Sampling, resampling and colour constancy

D.A. Forsyth

Computer Science Division
U.C. Berkeley
Berkeley, CA 94720
daf@cs.berkeley.edu

Abstract

We formulate colour constancy as a problem of Bayesian inference, where one is trying to represent the posterior on possible interpretations given image data. We represent the posterior as a set of samples, drawn from that distribution using a Markov chain Monte Carlo method. We show how to build an efficient sampler.

This approach has the advantage that it unifies the constraints on the problem, and represents possible ambiguities. In turn, a good description of possible ambiguities means that new information, instead of producing contradictions, is easily incorporated by resampling existing samples. The method is demonstrated on the case where surfaces seen in two distinct images are later discovered to be the same. We show examples using images of real scenes. **Keywords:** *Colour constancy, Markov chain Monte Carlo, Probabilistic reasoning, Inference*

The image appearance of a set of surfaces is affected both by the reflectance of the surfaces and by the spectral radiance of the illuminating light. Recovering a representation of the surface reflectance from image information is called *colour constancy*. Computational models customarily model surface reflectances and illuminant spectra by a finite weighted sum of basis functions and use a variety of cues to recover reflectance, including (but not limited to!): specular reflections [9]; constant average reflectance [2]; illuminant spatially frequency [8]; low-dimensional families of surfaces [10]; physical constraints on reflectance and illumination coefficients [4, 3]. Each cue has well-known strengths and weaknesses. The most complete recent study appears to be [1], which uses the cues to make Bayesian decisions that maximise expected utility, and compares the quality of the decision; inaccurate decisions confound recognition [5].

A reliable estimate of surface colour is possible only under highly specialised conditions, but representations of possible error and ambiguity are far more important for system building. For example, the hypothesis that a fire engine is present is probably insupportable unless the colour representation could be red (or green, in some states!). Furthermore, if we decide a fire-engine really is present, we should be able to use that information adjust our estimate of surface colour for other surfaces. These observations justify studying colour constancy in a framework of probabilistic inference. This approach has the advantage that, instead of comparing cues, we can *combine* them.

Notation: we write $\underline{\epsilon}$ for a vector, whose i'th component is ϵ_i and $\underline{\underline{M}}$ for a matrix whose i, j'th component is M_{ij}.

Comment: Current CVPR production does not allow for colour figures. A version of this paper with colour figures can be found at the author's home page `http://www.cs.berkeley.edu/~daf`.

1 Bayesian inference by sampling

The great importance of Bayes' rule is that *generative* models — which give the way that data is produced, given the state of the world — can be turned into recognition models just by multiplying by the prior. All but the silliest choices of prior are overwhelmed by data in the kinds of problem we wish to solve. Information can be extracted from the posterior by drawing a large number of samples from that distribution using Markov chain Monte Carlo methods. A typical MCMC algorithm is the Metropolis-Hastings algorithm, which would produce in this case a sequence of hypotheses, by taking an hypothesis of state T_i and proposing a revised version, T_i'. The new hypothesis T_{i+1} is either T_i or T_i', depending (randomly) on how much better the posterior associated with T_i' is. In particular, we accept the new state with

0-7695-0149-4/99 $10.00 © 1999 IEEE

probability

$$\alpha = \min(1, \frac{\text{Prop}(T_i' \rightarrow T_i)\text{Posterior}(T_i')}{\text{Prop}(T_i \rightarrow T_i')\text{Posterior}(T_i)}$$

where $\text{Prop}(T_i \rightarrow T_i')$ is the probability of proposing state T_i' given one is in state T_i.

Once sufficient iterations have completed, all subsequent T_i are samples drawn from the posterior; the number of iterations required to achieve this is often called the *burn in* time. These samples may or may not be correlated; if this correlation is low, the method is said to *mix* well. The formalism is extremely broad with respect to the structure of the state space (for example, the dimension of the state space may not be known *a priori*) [6]. Note the convenient fact that the algorithm uses only ratios of posteriors, so that the normalising constant is irrelevant.

Metropolis-Hastings algorithms should be viewed as a kind of souped up hypothesize and test process, so that current vision algorithms are a natural source of proposals. The crucial improvement is that we can use different, incompatible algorithms as distinct sources of proposals, and the samples we obtain represent the posterior and so incorporate all available measurements.

2 Probabilistic colour constancy

We assume that surfaces are flat, so that there is no shading variation due to surface orientation and no interreflection. There are four components to our model:

A viewing model: we assume a perspective view of a flat, frontal surface, with the focal point positioned above the center of the surface. As spatial resolution is not a major issue here, we work on a 50 x 50 pixel grid for speed.

A spatial model of surface reflectances: because spatial statistics is not the primary focus of this paper, we use a model where reflectances are constant in a grid of boxes, where the grid edges are not known in advance. A natural improvement would be the random polygon tesselation of [7].

A spatial model of illumination: for the work described in this paper, we assume that there is a single point source whose position is uniformly distributed within a volume around the viewed surface.

A rendering model: which determines the receptor responses resulting from a particular choice of illuminant and surface reflectance; this follows from standard considerations.

2.1 The rendering model

We model surface reflectances as a sum of basis functions $\phi_j(\lambda)$, and assume that reflectances are piecewise constant:

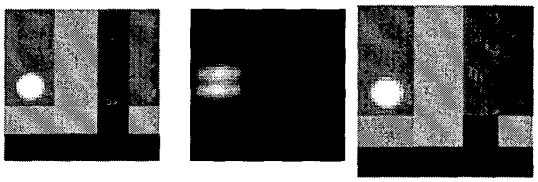

Figure 1: **Left:** *a typical synthetic Mondrian, rendered using a linear intensity scale that thresholds the specularity.* **Center:** *the proposal distribution for x and y position of the specularity, obtained by image filtering and shown with the highest value white.* **Right:** *a rendering of a typical sample for this case, using the sample's illuminant; a successful sampler produces samples that look like the image.*

$$s(x, y, \lambda) = \sum_{j=0}^{n_s} \sigma_j(x, y)\phi_j(\lambda)$$

Here $\sigma_j(x, y)$ are a set of coefficients that vary over space according to the spatial model.

Similarly, we model illuminants as a sum of (possibly different) basis functions ψ_i and assume that the spatial variation is given by the presence of a single point source positioned at $\underline{p}$. The diffuse component due to the source is $e_d(x, y, \lambda, \underline{p}) = d(x, y, \underline{p}) \sum_{i=0}^{n_e} \epsilon_i \psi_i(\lambda)$ where ϵ_i are the coefficients of each basis function and $d(x, y, \underline{p})$ is a gain term that represents the change in brightness of the source over the area viewed. The specular component due to the source is: $e_m(x, y, \lambda, \underline{p}) = m(x, y, \underline{p}) \sum_{i=0}^{n_e} \epsilon_i \psi_i(\lambda)$ where $m(x, y, \underline{p})$ is a gain term that represents the change in specular component over the area viewed.

Standard considerations yield a model of the k'th receptor response as:

$$p_k(x, y) = d(x, y, \underline{p}) \sum_{i,j} g_{ijk}\epsilon_i\sigma_j(x, y) + m(x, y, \underline{p}) \sum_i h_{ik}\epsilon_i$$

where

$g_{ijk} = \int \rho_k(\lambda)\psi_i(\lambda)\phi_j(\lambda)d\lambda$, $h_{ik} = \int \rho_k(\lambda)\psi_i(\lambda)d\lambda$ and $\rho_k(\lambda)$ is the sensitivity of the k'th receptor class. The illuminant terms $d(x, y, \underline{p})$ and $m(x, y, \underline{p})$ follow from the point source model; $m(x, y, \underline{p})$ is obtained using Phong's model of specularities.

Our model of the process by which an image is generated is then: sample the number of reflectance steps in x and in y (k_x and k_y respectively) from the prior; now sample the position of the steps ($\underline{e}_x$ and $\underline{e}_y$ respectively) from the prior; for each tile, sample the reflectance for that interval from the prior (σ_j^m for the m'th tile; sample the illuminant coefficients ϵ_i from

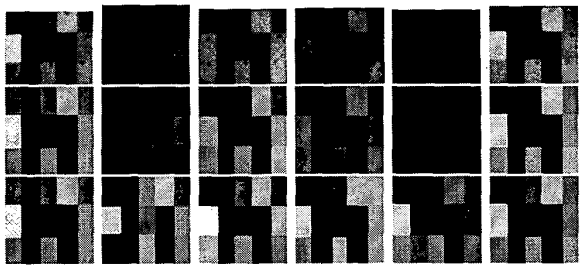

Figure 2: **Top:** images of the same set of patches on a Mondrian of coloured paper patches, photographed under white, blue, purple, red, aqua and yellow light and scanned from [4], used as inputs to the sampler. **Center:** renderings of typical representations obtained by the sampler, in each case shown under the coloured light inferred (so that in a successful result, the inferred representation looks like the image above it). Note the accuracy of the spatial model, and the robustness to image noise. **Bottom:** renderings of typical representations under the same light, so that a successful result implies similar renderings.

the prior; sample the illuminant position $\underline{p}$ from the prior; and render the image, adding Gaussian noise of known standard deviation. This gives a likelihood, $P(\text{image}|k_x, k_y, \underline{e}_x, \underline{e}_y, \sigma_j^m, \epsilon_i, \underline{p})$ The posterior is proportional to:

$$P(\text{image}|k_x, k_y, \underline{e}_x, \underline{e}_y, \sigma_j^m, \epsilon_i, \underline{p})\pi(kv_x)\pi(kv_y)\pi(\underline{s}_x) \times \pi(\underline{s}_y) \prod_{m \in \text{tiles}} \pi(\sigma_j^m)\pi(\epsilon_i)\pi(\underline{p})$$

where we write π for a prior.

2.2 Priors and practicalities

We specify the spatial model by giving the number of edges in the x and y direction separately, the position of the edges, and the reflectances within each block. We assume that there are no more than seven edges (8 patches) within each direction, purely for efficiency. The prior used is a Poisson distribution, censored to ensure that all values greater than seven have zero prior, and rescaled. Edge positions are chosen using a hard-core model: the first edge position is chosen uniformly; the second is chosen uniformly, so that the number of pixels between it and the first is never fewer than five; the third is chosen uniformly so that the number of pixels between it and the second and between it and the first is never fewer than five; and so on. This hard-core model ensures that edges are not so close together that pixel evidence between edges is moot.

The constraints on reflectance and illuminant coefficients are encoded in the prior. We use a prior that is constant within the constraint set and falls off exponentially with an estimate of distance from the constraint set. Because the constraint sets are convex,

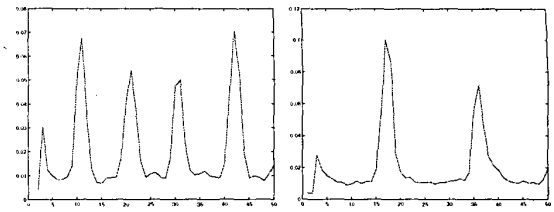

Figure 3: The proposal distribution for edge birth in the x direction (left) and the y direction (right) for the Mondrian imaged under white light in figure 2. Note that the filtering process leads to strong peaks near the edges; this means that the proposal process is relatively efficient, but does not completely rule out edges away from strong responses, if other evidence can be found for their presence (the likelihood component of the posterior).

they can be expressed as a set of linear inequalities; for surface reflectance we have $\underline{\underline{C}}_s\underline{\sigma} + \underline{b} > \underline{0}$ and for illuminant we have $\underline{\underline{C}}_i\underline{\epsilon} > \underline{0}$. If the coefficients in these inequalities are normalised (i.e. the rows of the matrices are unit vectors), then the largest negative value of these inequalities is an estimate of distance to the constraint set.

We use six basis elements for illumination and reflectance so that we can have (for example) surfaces that look different under one light and the same under another light. This phenomenon occurs in the real world; our exploration of ambiguities should represent the possibility. We *represent* surface colour by the colour of a surface rendered under a known, white light.

3 Sampling the posterior

Proposals are made by a mixture of five distinct move, chosen at random. The probability of proposing a particular type of move is uniform, with the exception that when there are no edges, no deaths are proposed, and when the number of edges in a particular direction is at a maximum, no births are proposed. An important advantage to this approach is that, *within each move*, we can assume that the values of variables that we are not changing are correct, and so apply standard algorithms to estimate other values. We omit calculations — which follow the lines of [6] — for lack of space.

Birth of an edge: For each direction, we apply a derivative of Gaussian filter to the red, green and blue components of the image and then divide the response by a weighted average of the local intensity; the result is squared and summed along the direction of interest. This is normalised to 0.8, and 0.2 of a uniform distribution is added. This process produces a proposal distribution that has strong peaks at each edge, and

at the specularity, but does not completely exclude any legal edge point. For a given state, this proposal distribution is zeroed for points close to existing edges (for consistency with the hard core model), and a proposed new edge position is chosen from the result. A new reflectance must be chosen for each of the new patches created by the birth of an edge. Currently, we average the receptor responses within each new patch, and then use the (known) illuminant to estimate a reflectance that comes as close as possible to achieving this average value, while lying within the constraint set. We then add a Gaussian random variable to the estimated reflectance value.

Moving the light: Proposals for a new x, y position for the light are obtained by filtering the image. We apply a filter whose kernel has the same shape as a typical specularity and a zero mean to the r, g and b components separately; the responses are divided by mean intensity, and the sum of squared responses is rescaled to form a proposal distribution. The kernel itself is obtained by averaging a large number of specularities obtained using draws from the prior on illuminant position. Proposals for a move of the light in z are uniform, within a small range of the current position.

Death of an edge: The edge whose death is proposed is chosen uniformly at random. The death of an edge causes pairs of surface patches to be fused; the new reflectance for this fused region is obtained using the same mechanism as for a birth.

Moving an edge: An edge to move is chosen uniformly at random. Within the region of available points (governed by the hard-core model — the edge cannot get too close to the edges on either side of it) a new position is proposed uniformly at random. This is somewhat inefficient, compared with the use of filter energies as a proposal distribution. We use this mechanism to avoid a problem posed by a hard-core model; it can be difficult for a sampler to move out of the state where two edges are placed close together and on either side of a real edge. Neither edge can be moved to the real edge — the other repels it — and a new edge cannot be proposed in the right side; furthermore, there may be little advantage in killing either of the two edges. Proposing uniform moves alleviates this problem by increasing the possibility that one of the two edges will move away, so that the other can move onto the right spot.

Change reflectance and illumination: It is tempting to use a Gibbs sampler, but the chain moves extremely slowly if we do this, because Gibbs sampling is well known to behave badly with correlated

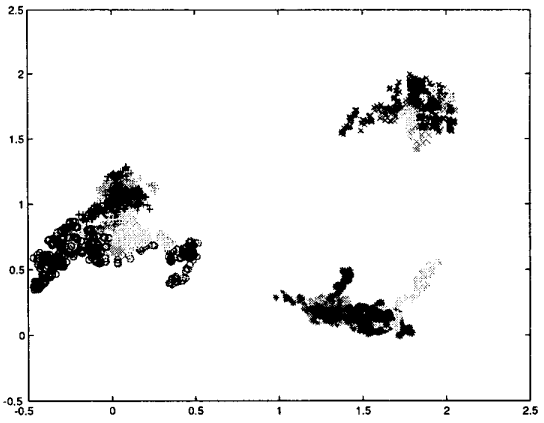

Figure 4: The first two components of surface reflectance samples, plotted on the same axes for four different surfaces. Each sample is colour keyed to the image from which it was obtained; red samples for the red image, etc, with black corresponding to the white image. The circles show samples of the reflectance coefficients for the blue surface at the top left corner of the Mondriaan; the stars for the yellow surface in the second row; the plusses show samples for the orange surface in the top row of the Mondriaan and the crosses for the red surface in the bottom row. Notice that the smear of samples corresponding to a particular surface in one image intersects, but is not the same as, the smear corresponding to that surface in another. This means that the representation envisages the possibility of their being the same, but does not commit to it.

variables (which explains its poor reputation in vision circles). Instead, we sample reflectance and illumination simultaneously using a method due to [12]. The state space for this proposal is reflectance and illumination; we append a set of independent Gaussian random variables ($\underline{w}$) of no external significance to obtain a posterior of the form $exp(-G(\underline{\sigma}_1, \underline{\sigma}_2 \ldots \underline{\epsilon}_1, \underline{\epsilon}_2 \ldots) - (1/2)\underline{w}^T\underline{w})$. The log of this posterior has the form of the Hamiltonian for a particle in an energy field. We now use two types of proposal move: advance time for this particle; and choose new momenta (which can be done by Gibbs sampling, because each w_i is independent of every other, and of the state variables). This method moves to maxima of the posterior about as fast as gradient descent, and then samples around the maxima. If the state is far from a maximum, then the state moves down the energy field, gathering momentum, which is then thrown away by the second type of move, so the particle will tend to get trapped in maxima and explore them.

Burn in and mixing: The sampler described here has been run on many synthetic images where "ground truth" is known, and in each case reaches a small neighbourhood of ground truth — i.e. "burns in" — within about 1000 samples. The experimen-

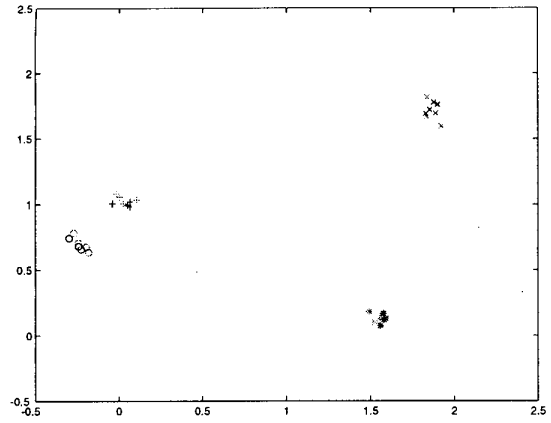

Figure 5: The first two components of surface reflectance samples, plotted on the same axes for four different surfaces. These come from the samples of figure 4, *resampled* under the assumption that the blue surface in the top left hand corner of the Mondriaan is the same for each image. We use the same representation and axes as in that figure. Notice that this single piece of information hugely reduces the ambiguity in the representation.

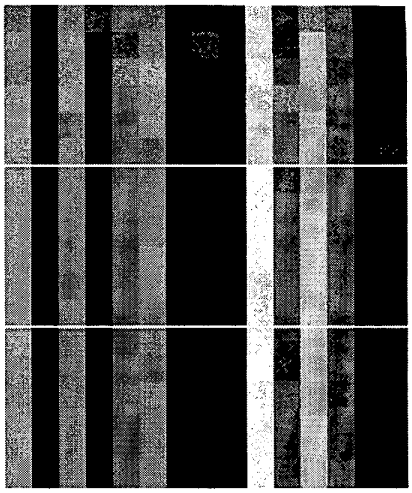

Figure 6: On the top, samples of reflectances returned for each patch on the Mondrian using the images of figure 2, under each light, rendered under white light. There are four hundred samples per patch and per illuminant, each rendered as a small square; thus, a patch for which there is very little information shows a salt-and-pepper style texture. The columns show samples for the same patch under different illuminants; each row corresponds to an illuminant (in the order aqua, blue, purple, red, white and yellow). Notice the very substantial variation in appearance; white pixels denote samples which saturated. Notice also that for each patch there are samples that look similar. The center figure illustrates the samples obtained when all samples are resampled, assuming that the right (blue) patch is the same patch in each image. The bottom figure illustrates the samples obtained when all samples are resampled, assuming that the sixth (yellow) patch is the same patch in each image. Notice the substantial reduction in variance; while this constraint does not force the other patches to look the same, they do because they are in fact the same surface.

tal data shown below suggests the sampler mixes well. The sampler converges if started from a random sample from the prior, but this is slow and unnecessarily inefficient. A good guess at edge positions follows by choosing a set of edges at maxima of the edge proposal distributions, censored to ensure the hardcore model applies. Similarly, a start point for the light position follows by choosing the maximum likelihood position from the proposal distribution; once the specular position is known, an estimate of illuminant colour follows. Finally, for each patch we obtain a reflectance estimate from the average colour within the patch and the illuminant colour. This yields a start point from which the sampler converges relatively quickly.

3.1 Resampling

Assume that we have a sampled representation of the posterior for two distinct images. We are now told that a patch in one image is the same as a patch in another — this should have an impact on our interpretation of both images. The sampled representation is well suited to determining the effect of this information.

In particular, we have samples of $P(\underline{\sigma}_a, \text{state a}|\text{image a})$ and $P(\underline{\sigma}_b, \text{state b}|\text{image b})$, where we have suppressed the details of the rest of the state in the notation. We interpret "the same" to mean that each patch is a sample from a Gaussian distribution with some unknown mean $\underline{\alpha}$ and a known standard deviation. We would like to obtain samples of $P(\underline{\alpha}, \text{state a}, \text{state b}|\text{image a}, \text{image b})$. Now we have that $P(\text{image a}, \text{image b}|\text{state a}, \text{state b}, \underline{\alpha})$ is proportional to

$$\int \begin{pmatrix} P(\text{image a}, \text{state a}|\underline{\sigma}_a)P(\underline{\sigma}_a|\underline{\alpha})\times \\ P(\text{image b}, \text{state b}|\underline{\sigma}_b)P(\underline{\sigma}_b|\underline{\alpha}) \end{pmatrix} d\sigma_a d\sigma_b \pi(\alpha)$$

Now the term inside the integral is:

$$\frac{P(\text{state a}, \underline{\sigma}_a, \text{image a})}{\pi(\underline{\sigma}_a)} \frac{P(\text{state b}, \underline{\sigma}_b, \text{image b})}{\pi(\underline{\sigma}_b)} P(\underline{\sigma}_b|\underline{\alpha})P(\underline{\sigma}_a|\underline{\alpha})$$

We have two sets of samples, Σ^a and Σ^b. We ensure that these samples are independent and identically distributed by shuffling them (to remove the correlations introduced by MCMC). This means that, for the conditional density for the i'th sample, we have $P(\Sigma_i^a|i) = P(\text{state a}, \underline{\sigma}_a, \text{image a})$. Now we construct a new sampler, whose state is $\{i, j, \alpha\}$. We ensure this produces samples of the distribution

$$\Pi(i, j, \alpha) = \frac{P(\underline{\sigma}_a(i)|\underline{\alpha})P(\underline{\sigma}_a(j)|\underline{\alpha})\pi(\underline{\alpha})}{\pi(\underline{\sigma}_a(i))\pi(\underline{\sigma}_b(j))}$$

We now use the i's and j's as indexes to our previous set of samples. We can marginalise with respect

to $\underline{\sigma_a}$ and $\underline{\sigma_b}$ by simply dropping their values from the sample. The result is a set of samples distributed according to the desired distribution. Building a sampler that obtains samples of $\{i, j, \alpha\}$ space according to the desired distribution involves technical difficulties beyond the scope of this paper.

4 Experimental results

The data set shown in figure 2 consists of images originally used in [4]. These images of the same set of patches on a Mondriaan of coloured paper patches, photographed under white, blue, yellow, purple, red and cyan light. There are no specularities, so we used a diffuse model for this data set.

The original data has been lost, so we used versions scanned from the paper; these images were displayed on a CRT, photographed from that display, subjected to four-colour printing and then scanned; it is remarkable that any constancy is possible under the circumstances. A basis was obtained using the bilinear fitting procedure of [11]. Constraint regions are more difficult; we obtained a natural coordinate system using principal components, and then constructed a bounding box in this coordinate system. The box was grown 10 % along each axis, on the understanding that none of the colours in the Mondriaans of [4] were very deeply saturated.

Figures 2 and 4 show a set of typical results from these images. The spatial model identifies edges correctly. Groups of samples drawn for the same surface reflectance under different lights intersect, as we expect. Furthermore, groups of samples drawn for different surface reflectances under the same light tend not to intersect, meaning that these surfaces are generally seen as different. Figure 6 shows a rendering of samples under white light, to give some impression of the variation in descriptions that results. The sampler is relatively slow (about one hour for 1000 samples on a Macintosh 300Mhz G3 in compiled Matlab).

The resampling results are pleasing. Figure 5 shows results obtained by assuming that a single surface patch in each of the six images is the same. This results in very much reduced variance in the rendering of that patch because the error balls for this surface patch intersect in a relatively small region. This reduced uncertainty about the reflectance of the patch means that our uncertainty about the illuminant in each case is also reduced. This means that *for a given illuminant* the variance in reports of the colour of every patch must be strongly reduced. What demonstrates the strength of the model is the fact that *from illuminant to illuminant* the variance in reports of patch colour is also reduced; thus, the algorithm has been able to use information that one patch is the same in each image to obtain a representation that strongly suggests the other patches are the same, too (figure 6).

5 Discussion

Because MCMC algorithms produce information about the range of possibilities implied by an image, integrating other forms of information is simpler. These algorithms are an attractive tool for integrating current vision techniques, because the proposal mechanism makes it possible to use distinct, apparently incompatible algorithms and obtain a unified representation; furthemore, standard vision techniques — like looking at gradient magnitude to find edges — have a natural role as proposal processes. Resampling means that new information can be seamlessly incorporated into a representation, given a rather natural conditional independence property applies.

Acknowledgements: Thanks to Stuart Russell for pointing out the significance of MCMC as an inference technique.

References

[1] D.H. Brainard and W.T. Freeman. Bayesian colour constancy. *J. Opt. Soc. Am.-A*, 14:1393–1411, 1997.

[2] G. Buchsbaum. A spatial processor model for object colour perception. *J. Franklin Inst.*, 310:1–26, 1980.

[3] G. Finlayson. Colour in perspective. *IEEE T. Pattern Analysis and Machine Intelligence*, 18:1034–1038, 1996.

[4] D.A. Forsyth. A novel algorithm for colour constancy. *Int. J. Computer Vision*, 5:5–36, 1990.

[5] B.V. Funt, K. Barnard, and L. Martin. Is machine colour constancy good enough? In *ECCV*, pages 445–459, 1998.

[6] P.J. Green. Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, 82(4):711–732, 1995.

[7] P.J. Green. Mcmc in image analysis. In W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, editors, *Markov chain Monte Carlo in practice*, pages 381–400. Chapman and Hall, 1996.

[8] E.H. Land and J.J. McCann. Lightness and retinex theory. *J. Opt. Soc. Am.*, 61(1):1–11, 1971.

[9] H.C. Lee. Method for computing the scene-illuminant chromaticity from specular highlights. *J. Opt. Soc. Am.-A*, 3:1694–1699, 1986.

[10] L.T. Maloney and B.A. Wandell. A computational model of color constancy. *J. Opt. Soc. Am*, 1:29–33, 1986.

[11] D.H. Marimont and B.A. Wandell. Linear models of surface and illuminant spectra. *J. Opt. Soc. Am.-A*, 9:1905–1913, 1992.

[12] R.M. Neal. Probabilistic inference using markov chain monte carlo methods. Computer science tech report crg-tr-93-1, University of Toronto, 1993.