L$ in the global energy to be minimised, we assume implicitly that the prior is effectively Ising with slightly weaker interaction parameter $\gamma - \kappa$. Since we have seen that graph cut is relatively insensitive (fig. 3) to perturbations in γ, this justifies neglect of the extrinsic partition fn Z_L and the application of graph cut to the Gibbs energy alone.

Adding parameters to the MRF. It might seem that segmentation performance could be improved further by allowing more general MRF models. They would have greater numbers of parameters, more than could reasonably be set by hand, and this ought

to press home the advantage of the new parameter learning capability. We have run preliminary experiments with i) spatially anisotropic clique potentials, ii) larger neighbourhoods (8-connected) and iii) independent, unpooled foreground and background texture parameters β_0 and β_1 (using min cut over a directed graph). However, in all cases error rates were substantially worsened. A detailed analysis of these issues is part of future research.

Acknowledgements. We gratefully acknowledge discussions with and assistance from P. Anandan, C.M. Bishop, B. Frey, T. Werner, A.L. Yuille and A. Zisserman.

References

1. Chuang, Y.Y., Curless, B., Salesin, D., Szeliski, R.: A Bayesian approach to digital matting. In: Proc. Conf. Computer Vision and Pattern Recognition. (2001) CD–ROM
2. Ruzon, M., Tomasi, C.: Alpha estimation in natural images. In: Proc. Conf. Computer Vision and Pattern Recognition. (2000) 18–25
3. Boykov, Y., Jolly, M.P.: Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images. In: Proc. Int. Conf. on Computer Vision. (2001) CD–ROM
4. Greig, D., Porteous, B., Seheult, A.: Exact MAP estimation for binary images. J. Royal Statistical Society **51** (1989) 271–279
5. Kolmogorov, V., Zabih, R.: What energy functions can be minimized via graph cuts? IEEE Trans. on Pattern Analysis and Machine Intelligence **in press** (2003)
6. Besag, J.: On the statistical analysis of dirty pictures. J. Roy. Stat. Soc. Lond. B. **48** (1986) 259–302
7. Winkler, G.: Image analysis, random fields and dynamic Monte Carlo methods. Springer (1995)
8. Kumar, S., Hebert, M.: Discriminative random fields: A discriminative framework for contextual interaction in classification. In: Proc. Int. Conf. on Computer Vision. (2003) CD–ROM
9. Descombes, X., Sigelle, M., Preteux, F.: GMRF parameter estimation in a non-stationary framework by a renormalization technique. IEEE Trans. Image Processing **8** (1999) 490–503
10. Malik, J., Belongie, S., Leung, T., Shi, J.: Contour and texture analysis for image segmentation. Int. J. Computer Vision **43** (2001) 7–27
11. Geman, S., Geman, D.: Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Trans. on Pattern Analysis and Machine Intelligence **6** (1984) 721–741