Journal of Machine Learning Research 4 (2003) 1499-1519 Submitted 10/02; Published 12/03

A Generative Model for Separating Illumination and Reflectance from Images

Inna Stainvas INNA-S@ORBOTECH.COM
FPD Division
Orbotech Ltd.
P.O. BOX 215, Yavne 81102, ISRAEL

David Lowe D.LOWE@ASTON.AC.UK
Information Engineering
Aston University
Aston Triangle, Birmingham. B4 7ET
United Kingdom, Fax: 0121 333 4586

Editors: Te-Won Lee, Jean-François Cardoso, Erkki Oja and Shun-Ichi Amari

Abstract

It is well known that even slight changes in nonuniform illumination lead to a large image variability and are crucial for many visual tasks. This paper presents a new ICA related probabilistic model where the number of sources exceeds the number of sensors to perform an image segmentation and illumination removal, simultaneously. We model illumination and reflectance in log space by a generalized autoregressive process and Hidden Gaussian Markov random field, respectively.

The model ability to deal with segmentation of illuminated images is compared with a Canny edge detector and homomorphic filtering. We apply the model to two problems: synthetic image segmentation and sea surface pollution detection from intensity images.

Keywords: Segmentation, illumination, reflectance, Potts model, general autoregressive model

1. Introduction

It is well known that even slight changes in nonuniform illumination lead to large image variability and are crucial for many visual tasks. Due to the impossibility of controlling illumination conditions in real-world applications, image recognition, retrieval, segmentation and tracking, which are all invariant to illumination, emerge as challenging problems.

There are three main approaches to handling visual tasks under illumination: (i) feature-based, (ii) appearance based and (iii) illumination removal. The first approach is based on selecting local features that are insensitive or at least robust to illumination. The method works by finding discontinuities in the image intensity with the subsequent searching of intensity related features along the edges. When images are colored, the RGB space is transformed to HSI color space (Gonzalez and Wintz, 1993) and intensity is replaced by the hue (chromaticity) value that is less sensitive to illumination (Ohba et al., 2000, Tauber et al., 2000). Though the feature-based methods have been successfully applied to object recognition, segmentation and retrieval, it is clear that local features do not contain all of the required information and it is easy to imagine scenarios when this approach fails.[1]

The second approach is a generative appearance-based method, that exploits the fact that images of an object under all possible illumination variations can be represented as a linear combination of a small set of basis images (Belhumeur and Kreigman, 1998). This approach has been used to address object and face recognition problems. Recognition is based on the smallest projection error onto the specific object representation manifold. Being a generative model approach, it requires the availability of the images taken under different viewpoints and several illumination conditions.

The idea of the third approach is to recover illumination and reflectance from the images (Barrow and Tenenbaum, 1978, Marr, 1982, Horn, 1986, Adelson and Pentland, 1996). It has been suggested that the early steps of visual processing are driven by the goal of recovering "intrinsic" images such as orientation, depth, reflectance and illumination (Barrow and Tenenbaum, 1978). The latter consequently can be used to address high-level visual tasks such as segmentation, classification, recognition, etc. Though it is obvious that recovery of the intrinsic images is an ill-posed problem, it has been contended that this is possible if constraints derived from assumptions about the nature of the scene and the physics of imaging process are taken into account.

However, the exact physics is extremely difficult since it involves many unknown factors such as description of the lighting in the scene, geometry of the real world, optical material properties and camera parameters; secondly, the generative image model is governed by a combination of the Fredholm integral equation (rendering equation) and camera model (Laszlo, 2000). Even a direct rendering problem to provide the illusion of watching real objects on the computer screen is far from being trivial see Laszlo (2000), Schroder (1994) for review.

Currently, reasonable simplifications of the real physics are required. The simplest one used in practice by engineers is to fit an image by a linear surface. This simplification sometimes leads to a certain improvement in visual task performance, but is, obviously, heuristic. Instead another more realistic simplified model that explains an image $f(x,y)$ as:

$$f(x,y) = i(x,y)r(x,y)\cos\alpha(x,y), \tag{1}$$

is usually considered (Barrow and Tenenbaum, 1978). In Equation (1), $i(x,y)$ and $r(x,y)$ stand for the illumination and reflectance components, respectively, and α is an angle between the incident illumination flux and surface normal at the point (x,y). In image processing, this is simplified further to get: $f(x,y) = i(x,y)r(x,y)$, which is valid up to an unimportant normalization constant for almost planar scenes (Gonzalez and Wintz, 1993, see pages 29-31) or, in log space:

$$ln\ f(x,y) = ln\ i(x,y) + ln\ r(x,y) \tag{2}$$

Due to the fact that the logarithm of the image intensity is the sum of the logarithms of the illumination and reflectance components, a high-pass filter of the log of the image intensity can be applied to restore the reflectance component; this processing is called homomorphic filtering (Gonzalez and Wintz, 1993).

Looking more closely at Equation (2), we notice that on the logarithmic scale an image is a linear mixture of the two independent components and finding the reflectance is, in fact, an ICA (Independent Component Analysis) related problem (Everson and Roberts, 2001b). We, therefore,

1. For example, see how a Canny edge detector fails under illumination (Figure 2:c2)

formulate the problem of illumination removal as an ICA problem,[2] where the number of independent sources is larger than the number of sensors. This is clearly an ill-posed problem[3] and an additional prior should be imposed to make the problem well behaved. This prior should express our beliefs about the nature of the illumination and reflectance sources and additionally the problem at hand.

In particular, we are interested in segmentation and tracking problems and assume that illumination is a smooth function while reflectance may be well modeled by a GHMRF (Gaussian Hidden Markov Random Field) model (Zhang et al., 2000). The latter means that the reflectance component as an image consists of the homogeneous objects and background. This is quite a moderate assumption, in that the GMM (Gaussian Mixture Model), which is the degenerative form of the GHMRF, has been successfully used to tackle segmentation and tracking (Carson et al., 1999, Stauffer and Grimson, 1999).[4]

The GHMRF is widely used in image processing in the context of image segmentation (Besag, 1986, Chellappa and Jain, 1993, Celeux et al., 2003). It imposes an important continuity constraint at the segmentation level and additionally controls the image scene structure (as reflectance) rather than just constraining the image intensity distribution to a Gaussian mixture.

The smoothness of the illumination component seems to be a natural assumption for convex objects and remote light source and is an implicit basis of homomorphic filtering.

In summary, we assume that (i) illumination is a smooth function; (ii) reflectance is almost homogeneous within segmentation areas; and (iii) these two processes are independent; and use these assumptions to built a new ICA related model for simultaneous image segmentation and illumination removal, which we refer to as the Illumination/Segmentation Model (ISM).

In recent years, ICA has been widely used to address many problems arising in signal and image processing such as data segmentation, analysis and data decomposition, in particular. According to the ICA approach, the observed data are generated from a mixture of unknown (latent) independent sources. In the context of illumination removal, the number of sources exceeds the number of sensors. This motivates us to consider a generative probabilistic ICA approach. This approach has the advantage that priors regarding the mixing process (the number of sources and the structure of the mixing matrix), the nature of the sources and the noise may be handled in a more principled way under the Bayesian framework.

For example, *positivity* constraints coupled with the building of hierarchical models have been previously used to perform data partitioning and decomposition (Roberts and Choudrey, 2003). Imposing sparseness constraints as a Laplacian source prior has been employed to learn overcomplete representations of signals (M.S.Lewicki and Sejnowski, 2000). In another version of ICA known as the Independent Factor Analysis (IFA) model, the probability density functions of the sources are modeled by a mixture of Gaussians (Attias, 1999).

In these approaches, the ICA model parameters are estimated using the maximum likelihood principle, with the observed data being assumed independently and identically distributed. In our case the observational data is a single non-stationary 2D image with intensities of the neighboring

2. There is no need for a mixing matrix in our formulation, though it may be introduced as scaling coefficients for normalized illumination and reflectance components.

3. For example, if there are two independent light sources, there are at least three explanations to decompose an image as a sum of two independent sources.

4. Stauffer and Grimson (1999) also update the means and variances of GMM by a Kalman filter that tries to cope with varying illumination between video frames. We, however, are more concerned about the varying illumination, when the intensity of the same object or background changes significantly within the scene.

pixels being highly correlated. This fact invalidates an application of the above mentioned methods for our problem.

An alternative image representation, as a collection of small patches, may also be considered but leads to a non-trivial problem of identifying basis functions as representing illumination. Additionally, a model selection problem should be properly addressed in this case, as the number of sources is unknown and may be quite large due to the 2D image being non-stationary. In addition, the positivity constraint is not very well justified for the illumination removal problem since the ICA approach for the image decomposition problem has to be formulated in the logarithmic domain. So, for the illumination removal domain considered in this paper, alternative non-stationary ICA probabilistic models should be considered.

Diverse methods incorporating dynamical dependency have been proposed to generalize ICA for non-stationary time series, (Pearlmutter and Parra., 1997, Attias, 2000, Penny et al., 2000, Everson and Roberts, 2001a). For instance Attias (2000) captures temporal statistical properties of the observed data by describing each source as a hidden Markov Model (HMM) where the source emission probabilities are assumed to be Gaussian. Alternatively Penny et al. (2000) use an HMM to switch between independent components which themselves are modeled as generalized autoregressive processes. Another non-stationary ICA generative model assumes the data to be an instant mixture of independent components with the mixing matrix dynamically evolving in time; and it is additionally contaminated by normally distributed noise (Everson and Roberts, 2001a). The dynamics of the mixing matrix is modeled by a first order Markov process and the source densities are modeled with generalized exponents.

Since an image is a mapping from the 2D spatial domain to the intensity domain, the above mentioned non-stationary ICA methods can not be used for illumination removal. However, they provide us with many ideas that may be adopted to generalize a non-stationary ICA in the context of our problem (see the next section). Below we also briefly review some papers that are relevant to our model.

We note that ICA has been used previously to separate mirror like reflections from images when objects are photographed behind glass (Farid and Adelson, 1999). In this case, the problem is formulated as a standard ICA problem with the number of photographed images (sensors) equal to the number of independent sources[5] and is written for image intensities and not on the log-scale as in Equation (2).

Another model has been proposed by Wells et al. (1996) for adaptive segmentation of MR images. Similar to this paper, a log transform is used to eliminate a slowly varying intensity bias field. However, the Wells et al. model is different from our ISM in that: (i) reflectance in the ISM is modeled as a HGMRF rather than a GMM, (ii) the illumination smoothness constraints are expressed differently; (iii) the ISM parameters are fully estimated in the generalized EM algorithm with the variational E-step, while the GMM means, variances and illumination parameters in the Wells et al. model are assumed to be known or heuristically estimated. It has also been proposed to separate illumination and reflectance from a sequence of images using the fact that illumination images give rise to sparse filter outputs (Weiss, 2001) and assuming reflectance to be practically stationary. This leads to a simple procedure for reflectance recovered as a pixel-wise median through time in the space of the filtered images. However, this method is constrained to its underlying

5. The latter is a picture and a number of objects before the glass.

assumptions: the availability of several images of the same scene and a different temporal scale of changes in illumination and reflectance.

Our illumination/segmentation model (ISM) is introduced in the next section. Section 3 presents a framework for model parameter estimation via the generalized expectation maximization algorithm (GEM). Technical details for the GEM application to the ISM are given in the Appendix. The feasibility of the ISM is studied in Section 4. Two problems with two types of data: (i) a synthetic problem with artificially simulated illumination and (ii) real data of sea-surface images with natural varying illumination across the scene are considered. The paper's results and findings are finally summarized in Section 5.

2. Illumination/Segmentation Model (ISM)

We consider the problem of recovering illumination and reflectance components as an inference of hidden independent sources in a specially introduced generative probabilistic model (see Figure 1). Similar to Attias, (Everson and Roberts, 2001b, pages 95-113), observable data are modeled as a

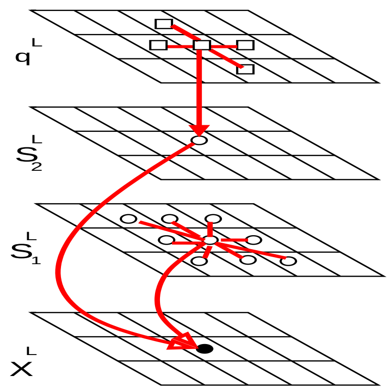

Figure 1: **Illumination/Segmentation Model**: An observed image x^L is a sum of the illumination s_1^L and reflectance s_2^L on the log-scale. The graph is a mixed directed/undirected graph; directed edges (arrows) stand for causal probabilistic dependencies; undirected edges show pixel neighborhood. Only one representative clique per pixel is shown, similar graphs pierce through the observation and source image lattices. The upper segmentation level q^L explains the reflectance as consisting of the homogeneous areas.

sum of two independent sources that are contaminated with isotropic Gaussian noise:

$$
\begin{aligned}
p(x^L, s_1^L, s_2^L, q^L) &= p(x^L|s_1^L, s_2^L)\ p(s_1^L)\ p(s_2^L, q^L), \qquad (3)\\
p(x^L|s_1^L, s_2^L) &= \prod_l^{|L|} p(x^l|s_1^l, s_2^l),\\
p(x^l|s_1^l, s_2^l) &\sim \mathcal{G}(s_1^l + s_2^l, \sigma^2)
\end{aligned}
$$

where x^L, s_1^L, s_2^L stand for intensities of the observable image, illumination and reflectance image components, respectively, in logarithmic space and the upper segmentation level q^L explains reflectance as consisting of homogeneous regions; $\mathcal{G}(\mu,\sigma^2)$ is a normal distribution with a mean μ and variance σ^2 and L is a two-dimensional lattice on which images are defined.

Note that a Gaussian noise model in logarithmic space corresponds to a multiplicative, or so called speckle, noise in the original domain see Chapter 8.13 of Jain (1989), and is a result of the sensitivity variations between the sensors in the CCD cameras.

As has already been discussed in the Introduction, strong constraints on the nature of the sources need to be imposed to solve the ICA problem in the ill-posed case. We assume the illumination to be a smooth and predictable process defined on the lattice L that is expressed by the generalized autoregressive model (GAR):

$$p(s_1^L) \approx \prod_{l=1}^{|L|} p(s_1^l|\mathcal{N}(s_1^l)), \tag{4}$$

$$p(s_1^l|\mathcal{N}(s_1^l)) = \mathcal{G}(e^l,R,\beta) = \frac{R\beta^{1/R}}{2\Gamma(1/R)}\exp(-\beta|e^l|^R),$$

$$e^l = s_1^l - \hat{s}_1^l \quad \hat{s}_1^l = h(v^l) + \sum_{k\in\mathcal{N}(s_1^l)} c_k(s_1^k - h(v^k)). \tag{5}$$

In Equation (5) $h(v^l)$ is a trend of the illumination component in the pixel $l \in L$ with spatial coordinates $v^l \in \mathcal{R}^2$, that is modeled by an RBF network introduced by Broomhead and Lowe (1988). For a textbook reference see Bishop (1995), chapter 5. We use either Thin-Plate Spline (TPS) or Gaussian RBF kernels. In the latter case the Gaussian kernel centers are fixed and located symmetrically outside the lattice L on which the image is defined; kernel variances are also fixed so that only Gaussian kernel basis function tails cover the image lattice.[6] The role of the RBF is to describe a mean of illumination. Modeling illumination by the RBF leads to a very simple and intuitive interpretation that lighting is generated by a set of light sources arranged around the image lattice and with the energy decaying (due to the energy dissipation process) with distance. RBF kernel centers and variances control how far the sources are located from the scene and the choice of the kernel corresponds to the energy dissipation process modeling. The RBF weights correspond to the maximal power of the sources; the weights, however, are not constrained to be non-negative. In Equation (5) $\mathcal{N}(s_1^l)$ are neighbors of the l^{th}-pixel, participating in its intensity prediction and c_k are autoregressive coefficients. Parameter $1/\beta$ controls the width of the exponential distribution for prediction errors e^l and R controls the weight of the distribution tails, see Everson and Roberts (2001b), pages 280–299.

The GAR model with $R = 2$ corresponds to a standard autoregressive (AR) model with Gaussian noise in the error-prediction model. The latter is clearly far from being a true model and, moreover, is known to be very sensitive to outliers (Penny and Roberts, 2002). Therefore, a GAR model that is more flexible than the standard AR model with an arbitrary degree R is considered by us. Intuitively, it is clear that smoothness grows with R, as even moderate errors (> 1) will be amplified. Alternatively, small values of R relax the smoothness constraints. Though it may be beneficial to use the smoothness priors on the GAR model parameters $\mathbf{c}$, R, they are currently assumed to have a uniform, non-informative, prior. We also note that though the RHS of Equation (4) is not properly normalized, this is a widely used approximation in the Markov random field area; see Li (1995), chapter 6, pages 146-148.

6. This is done to satisfy our smoothness prior for illumination surface.

The second reflectance source s_2^l is modeled by a GHMRF model (Zhang et al., 2000):

$$p(s_2^L, q^L) = p(s_2^L|q^L)p(q^L), \quad p(s_2^L|q^L) = \prod_{l=1}^{|L|} p(s_2^l|q^l),$$

where $q^l \in \{1 \dots K\}$ is a discrete hidden variable standing for a segmentation label (a certain object or background) and the emission probability $p(s_2^l|q^l = k)$ is Gaussian $\mathcal{G}(\mu_k, \sigma_k^2)$. The random field for a segmentation level q^L, $p(q^L)$ is given by the Potts model Morris et al. (1996):

$$p(q^L) = \frac{1}{Z(\gamma)} \exp(\gamma \sum_{l=1}^{|L|} \sum_{m \in \mathcal{N}(q^l)} \delta(q^l - q^m),) \tag{6}$$

where $\delta(q^l - q^m) = 1$, when the segmentation latent variables q^l, q^m are in the same state and is zero otherwise. The Potts model imposes contextual constraints on the reflectance; it encourages a representation where neighboring pixels tend to belong to the same segmentation area.

The γ parameter controls the size of the segmentation regions, that are also referred to as blobs, and is assumed to be known. We have also considered a simplified, degenerative case of the Potts model with $p(q^L) = \prod_{l=1}^{|L|} p(q^l)$ and $p(q^l = k)$ being independent of the pixel l, but it leads to worse results and is not discussed here.

3. Learning the ISM

It is typically impossible to conduct an exact Bayesian analysis of model hidden variables and parameters. Below, we briefly review the main results of using variational methods for a maximum likelihood (ML) to approximate an inference of the hidden variables and a parameter point estimation (Section 3.1). We then show how to apply this technique for our learning model (Appendix: Section A).

3.1 Background on the EM Algorithm in Graphical Models

A generalized expectation maximization (GEM) algorithm is used for the ISM parameter estimation and hidden source inference (Jordan, 1999, Opper and Saad, 2001). GEM exploits the graphical structure of the joint probability (probability of the complete data) and is based on variational principles. Let the joint probability of the hidden x_h and observable x_v variables be denoted $f_\Theta(x_v, x_h)$ and let it be factorized into $f(x_v, x_h) = \prod \Psi_{i,\Theta_i}(x_i)$, where x_i is a set of hidden and observable variables in the i-clique ($\{x_v, x_h\} = \bigcup_i x_i$) and Θ_i is a part of the Θ model parameters on which the i-clique potential Ψ_i depends. Then GEM consists of two iterative E (Expectation) and M (Maximization) steps. The *M-step* is given by:

$$\theta_s = \arg\max_{\theta_s} \sum_{i \in C(s)} \langle \ln \ \Psi_{i,\Theta_i}(x_i) \rangle_{Q(x_i)} \tag{7}$$

where $C(s)$ contains all the cliques with potentials dependent on a θ_s parameter and $Q(x_i)$ is a marginal distribution of a distribution $Q(x_h)$ that approximates the true posterior distribution $p(x_h|x_v)$. The generalized E-step is an approximate inference step; an approximate distribution is sought in the form $Q(x_h) = \prod_r Q(x_r)$, where r splits the hidden variables into disjoint groups of "meta" independent variables $\{x_h\} = \bigsqcup_r x_r$ (note that splitting into cliques and independent "meta" variable sets are not the same, in general). The main *variational mean-field* result is given by:

$$Q(x_r) \propto \exp \sum_{i \in D(r)} \langle \ln \Psi_{i,\Theta_i}(x_i) \mid x_r \rangle_{Q(x_i \setminus x_r)}, \tag{8}$$

where $D(r)$ contains all the cliques including at least one of the components of x_r and $x_i \setminus x_r$ is a Markov blanket of x_r in the i-clique (i.e. all hidden variables in the i-clique that do not contain the variable x_r or its components). Equation (8) should be iterated over r until consistency in $Q(x_h)$ is reached and Θ is the set of current parameters found in the M-step. Maximization and inference steps Equations (7) and (8) should be iterated until convergence in $Q(x_h)$ and Θ.

Therefore, to apply GEM one should explore the graphical structure of the joint distribution, consider the log of the potentials and decide on the form of factorization for an approximate posterior distribution $Q(x_h)$; the rest is a technical issue. In our model, cliques are naturally defined and shown graphically in Figures 1 and 6. For the inference step, hidden variables are split into disjoint sets of independent variables according to: $\{s_1^L, s_2^L, q^L\} = \bigsqcup_{l=1}^{|L|}(s_1^l \bigsqcup \{s_2^l, q^l\})$; in other words, illumination and reflectance are assumed to be approximately pixel-wise independent given the observed image. We write down EM in terms of functions $\mathcal{F}_i,\ i = 1 \ldots 3$ corresponding to the log of the probabilities in Equation (3). Technical details are given in the Appendix.

4. Experiments

This section presents two problems with two types of data: (i) a synthetic problem with artificially simulated illumination and (ii) real data of sea-surface images with natural varying illumination across the scene.

4.1 Artificial Illumination

Variable illumination across an image scene is simulated by creating an image according to a three step procedure described below (see Figure 2:a1-a3). First, reflectance in log space is generated by contaminating a binary image with Gaussian noise. A clean binary image with two levels of intensity ($\mu_1 = 0.1$ and $\mu_2 = 0.9$) and Gaussian noise with variable standard deviations per regions $\sigma_1 = 0.01,\ \sigma_2 = 0.02$ are chosen. Secondly, illumination on the log-scale is generated as a polynomial function of degree P with random coefficients: $z = \sum_{p=0}^{P} \sum_{j=0}^{p} a_{j,p-j}^{p} v_1^i v_2^{p-j}$, where $a_{j,p-j}^{p} \sim \mathcal{G}(0,1)$. Illumination z is subsequently normalized according to: $Kz/\max(|z|)$, where K is a normalization constant.[7] Finally, sources are added and contaminated with a global Gaussian noise of zero-mean and $\sigma = 0.05$.

It is easy to see that due to the global noise being Gaussian, the estimation of variances $\sigma_1^2,\ \sigma_2^2,\ \sigma^2$ is an ill-posed problem, as any consistent shift $\sigma_1^2 + a,\ \sigma_2^2 + a,\ \sigma^2 - a$ with any parameter a leaving noise variances positive, leads to the same data generation. The same is true regarding means μ_1, μ_2 and a global constant shift of the illumination surface. As our main goal is to remove illumination

7. We have run experiments with $P \leq 3$; a typical example given in the section below corresponds to $P = 3$ and $K = 5$.

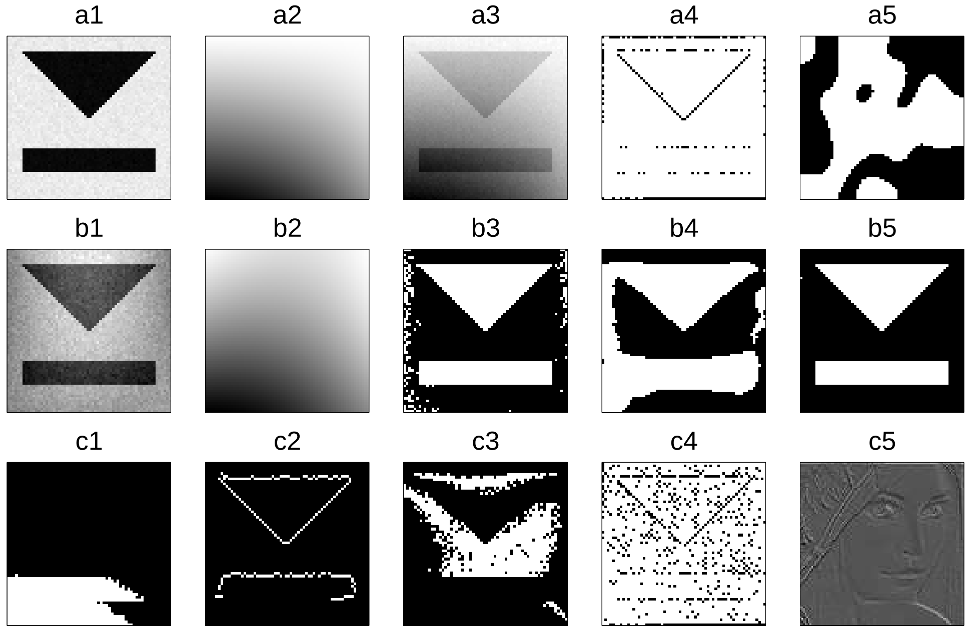

Figure 2: **Artificial illumination**: a1. clean binary image with different Gaussian noise per regions; a2. artificial illumination; a3. artificially illuminated image; *The ISM segmentation for experiments (A):* a4. $\gamma = 0.01$; a5. $\gamma = 1$; *Experiments (B),* $n_h = 32$*:* b1. posterior mean of the reflectance ($\gamma = 0.01$); b2. the RBF trend ($\gamma = 0.01$); b3. the ISM segmentation ($\gamma = 0.01$); b4. the ISM segmentation ($\gamma = 0.2$); b5. segmentation by the GMM for the difference between the observed image and the estimated RBF trend ($\gamma = 0.01$); c1. segmentation by the GMM for the raw data; c2. Canny edge detector; c3. the GMM of the output of the BHPF with $d_0 = 4$ and $g_l = 0.2$; *Experiments (C):* c4. the ISM segmentation ($\gamma = 0.01$); c5. the GAR coefficients as a filter. See also text for an explanation.

and to reveal an image for which (hopefully) segmentation is better than for the raw data, we do not care too much about this type of degeneracy.[8]

4.1.1 RESULTS

Several experiments with different learning and model complexity parameters have been run. The maximal number of iterations per E and M steps have been set to 50, but mean-field usually converges after 10 steps on average (excluding the first two EM steps). The algorithm converges after $5-7$ EM steps dependent on other parameters. The EM algorithm is initialized from the M-step with an observation posterior mean being the observed image; other parameters are randomly initialized. To separate an effect of the RBF trend in the GAR model, experiments are divided into three

8. One can impose constraints on the global variance and illumination surface mean to get a unique solution. In fact, the sensor noise variance is considered to be known and it is taken so that it does not exceed $\sigma^2 + \min(\sigma_1^2, \sigma_2^2,) - \delta$, i.e. a certain part of the image noise is explained by the sensor noise.

groups: (A) without the RBF trend; (B) with the RBF trend and zero autoregressive coefficients; (C) full model with all parameters being free.

It turns out that the ISM without the RBF trend (experimental group A) behaves as an edge detector for small $\gamma < 0.1$ (Figure 2:a4). Indeed, with the omitted RBF trend, the GAR model tries to find the best possible coefficients explaining the given illuminated image itself. Therefore, the GAR model (being responsible for predictability) leads to large errors where sudden (unpredictable) events occur and this obviously happens along the image edges. This error information is then propagated to the posterior reflectance, that also gets unexpected values[9] along the edges and, further on, the GHMRF model is able to detect where sudden (unpredictable) events occur. The Potts model additionally links the found edges with the degree dependent on the blobness parameter γ. The Potts model parameter should be properly set for the segmentation to be successful; the ISM with too large $\gamma > 0.5$ leads to bizarre and artificial segmentation (Figure 2:a5). So far the predictability alone with a non-informative prior on the GAR coefficients has not been sufficient to reveal illumination and reflectance.

Experiments in group (B) have been run with a different number of kernels: $n_h \in \{16, 32, 64\}$ and with a varying γ parameter. The best perceptual segmentation result is obtained for $n_h = 32$. The Potts model parameter γ is set empirically; it was found that for $\gamma \in [0.005\ 0.5]$ segmentation results are quite good (Figure 2:b3); for values larger than $\gamma = 0.5$ regions start to merge (Figure 2:b4) and segmentation becomes worse. The reflectance source (as a posterior mean of the reflectance) is presented in Figure 2:b1; it has more contrast and is more homogeneous than the original image under illumination (Figure 2:a3); the RBF trend (Figure 2:b2) is very similar to the artificial illumination (Figure 2:a2). We have also considered a heuristic segmentation based on the GMM for a difference between the observed image and the RBF trend (Figure 2:b5) and found it to be superior to other segmentation results.

The ISM with the RBF trend and non-zero autoregressive (experimental group C.) coefficients acts as an edge-detector similar to experiments in group (A) (see Figure 2:c4). The RBF trend is not coherent with the illumination in this case; this may be a consequence of the existence of many-local minima in the many-parameter models, i.e. only one possible explanation and not the best one is found. To assess an effect of the GAR coefficients, the errors (e^l evaluated with $h(v^l) = 0$ according to Equation (5)), i.e. the difference between the image and its convolution with the filter dependent on the GAR coefficients $\mathbf{c}$ is presented in Figure 2:c5 for the “Lena” image. This image has been chosen due to its rich edge structure; the presence of many edges of different orientation and intensity enables us to perceive the result of the GAR coefficients visually. Perceptually, this operation is equivalent to band-pass filtering.

In the experiments reported here, the neighborhood system for the GAR model $\mathcal{N}(s_1^l)$ has been defined as the first nearest vertical and horizontal neighbor pixels with respect to the pixel l; and for the Potts model, all the eight nearest pixels around the pixel l.

The ISM's ability to deal with illumination is compared with two other base-line approaches: (i) feature-based with object localization based on edge-detection and (ii) homomorphic filtering with the Butterworth high-pass filter (BHPF) (Gonzalez and Wintz, 1993): $h(\omega) = g_l + \frac{g_h - g_l}{1+[d_0/\sqrt{\omega_1^2+\omega_2^2}]^{2n}}$,

9. If the last two terms under the exponent of the expression for $Q(s_1^l)$ Equation (20) are assumed to be the most significant and the RBF trend is omitted, then $\langle s_1^l / x^L \rangle$ may be shifted to either very large or very small values. This consequently will lead to $\bar{s}_2^l$ in Equation (18) being either very large or small and consequently, to unexpected a_k^l in Equation (19) and, finally, to unexpected $\langle s_2^l / x^L \rangle$ values.

where ω is a complex frequency variable, d_0 is a cutoff frequency; g_h and g_l are maximal and minimal amplitudes of the BHPF corresponding to very high and low (zero) frequencies, respectively. The BHPF parameters are selected based on the best perceptual segmentation. Segmentation results with the Canny edge detector (Canny, 1986) and the BHPF filter are presented in Figures 2:c2-c3, respectively. The best segmentation results are obtained for the ISM with $\gamma = 0.01$, $n_h = 32$ and with segmentation based on the GMM for the difference between the observation and RBF trend.

4.2 Illumination Removal and Segmentation of Sea-surface Images

It is well known that water-borne pollutants, including both natural, such as algal bloom, bacteria and fish oil and leakage from the sea bed, and artificial, such as caused intentionally by ships, generate oily spills (*slicks*) on the water surface. Monitoring and tracking of slick regions is an an important environmental problem. Many government authorities are interested in automatic pollution detection and general assessment of water quality.

As has been demonstrated by Stainvas and Lowe (2001), near-shore slick regions may be efficiently detected from remotely sensed low-platform mounted visible band camera images. This becomes possible due to different light reflectance of the slick and surrounding water surfaces and due to different turbulent water motion characteristics of slick and non-slick regions (slicks have a damping effect on the turbulent water motion). As a result, slick regions generally appear brighter in images than normal wave regions, as they reflect the sky intensity more coherently.

This finding has led to a successful segmentation of sea-surface images based on the Gaussian mixture model (GMM) applied to the tonal (intensity) information (Stainvas and Lowe, 2001). However, this method is very sensitive to illumination varying across the scene when objects with elongated shapes or the background appear inhomogeneous. In vision problems with uncontrolled environments, such as water pollution detection, removal of illumination variations emerges as an essential problem.

As an example, let us consider a typical sea surface image under illumination (Figure 3a). This image has been automatically cropped from the larger image to eliminate sky and clouds and to concentrate on the sea surface region alone. The final cropped image has attributes 313×534 pixels and is longer in the direction parallel to the shore than in the direction from the shore to the horizon. The slicks are the bright regions in the top left part of the image, but also exist in the foreground. The slicks perceptually seem to appear brighter, but the average vertical and horizontal profiles of the image intensity (Figure 3b) clearly reveal the inhomogeneity of the illumination across the scene. It is easy to see (Figure 3a), that the information required for segmentation is rather located in the vertical direction than in the horizontal one. At the same time the vertical intensity profile is badly affected by varying illumination (thin red line in Figure 3b). The varying illumination affects the performance of the GMM considerably and leads to unsatisfactory segmentation (see Figure 3c). In this example, the number of mixture components has been set to $k = 3$; segmentation with $k = 2$ leads to similar or worse results and is not considered here.

In order to separate illumination from the image, the ISM with the RBF trend and zero autoregressive coefficients has been applied. We have used an RBF with Gaussian kernels and the number of kernels set to $n_h = 32$. The neighborhood system is the same as in the experiments with artificial data and the blobness parameter γ has been set to 0.01. The illumination source given by the RBF trend recovered by the ISM as an image and surface are shown in Figures 5a-b, respectively. Obviously, illumination is a nonlinear surface, so that a conventional engineering approach to fit

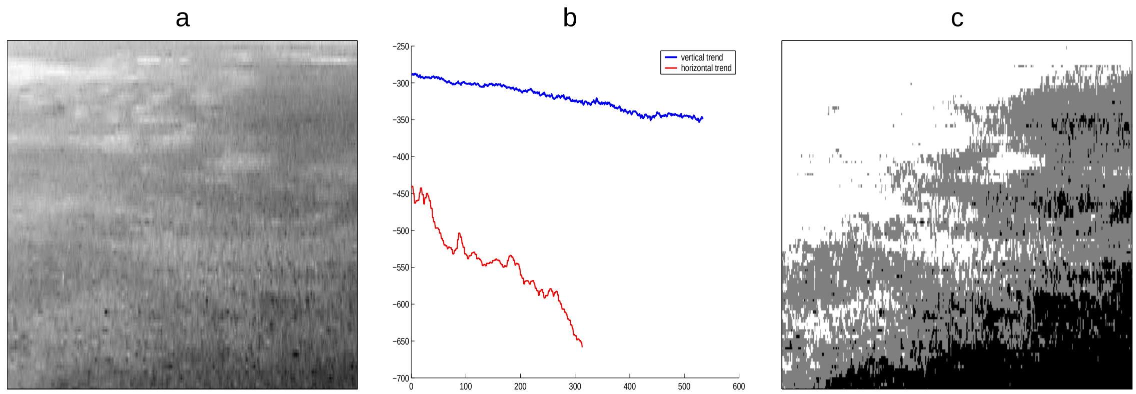

Figure 3: Sea Surface under Illumination: a. A typical image under illumination on the log-scale; b. The intensity profiles for the image in Figure 3a; blue and red colors (thick and thin lines in white/black printing, respectively) correspond to averaging over vertical and horizontal directions, respectively. c. Segmentation of the image in Figure 3a by the GMM with three mixture components. The images (a,c) of Figure 3 have attributes 313×534 pixels, but are scaled to appear square in size.

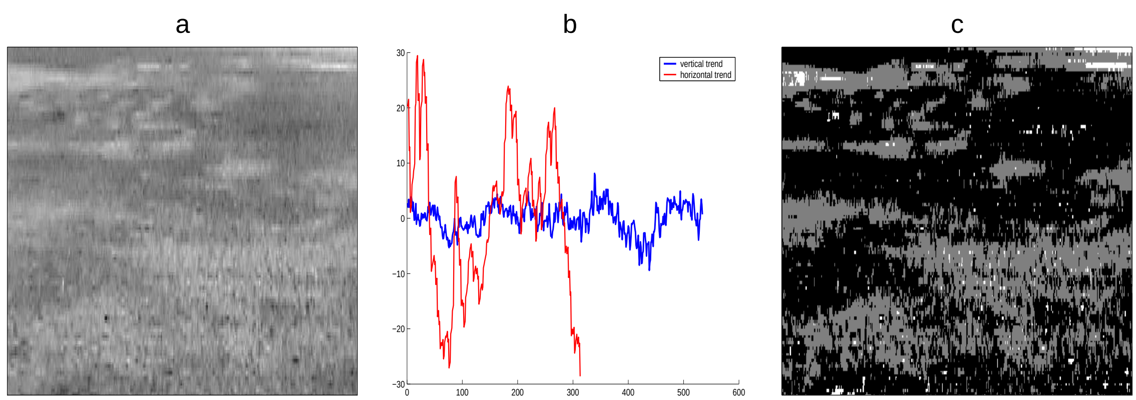

Figure 4: a. Difference between the observed image and the RBF trend; b. Intensity profile for the image in Figure 4a; blue and red colors (thick and thin lines in white/black printing) correspond to averaging over vertical and horizontal directions, respectively. c. Segmentation of the image in Figure 4a by the GMM with three mixture components. The images (a,c) of Figure 4 have attributes 313×534 pixels, but are scaled to appear square in size.

the image with a linear surface will fail. The recovered reflectance as the difference between the observed image and the RBF trend $\vec{s}_2^{\,l}$ is presented in Figure 4a. The obtained "reflectance" image is composed of more uniform regions than the original illuminated image (Figure 3a). The intensity profiles of the "reflectance" image are shown in Figure 4b. This figure clearly demonstrates that the illumination component has been removed from the image. These profiles approximately follow the average locations of the slick regions across the vertical and horizontal directions as corresponding to larger values of the log-intensity. Segmentation by the GMM with three mixture components (Figure 4c) of the "reflectance" image is undoubtedly better than for the original raw image and is coherent with our perceptual expectations.

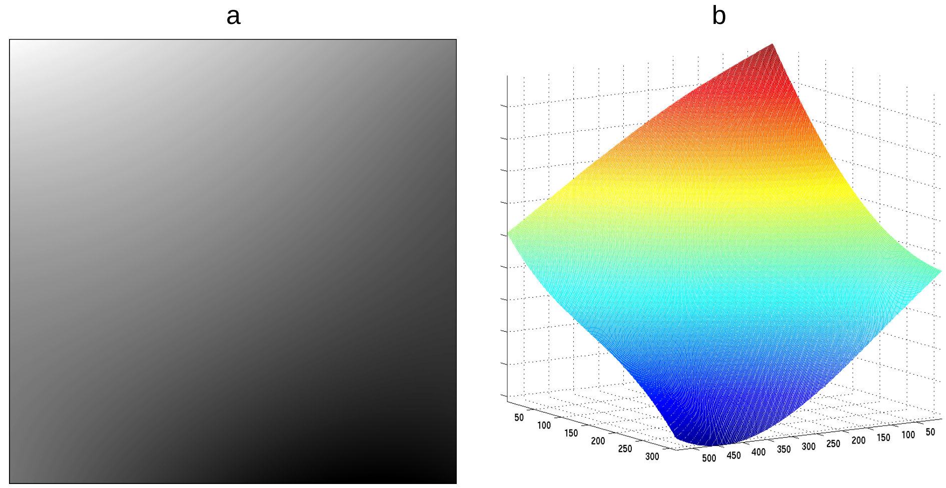

Figure 5: Illumination on the log scale: a. the RBF trend; b. the RBF trend as a surface.

5. Summary and Conclusion

We have introduced a new Illumination/Segmentation Model (ISM) as a probabilistic ICA model for the illumination and reflectance separation. The ISM can be recognized as an ICA model with spatial context and with the number of sources exceeding the number of observations. The ISM works in two regimes: (i) it detects edges for non-zero autoregressive coefficients and (ii) performs segmentation for the zero-autoregressive coefficients, revealing the illumination as the RBF trend. It has been shown that in the regime (ii) the ISM is superior to the Canny edge detector and homomorphic filtering; conventional methods used to deal with illumination.

The ISM has been trained using the GEM with the approximate E and M steps; the quality of this approximation in conjunction with the pseudo-likelihood for the GAR model is an open question. The Potts model parameter γ has been set by trial and error. Unfortunately, it is impossible to calculate the partition function for the Potts model analytically. This makes the estimation of the γ parameter problematic; though it still may be approximately estimated in the M-step if the partition function $Z(\gamma)$ in Equation (6) is ignored. The ISM inference (M-step) can also be performed based on Monte-Carlo methods (Jordan, 1999), but in our opinion, it is preferable to have a tractable analytical expressions instead as the Monte-Carlo methods are slow and the assessment of their convergence is difficult.

It may be worth replacing the GAR model with a Gaussian process (GP). Using GPs solves the problem of working with pseudo-likelihood and does not require approximation in the mean-field and M-steps, so that all analytical expressions are simpler in this case. Moreover, it is interesting to find out how the choice of the GP kernels affects the ISM, since GP kernels implicitly impose different smoothness constraints on the stochastic process.[10]

To control the smoothness of the RBF trend, the Gaussian kernel centers were located outside an image lattice so that only the smooth basis function tails (without "bumps") cover the image plane. It may be interesting to constrain the RBF weights to be positive as it may lead to sparseness, i.e. most of the weights are expected to be zero, giving the most simple explanation of lighting. Another possibility is to use RBF networks with all parameters being flexible, but to regularize the network curvature (Bishop, 1993).

In summary, to be able to reveal illumination: firstly, we assume a prior on the smoothness; secondly, the blobness of the world should be taken into account. As has been shown, predictability enables us only to find contours between the world blobs (segmentation areas). Unfortunately, predictability and smoothness joined together in the ISM do not cooperate; in fact one of them dominates the other. It is an interesting aspect how these two mechanisms cooperate in the human visual processing system.

We are currently working on two alternative approaches to the illumination removal problem based on: (i) using a regularized mixture density network (MDN) (Bishop, 1994, Hjorth and Nabney, 1999); regularization is required as the MDN should work in the switching regime for mixing coefficients, while the means ought to be smooth enough. (ii) using the wavelet best-basis approach (Coifman and Wickerhauser, 1992, Saito, 2001). The idea of the last approach is to find the least dependent bases and then to separate the linear combination of basis functions into two components corresponding to illumination and reflectance sources, where it is known that illumination is smooth and reflectance is good for clustering. The same approach may also be applied to the ICA algorithms proposed by M.S.Lewicki and Sejnowski (2000), Everson and Roberts (2001b), chapter 7.

Acknowledgements

Many thanks to Manfred Opper for helpful discussions and to Ian Nabney for the Netlab software. This work was supported by the EU Framework **V** funded "Blue Water" project.

Appendix A. Technical Details on the ISM Learning

There are four types of clique per pixel in our model (see Figure 6) with the log potentials ($\Phi = \ln \Psi$) given by:

$$\Phi_1^{l;\sigma}(x^l, s_1^l, s_2^l) = -\frac{1}{2\sigma^2}(x^l - (s_1^l + s_2^l))^2 - \ln\sigma \tag{9}$$

$$\Phi_2^{l;\beta,R,\mathbf{c},\mathbf{w}}(s_1^l, \mathcal{N}(s_1^l)) = -\beta|e^l|^R + \ln\frac{R\beta^{1/R}}{2\Gamma(1/R)} \tag{10}$$

$$\Phi_3^{l;\sigma_k,\mu_k}(s_2^l, q^l) = -\frac{1}{2}\left[\frac{(s_2^l - \mu_{q^l})^2}{\sigma_{q^l}^2} + 2\ln\sigma_{q^l}\right] \tag{11}$$

$$\Phi_4^{l;\gamma}(q^l, \mathcal{N}(q^l)) = \gamma \sum_{m \in \mathcal{N}(q^l)} \delta(q^l - q^m), \tag{12}$$

10. Computationally, however, it will be required to find the inverse of the very large kernel dependent covariance matrix, with the size equal to the number of pixels in the image; the inversion operation may be replaced by a quadratic optimization problem (Gibbs and MacKay, 1997).

where $\mathbf{w}$ stands for the standard parameters of the RBF network (Bishop, 1995) and all irrelevant terms such as constants and $Z(\gamma)$ are dropped; the log-potentials $\Phi_1^l, \Phi_2^l, \Phi_3^l, \Phi_4^l$, correspond to $\ln p(x^l|s_1^l, s_2^l)$, $\ln p(s_1^l|\mathcal{N}(s_1^l))$, $\ln p(s_2^l|q^l)$ and l^{th}-potential of the $\ln p(q^L)$, respectively. The sum of the log-potentials $\mathcal{F} = \sum_{k=1}^4 \sum_{l=1}^{|L|} \Psi_k^l$ corresponds to the log-likelihood of the complete data and four additional functions: $\mathcal{F}_k = \sum_{l=1}^{|L|} \Phi_k^l,\ k = 1, \ldots 4$, are introduced for convenience.

A.0.1 M-STEP

The model parameters are found in the M-step by maximizing the expectation of the complete log-likelihood $\mathcal{F}$ over the approximate posterior distribution $Q(s_1^L, s_2^L, q^L)$ and ultimately, according to Equation (7).

It turns out that the sensor noise σ and reflectance model parameters μ_k, σ_k may be found analytically as critical points of $\langle \mathcal{F}_1 /\ x^L \rangle$ and $\langle \mathcal{F}_3 /\ x^L \rangle$ respectively and are given by:

$$
\begin{aligned}
\sigma^2 &= \frac{1}{|L|} \sum_{l=1}^{|L|} (x^l - \langle s_1^l + s_2^l /\ x^L \rangle)^2 + Var(s_1^l /\ x^L) + Var(s_2^l /\ x^L) \\
\mu_k &= \frac{\sum_{l=1}^{|L|} Q(q^l = k) \langle s_2^l /\ q^l = k, x^L \rangle}{\sum_{l=1}^{|L|} Q(q^l = k)} \\
\sigma_k^2 &= \frac{1}{\sum_{l=1}^{|L|} Q(q^l = k)} \sum_{l=1}^{|L|} Q(q^l = k) \cdot \\
&[(\mu_k - \langle s_2 /\ q^l = k, x^L \rangle)^2 + Var(s_2^l /\ q^l = k, x^L)]
\end{aligned}
$$

where $\langle \cdot /\ U \rangle$, $Var(\cdot/U)$ are the conditional mean and variance given the event U and expectations are taken over approximate posterior distributions of the hidden variables.

Illumination model parameters $\beta, R, \mathbf{c}, \mathbf{w}$ are found by gradient like algorithms using the Netlab toolbox (Nabney, 2001). The required derivatives of $\langle \mathcal{F}_2 /\ x^L \rangle$ are given by:

$$\frac{\partial \langle \mathcal{F}_2 /\ x^L \rangle}{\partial \beta} = \frac{|L|}{R} \beta^{-1} - \sum_{l=1}^{|L|} \langle |e^l|^R \ /\ x^L \rangle \tag{13}$$

$$\frac{\partial \langle \mathcal{F}_2 /\ x^L \rangle}{\partial R} = \frac{|L|}{R} [1 + \frac{1}{R} (\phi[\frac{1}{R}] + \ln \beta)] - \beta \sum_{l=1}^{|L|} \langle |e^l|^R \ln |e^l| \ /\ x^L \rangle \tag{14}$$

$$\frac{\partial \langle \mathcal{F}_2 /\ x^L \rangle}{\partial c_k} = -\beta R \sum_{l=1}^{|L|} \langle |e^l|^{R-1} sign(e^l) \frac{\partial e^l}{\partial c_k} /\ x^L \rangle \tag{15}$$

$$\frac{\partial \langle \mathcal{F}_2 /\ x^L \rangle}{\partial w_r} = -\beta R \sum_{l=1}^{|L|} \langle |e^l|^{R-1} sign(e^l) \frac{\partial e^l}{\partial w_r} /\ x^L \rangle, \tag{16}$$

where *sign* stands for the sign function and derivatives of e^l are given according to:

$$
\begin{aligned}
\frac{\partial e^l}{\partial c_k} &= -(s_1^k - h(v_k)), \\
\frac{\partial e^l}{\partial w_r} &= -\frac{\partial h(v_l)}{\partial w_r} - \sum_{k \in \mathcal{N}(s_1^l)} c_k (s_1^k - \frac{\partial h(v_k)}{\partial w_r}).
\end{aligned}
$$

The derivatives $\frac{\partial h(v_k)}{\partial w_r}$ of the standard RBF network can be found in Bishop (1995, for example) and they are already implemented in the Netlab software (Nabney, 2001). Since β and R parameters should be non-negative the constrained optimization is replaced with the unconstrained one on the parameters $\tilde{\beta}$, $\tilde{R}$, such that $\beta = \exp\tilde{\beta}$, $R = \exp\tilde{R}$. The $\phi(\cdot)$ in Equation (14) is the digamma function (Press et al., 2002). In spite of the fact that a critical β value can be found analytically, see Equation (13), we prefer to use gradient like algorithms on all illumination model parameters at once.

The Equations (13-16) require the expectations to be taken over the approximate posterior distribution found in the E-step. Though this integration can be performed numerically, it is assumed that the posterior of the errors e^l is highly peaked around its mean and an approximation $\langle f(s^L)/\ x^L\rangle \approx f(\langle s^L/\ x^L\rangle)$ is used instead, so that only the first and second order statistics of the approximate posterior distribution are required for the next M-step. Intuitively, this approximation should be sufficient since e^l (Equation (5)) is the weighted sum of the neighboring s_1^k variables that are assumed to be a posterior independent and thus, tends to a highly peaked distribution as the neighborhood grows. Though the formulae look complicated due to the $\sum_{k\in\mathcal{N}(s_1^l)} c_k y_k$ terms, the latter are easy to calculate using convolutions.

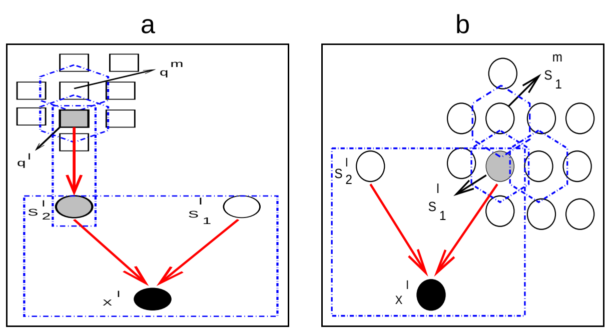

Figure 6: The ISM Cliques: The cliques participating in the inference (mean field step) are shown by dashed polygons. Shaded nodes stand for “meta” independent variables, the black node stands for the observable variable x^l. Non-filled nodes show the Markov blanket for the variable for which approximate posterior is updated. Note that to improve visualization performance not all upper potentials and corresponding blanket variables are shown. The hexagons mark variables on which functions Φ_4 depend, they are not the maximal complete subsets of the graph.

A.0.2 E-STEP: MEAN FIELD SOLUTION

As has already been discussed, approximate posterior distributions given by the mean field theory satisfy Equation (8) written for s_1^l and clipped $\{s_2^l, q^l\}$ variables.

“Meta” variable $\{s_2^l, q^l\}$: To implement Equation (8) the cliques dependent on s_2^l and q^l should be identified. These cliques are graphically shown in Figure 6a, so that it is easy to write down:

$$Q(s_2^l, q^l = k) \propto \exp\langle[\Phi_1^l + \Phi_3^l + \Phi_4^l + \sum_{m\in\mathcal{N}(q^l)} \Phi_4^m]/\ s_2^l, q^l = k\rangle, \tag{17}$$

where expectation is taken over the $\{s_2^l, q^l\}$ blanket that consists of the s_1^l and q^p variables, where index p sweeps over the neighbors of the l-pixel and neighbors of their neighbors; i.e. neighbors of the second order. Since both potentials Φ_1^l, Φ_3^l are quadratic forms on the s_2^l, and q^l is a discrete variable, it is easy to see that the approximate posterior distribution $Q(s_2^l)$ is a mixture of Gaussians (MoG). This is one of the advantages of the graphical model formulation, since it allows us to control and anticipate the final results by looking at the graphical structure of the distribution and the form of the log-potentials.

Substitution of Equations (9-12) into Equation (17) leads to:

$$Q(s_2^l, q^l = k) \propto \exp\{-\frac{1}{2\sigma^2}[(s_2^l)^2 - 2s_2^l\bar{s}_2^l] - \frac{1}{2}[\frac{(s_2^l - \mu_k)^2}{\sigma_k^2} + 2\log\sigma_k] + 2\gamma \sum_{m\in\mathcal{N}(q^l)} Q(q^m = k)\}, \quad \bar{s}_2^l = x^l - \langle s_1^l /\, x^L\rangle. \tag{18}$$

As can be easily seen, Equation (18) can be rewritten in the form $Q(s_2^l, q^l = k) \propto b_k^l\, \mathcal{G}(a_k^l, \Sigma_k^l)$ leading to the following posterior distributions:

$$Q(s_2^l, q^l = k) = Q(q^l = k)Q(s_2^l|\, q^l = k)$$
$$Q(q^l = k) = \frac{b_k^l}{\sum_{k=1}^K b_k^l}, \quad Q(s_2^l|\, q^l = k) = \mathcal{G}(a_k^l, \Sigma_k^l)$$

where parameters b_k^l, a_k^l, Σ_k^l are defined according to:

$$\Sigma_k^l \equiv \Sigma_k = 1/(1/\sigma^2 + 1/\sigma_k^2), \quad a_k^l = \Sigma_k(\frac{\bar{s}_2^l}{\sigma^2} + \frac{\mu_k}{\sigma_k^2})$$
$$b_k^l = \frac{\sqrt{\Sigma_k}}{\sigma_k}\exp -\frac{1}{2}\{\frac{\mu_k^2}{\sigma_k^2} - \frac{(a_k^l)^2}{\Sigma_k} - 2\gamma \sum_{m\in\mathcal{N}(q^l)} Q(q^m = k)\}. \tag{19}$$

In summary, $s_2^l|x^L$ is a MoG with pixel-dependent means and with the same variances for all pixels. The terms $\sum_{m\in\mathcal{N}(q^l)} Q(q^m = k)$ in Equation (19) are calculated using a convolution operation.

In the M-step, the required first and second order statistics of s_2^l are found using the following expressions:

$$\langle s_2^l /\, x^L\rangle = \sum_{k=1}^K Q(q^l = k)a_k^l,$$
$$Var(s_2^l|x^L) = \langle (s_2^l)^2 /\, x^L\rangle - \langle s_2^l /\, x^L\rangle^2$$
$$\langle (s_2^l)^2 /\, x^L\rangle = \sum_{k=1}^K Q(q^l = k)\langle (s_2^l)^2 /\, q^l = k, x^L\rangle,$$
$$\langle (s_2^l)^2 /\, q^l = k, x^L\rangle = \Sigma_k + (a_k^l)^2,$$

or finally:

$$Var(s_2^l|x^L) = \sum_{k=1}^K Q(q^l = k)(\Sigma_k + (a_k^l)^2) - (\sum_{k=1}^K Q(q^l = k)a_k^l)^2.$$

Segmentation: It is well known (Bishop, 1995), that when misclassification loss functions are the same for all classes the best classifier minimizing the misclassification error is the Bayesian classifier that assigns class labels according to a maximal posterior distribution $p(q^L|x^L)$. The posterior distribution can not be calculated exactly and instead its approximation is used so that the approximate segmentation is performed pixel-wise according to a maximal approximate posterior $k^\star = \arg\max_k Q(q^l = k)$.

Illumination variable $\mathbf{s_1^l}$*:* Cliques dependent on the s_1^l variable are shown in Figure 6b. Similar to the Potts model, the Markov blanket of s_1^l consists of s_2^l and the neighbors of the second order for s_1^l. Equation (8) for the variable s_1^l is given by:

$$Q(s_1^l) \propto \exp\langle[\Phi_1^l + \Phi_2^l + \sum_{m \in \mathcal{N}(s_1^l)} \Phi_2^m] / \, s_1^l, x^L\rangle$$

Once again an approximation is used to calculate these expectations and the final result has the following form:

$$\begin{aligned} Q(s_1^l) &\propto \exp\{-\frac{1}{2\sigma^2}(s_1^l - \bar{s}_1^l)^2 - \beta|\tilde{e}^l|^R - \beta \sum_{m \in \mathcal{N}(s_1^l)} |\tilde{e}^m|^R\}, \\ \bar{s}_1^l &= x^l - \langle s_2^l / \, x^L\rangle, \end{aligned} \tag{20}$$

where $\tilde{e}^l$ is calculated the same as in Equation 4, but with

$$\hat{s}_1^l = h(v^l) + \sum_{k \in \mathcal{N}(s_1^l)} c_k(\langle s_1^k / \, x^L\rangle - h(v^k))$$

and $\tilde{e}^m = \langle s_1^m / \, x^L\rangle - \hat{s}_1^m$, where

$$\hat{s}_1^l = h(v^l) + c_{lm}(s_1^l - \langle s_1^l / \, x^L\rangle) + \sum_{k \in \mathcal{N}(s_1^m)} c_k(\langle s_1^k / \, x^L\rangle - h(v^k))$$

and c_{lm} is one of the GAR coefficients $\mathbf{c}$ which corresponds to the s_1^l term in the potential Φ^m. The expressions including summation over the neighborhood are calculated using convolution and shift operations. In summary, the approximate posterior of the illumination variable is a Gibbs distribution that is expected to be sharply peaked near $\bar{s}_1^l$.[11] An adaptive Simpson quadrature numerical algorithm[12] (Press et al., 2002) is used to calculate the first and second order statistics of the s_1^l with the integration interval defined as $[\bar{s}_1^l - \Delta \;\; \bar{s}_1^l + \Delta]$, where Δ is taken sufficiently large.[13]

References

E. H. Adelson and A. P. Pentland. The perception of shading and reflectance. In D. Knill and W. Richards, editors, *Perception as Bayesian Inference*, pages 409–423. New York: Cambridge University Press, 1996.

11. The first term under the exponent in the expression for $Q(s_1^l)$ (Equation (20)) is expected to be the most significant since the errors should be small.
12. The Matlab "quad.m" function has been adapted to calculate integrals at once as an image.
13. A value equal to 10 has been used in experiments here.

H. Attias. Independent factor analysis. *Neural Computation*, 11(4):803–851, 1999.

H. Attias. Independent factor analysis with temporally structured factors. In Leen et al., editor, *Advances in Neural Information Processing Systems*, volume 12, pages 73–80. MIT Press, Cambridge, MA, 2000.

H. G. Barrow and J. M. Tenenbaum. Recovering intrinsic scene characteristics from images. In A. Hanson and E. Riseman, editors, *Computer Vision Systems*, pages 3–26. New York: Academic Press, 1978.

P. Belhumeur and D. Kreigman. What is the set of images of an object under all possible illumination conditions? *Int. Journal of Computer Vision,*, 28(3):245–269, 1998.

J. Besag. On the statistical analysis of dirty pictures. *Journal Royal Statistical Society, Series B*, 48: 259–302, 1986.

C. Bishop. Curvature-driven smoothing: A learning algorithm for feed-forward networks. *IEEE Transactions on Neural Networks*, 4(5):882–884, 1993.

C. Bishop. Mixture density networks. Technical report, ncrg/94/004, NCRG, Aston University, 1994.

C. M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995. Available electronically from `http://neural-server.aston.ac.uk/NNPR/`.

D. S. Broomhead and D. Lowe. Multivariable functional interpolation and adaptive networks. *Complex Systems*, 2:321–355, 1988.

J. Canny. A computational approach to edge detection. *PAMI*, 8(6), November 1986.

C. Carson, S. Belongie, H. Greenspan, and J. Malik. Blobworld: A system for region-based image indexing and retrieval. In *Third Int. Conf. on Visual Information Systems*, pages 1–8. Springer-Verlag, Junes 1999. Available electronically from `http://www.cs.berkeley.edu/~carson/research/publications.html`.

G. Celeux, F. Forbes, and N. Peyrard. EM procedures using mean filed-like approximations for Markov model-based image segmentation. *Pattern Recognition*, 36:131–144, 2003.

R. Chellappa and A. Jain. *Markov Random Fields: Theory and Applications*. Academic Press, 1993.

R. R. Coifman and M. Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Trans. Info. Theory*, 38(2):713–719, 1992. Available electronically from `http://wuarchive.wustl.edu/doc/techreports/wustl.edu/math/papers/entbb.ps.Z`.

R. Everson and S. Roberts. Particle filters for non-stationary ICA. In Richard Everson and Stephen Roberts, editors, *ICA: Principles and Practice*, pages 280–298. Cambridge University Press, Cambridge, 2001a.

Richard Everson and Stephen Roberts, editors. *ICA: Principles and Practice*. Cambridge University Press, Cambridge, 2001b.

H. Farid and E. H. Adelson. Separating reflections from images using independent component analysis. *Journal of the Optical Society of America*, 9(1):2136–2145, 1999.

M. N. Gibbs and David J. C. MacKay. Efficient implementation of Gaussian processes. Available electronically from `http://www.cs.toronto.edu/~mackay/gpros.ps.gz`. Preprint, 1997.

R. C. Gonzalez and P. Wintz. *Digital Image Processing*. Addison-Wesley Publishing Company, 1993.

L. U. Hjorth and I. T. Nabney. Regularisation of mixture density networks. In *International Conference on Artificial Neural Networks, Edinburgh*, pages 521–526, 1999.

B. K. P. Horn. *Robot Vision*. The MIT Press, New York, 1986.

A. K. Jain. *Fundamentals of Digital Image Processing*. Prentice Hall, London, 1989.

M. Jordan, editor. *Learning in Graphical Models*. The MIT Press, Cambridge, Massachusetts, London, England, 1999.

S-K. Laszlo. Monte-Carlo Methods in Global Illumination. Available electronically from `http://www.iit.bme.hu/~szirmay/puba.html`. Book Preprint, 2000.

S. Z. Li. *Markov Random Field Modeling in Computer Vision*. Springer, 1995.

D. Marr. *Vision*. Imprint FREEMAN, New York, 1982.

R. Morris, X. Descombes, and J. Zerubia. Fully Bayesian Image Segmentation - an Engineering Perspective. Technical report, n 3017, INRIA, October 1996. Available electronically from `name@sophia.inria.fr`.

M.S.Lewicki and T. J. Sejnowski. Learning overcomplete representations. *Neural Computation*, 12 (2):337–365, 2000. Available electronically from `http://www.cnl.salk.edu/~lewicki/`.

I. T. Nabney. *Advances in Pattern Recognition*. Springer, London, 2001.

K. Ohba, Y. Sato, and K. Ikeuchi. Appearance-based visual learning and object recognition with illumination invariance. *Machine Vision and Applications*, 12(4), 189-196 2000. Available electronically from `http://www.aist.go.jp/MEL/soshiki/robot/biorobo/ooba/cv.html`.

Manfred Opper and David Saad, editors. *Advanced Mean Field Methods*. The MIT Press, Cambridge, 2001.

B. A. Pearlmutter and L. C. Parra. Maximum likelihood blind source separation: A context-sensitive generalization of ica. In D. S. Touretzky and R. P. Lippmann, editors, *NIPS*, volume 9, pages 613–619. MIT Press, 1997.

W. Penny, R. Everson, and S. Roberts. Hidden markov independent component analysis. In M. Girolami, editor, *Advances in Independent Component Analysis*. Springer, 2000.

W. Penny and S. Roberts. Variational Bayes for generalised autoregressive models. *To appear in IEEE Transactions on Signal Processing*, pages 1–22, 2002.

William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing*. Amazon, Cambridge, 2002. second edition.

S. Roberts and R. Choudrey. Data decomposition using independent component analysis with prior constraints. *Pattern Recognition*, 36(8), 2003. Available electronically from `http://www.robots.ox.ac.uk/~parg/publications.html`.

N. Saito. Image approximation and modeling via least statistically dependent bases. *Pattern Recognition*, 34:1765–1784, June 2001.

Peter Schroder. *Wavelet Algorithms for Illumination Computation*. PhD thesis, Computer Science, Princeton University, 1994. Available electronically from `http://www.math.tau.ac.il/~stainvas`.

I. Stainvas and D. Lowe. Towards sea surface pollution detection from visible band images. *IEICE Transactions on Electronics a Special Issue on New Technologies in Signal Processing for Electromagnetic-wave Sensing and Imaging*, E84C(12):1848–1856, December 2001. Available electronically from `http://www.www.ncrg.aston.ac.uk/~stainvai/spolb.ps.gz`.

C. Stauffer and W.E.L Grimson. Adaptive background mixture model for real-time tracking. In *IEEE Computer Society Conference om Computer Vision and Patten Recognition, Cat. No. PR00149*, volume 2, page IEEE Comput. Soc. part, 1999. Available electronically from `http://www.ai.mit.edu/projects/vsam/Tracking/`.

Z. Tauber, Ze-Nian Li, and M. S. Drew. Local-based visual object retrieval under illumination cnange. In *Proceedings of the Computer Vision and Pattern Recognition*, volume 4, pages 43–46, Barcelona, 2000.

Y. Weiss. Deriving intrinsic images from image sequences. In *Int. Conf. Computer Vision*, number 8, pages 1–8, Vancouver, Canada, 2001.

W. M. Wells, R. Kikins, W. Grimson, and F. Jolesz. Adaptive segmentation of MRI data. *IEEE Transactions on Medical Imaging*, (15):429–442, 1996.

Y. Zhang, S. Smith, and M. Brady. A Hidden Markov Random Field Model for Segmentation of Brain MR Images. In *SPIE Medical Imaging*, pages 1126–1137, San-Diego, USA, February 2000. Available electronically from `http://www.fmrib.ox.ac.uk/~yongyue/mriseg.html`.