Probabilistic Relaxation as an Optimizer

A. J. Stoddart, M. Petrou and J. Kittler
Department of Electrical and Electronic Engineering,
University of Surrey, Guildford, GU2 5XH, U.K.
a.stoddart@ee.surrey.ac.uk

Abstract

Probabilistic Relaxation with product support has been shown to have advantages over 'traditional' probabilistic relaxation. However it is less well understood in the sense that a cost function is not known. In this paper we present a cost function. This greatly improves our understanding of probabilistic relaxation with product support, and also leads us to propose a new class of probabilistic relaxation algorithms. We investigate two applications.

1 Introduction

Probabilistic relaxation [PR] is an established technique for pattern recognition. For a review of PR and applications of PR see [1]. The 'traditional' scheme for PR, which we will call HZ PR, was reviewed in a landmark paper by Hummel and Zucker [2] and is well understood in the sense that a cost function is known.

Kittler and Hancock [3] developed a version of PR in which the support is of product form. This version of PR, which we call KH PR, has significant advantages. It is based on an 'object centered' formula, in which evidence from neighboring objects is combined according to the axioms of probability theory, in order to decide the label of each object. The compatibility coefficients, instead of being heuristically chosen, are prescribed to be conditional probabilities. This support was shown to be unbiased and correct in certain limits. More recently Christmas, Kittler and Petrou [CKP] have developed a related PR algorithm [4] also of product form.

Although KH PR is superior to HZ PR in some respects, a cost function for KH PR has up to now not been known. A cost function can be useful in several ways. In particular it provides an enhanced theoretical understanding of what the algorithm achieves. In this paper we present cost functions for the KH and CKP PR algorithms. We are able to prove that unambiguous stable stationary points of the update algorithm are local maxima of this cost function. The detailed proofs will be presented elsewhere [8]. As a by-product we introduce a new class of relaxation algorithms that hold considerable promise. Work is in progress on applications, and we present two example applications in this paper.

The labelling problem has been studied in an optimization framework by others. Hummel and Zucker [2] study PR with sum rule support, however the final proof connecting the cost to the update rule had to wait until Pelillo [5]. Some early work by Faugeras and Berthod [6] also made optimization a key ingredient, however the cost function was a heuristic choice. Our cost function is firmly linked to the posterior probability, and as such we know precisely what we are computing. In a separate paper we have studied the connection between PR and the Baum Eagon theorem [7].

BMVC 1995 doi:10.5244/C.9.61

2 Notation

In this paper we consider labelling problems. We wish to match scene properties to those of a model. The scene is comprised of nodes or objects, where each node is denoted by i, j, or $k = 1..N$, for N objects. The M model nodes or labels are denoted by $\alpha = 1..M$. When object i has label α this is denoted by α_i. We denote a joint labelling of all objects by the set $\{\alpha_1, \alpha_2, \dots \alpha_N\}$. As a shorthand we define the multiindex $\vec{\alpha} = \{\alpha_1, \alpha_2, \dots \alpha_N\}$.

The probability of object i having label α is denoted $p(\alpha_i)$, and the probability of some global labelling $\vec{\alpha}$ is denoted $p(\vec{\alpha})$. The *prior* probability will be denoted by $\hat{p}(\vec{\alpha})$ to make it distinct from probabilities conditioned on measurements, or posterior probabilities.

In general, a probability of the form $p(\vec{\alpha})$ is an extremely complex quantity. It may take on a distinct positive real value for each of the M^N distinct global labellings. If we can establish or assume that the objects are conditionally independent then the probability may be factorized. An example is the prior probability for CKP PR, which is assumed to factorize as follows

$$\hat{p}(\vec{\alpha}) = \prod_{i=1}^{N} \hat{p}(\alpha_i) \tag{1}$$

In this case there are only MN distinct positive real values, subject to N constraints, namely that $\sum_{\alpha_i} \hat{p}(\alpha_i) = 1$ for each object.

The quantities that we use in update rules may or may not have an *interpretation* as probabilities. We will call these quantities *weighted labelings*. They will always factorize as in equation (1). We always denote these quantities with the letter r, to distinguish them from probabilities usually denoted p.

3 A review of KH & CKP PR

In this section we briefly introduce the Probabilistic Relaxation [PR] scheme of Kittler and Hancock [KH PR], and the later scheme of Christmas, Kittler and Petrou [CKP PR].

The KH PR algorithm applies to labeling problems where there is a physical adjacency or neighborhood relationship between objects, e.g. the objects may be pixels on a regular grid. For each scene node j we may define a neighborhood I_j which includes node j. If we exclude node j from I_j we denote this $I_j^* = I_j \setminus \{j\}$.

We assume that if $i \in I_j$ then $j \in I_i$. The set of all pairs i, j such that $i \in I_j$ is denoted $\mathcal{P}$. This could be for example the set of all 4-connected pairs of pixels in an image.

The world model is captured in the global prior probability, $\hat{p}(\vec{\alpha})$. The KH PR algorithm addresses problems in which the prior probability is of Markov type, in other words that

$$\hat{p}(\alpha_j | \vec{\alpha} \setminus \alpha_j) = \hat{p}(\alpha_j | \alpha_i, i \in I_j^*) \tag{2}$$

Kittler and Hancock went on to consider a variety of problems based on this.

Probabilities which are Markov random fields with respect to some neighborhood are subject to the Clifford-Hammersley theorem which states that they have an equivalence with a Gibbs distribution over that neighborhood.

In this paper we consider only a restricted subset of Markov-type models in which the prior probability may be expressed as a Gibbs distribution

$$\hat{p}(\vec{\alpha}) = \frac{1}{Z} \exp \left\{ - \sum_{\{i,j\} \in \mathcal{P}} G(\alpha_i, \alpha_j) \right\} \tag{3}$$

where the functions $G(\alpha_i, \alpha_j)$ may be chosen arbitrarily, while Z normalizes the sum over label space to one. One consequence of this assumption is that the probability factorizes into a product over pairs $\{i,j\} \in \mathcal{P}$. Another consequence is that the conditional prior probability factorizes,

$$\hat{p}(\alpha_j, \alpha_i \in I_j^*) = \hat{p}(\alpha_j) \prod_{i \in I_j^*} \hat{p}(\alpha_i | \alpha_j) \tag{4}$$

The KH PR algorithm considers that we make measurements of unary properties. We measure some property x_i, which depends on the label α_i, and presume that we can model the sensor using a probability density function $p(x_i|\alpha_i)$. Each unary measurement is assumed to be conditionally independent of measurements taken at other nodes.

We start off by considering the probability of label α_j given only the measurements in its immediate contextual neighbourhood. $p(\alpha_j|x_i, i \in I_j)$. The PR algorithm starts by deriving from the standard axioms of probability theory the rule that

$$p(\alpha_i|x_j, j \in I_i) = \frac{p(\alpha_i|x_i) \prod_{j \in I_i^*} \sum_{\alpha_j} \frac{\hat{p}(\alpha_i,\alpha_j)}{\hat{p}(\alpha_i)\hat{p}(\alpha_j)} p(\alpha_j|x_j)}{\sum_{\beta_i} p(\beta_i|x_i) \prod_{j \in I_i^*} \sum_{\alpha_j} \frac{\hat{p}(\beta_i,\alpha_j)}{\hat{p}(\beta_i)\hat{p}(\alpha_j)} p(\alpha_j|x_j)} \tag{5}$$

A derivation of this expression may be found in [3]. The above formula contains no heuristic elements, it is simply an application of Bayes rule and the axioms of probability theory.

To obtain a PR algorithm this formula is iterated. This adds a heuristic element, where we make the identification that $r^{(n)}(\alpha_j) \leftarrow p(\alpha_j|x_j)$, and use equation (5) as an update rule. In the notation of the standard update rule the support is given by

$$q^{(n)}(\alpha_i) = \prod_{j \in I_i^*} \sum_{\alpha_j} [\hat{p}(\alpha_i, \alpha_j) / \hat{p}(\alpha_i)\hat{p}(\alpha_j)] r^{(n)}(\alpha_j) \tag{6}$$

and the algorithm is initialized with $r^{(0)}(\alpha_i) = p(\alpha_i|x_i)$.

3.1 CKP PR

The CKP PR algorithm tackles a different class of problems. Here no adjacency relations are assumed. It is therefore a natural assumption that the prior probability factorizes as in equation (1).

The labelling process is based on a set of N conditionally independent unary measurements x_i for each object, and a further set of $N(N-1)/2$ conditionally independent binary measurements A_{ij} which are independent of the unary ones.

The neighborhood I_i is now the set of all objects $I_i = \{1..N\}$, and I_i^* and $\mathcal{P}$ are defined accordingly.

$$p(\alpha_i|x_j, j \in N_0, A_{ij}, j \in N_i) = \frac{p(\alpha_i|x_i) \prod_{j \neq i} \sum_{\alpha_j} p(A_{ij}|\alpha_i, \alpha_j) p(\alpha_j|x_j)}{\sum_{\beta_i} p(\beta_i|x_i) \prod_{j \neq i} \sum_{\alpha_j} p(A_{ij}|\beta_i, \alpha_j) p(\alpha_j|x_j)} \tag{7}$$

In the same way as KH PR the equality above is iterated to produce a probabilistic relaxation algorithm. The support now has the measurement conditional probabilities playing the role of compatibility coefficients and is given by

$$q^{(n)}(\alpha_i) = \prod_{j \neq i} \sum_{\alpha_j} p(A_{ij}|\alpha_i, \alpha_j) r^{(n)}(\alpha_j) \tag{8}$$

In structural terms the support functions for KH and CKP PR are very similar, although the compatibility coefficients relate to different quantities.

4 Understanding KH & CKP PR

In this paper we suggest a cost function for KH & CKP PR. Theorems in support of this cost function will be presented elsewhere [8]. The underlying mathematical structure will turn out to be similar in both cases, so that we will consider only one relaxation formula, namely an update formula with support of the form

$$q(\alpha_i) = D(\alpha_i) \prod_{j \in I_i^*} \sum_{\alpha_j} C(\alpha_i, \alpha_j) D(\alpha_j) r(\alpha_j) \tag{9}$$

The coefficients $D(\alpha_i)$ and $C(\alpha_i, \alpha_j)$ will be shown to relate to appropriate terms in KH and CKP PR. The cost function we propose is of the general form

$$E[\underline{r}] = \sum_{\vec{\alpha}} \left[\prod_i r(\alpha_i) \right] \left[\prod_i D(\alpha_i) \right] \left[\prod_{\{i,j\} \in \mathcal{P}} C(\alpha_i, \alpha_j) \right] \tag{10}$$

How do we motivate this cost? Essentially it is a guess backed up by some proofs. It is not altogether unexpected in form. The combined product over $D(\alpha_i)$ and $C(\alpha_i, \alpha_j)$ will later be seen to map on to the global posterior probability. The product over $r(\alpha_i)$ may seem a little unexpected. The $r(\alpha_i)$ are auxiliary variables whose primary role is to convert a discrete maximization problem over label space ($\vec{\alpha}$) to a maximization over a continuous weighting space.

We may state what we have proved [8] as follows. If the relaxation formula converges to an unambiguous labelling, then that labelling is a local maximum of the above cost function. The update rule does sometimes converge to ambiguous labellings. Some additional statements can be made in this case, these relate mainly to degenerate solutions.

We should make clear that we have not found a general proof that $E[\underline{r}^{(n+1)}] \geq E[\underline{r}^{(n)}]$, although for practical purposes we regard this to be the case.

We now consider the cost functions that the KH and CKP PR maximize. The KH algorithm, with support given by equation (6), finds a local maximum of the cost function

$$E[\underline{r}] = \sum_{\vec{\alpha}} r(\vec{\alpha}) \left[\prod_{\{i,j\} \in \mathcal{P}} \frac{\hat{p}(\alpha_i, \alpha_j)}{\hat{p}(\alpha_i)\hat{p}(\alpha_j)} \right] \tag{11}$$

An obvious cost function that we might attempt to maximize is a cost with coefficients given by the posterior probabilities of labellings

$$E[\underline{r}] = \sum_{\vec{\alpha}} r(\vec{\alpha}) \left[\prod_i p(x_i|\alpha_i) \right] \left[\prod_{\{i,j\} \in \mathcal{P}} \hat{p}(\alpha_i, \alpha_j) \right] \tag{12}$$

The global maximum of this cost function is the MAP estimate, i.e. the labelling that maximizes

$$p(\vec{\alpha}|x_i, i = 1..N) \propto p(x_i, i = 1..N|\vec{\alpha})\hat{p}(\vec{\alpha}) \tag{13}$$

This would lead to a modified update rule with support given by

$$q^{(n)}(\alpha_i) = p(x_i|\alpha_i) \prod_{j \in I_i^*} \sum_{\alpha_j} \hat{p}(\alpha_i, \alpha_j) p(\alpha_j|x_j) r^{(n)}(\alpha_j) \tag{14}$$

We will refer to this as SKH PR. How do we propose to initialize this new PR algorithm? The obvious choice is to set $r^{(0)} = 1/M$, and the immediate consequence of this it that the first iteration then coincides with KH PR, i.e.

$$r^{(1)}(\alpha_i) = p(\alpha_i|x_j, j \in I_i) \tag{15}$$

After this iteration the new algorithm balances the competing requirements of the data and the prior model to find the nearest local maximum of the cost as given in equation (12). In contrast succesive iterations of KH PR are dominated by the prior model and find the nearest local maximum of equation (11), given the starting point.

The same procedure may be applied to CKP PR and we see that CKP PR finds a local maximum of the cost

$$E[\underline{r}] = \sum_{\vec{\alpha}} r(\vec{\alpha}) \left[\prod_{\{i,j\} \in \mathcal{P}} p(A_{ij}|\alpha_i, \alpha_j) \right] \tag{16}$$

Once again it seems attractive to maximize a cost function with coefficients given by the posterior probabilities of labellings

$$E[\underline{r}] = \sum_{\vec{\alpha}} r(\vec{\alpha}) \left[\prod_i p(\alpha_i|x_i) \right] \left[\prod_{\{i,j\} \in \mathcal{P}} p(A_{ij}|\alpha_i, \alpha_j) \right] \tag{17}$$

The global maximum of this cost function is once again the MAP labelling, and this leads to a modified update rule [SCKP PR] with support given by

$$q^{(n)}(\alpha_i) = p(\alpha_i|x_i) \prod_{j \in I_i^*} \sum_{\alpha_j} p(A_{ij}|\alpha_i, \alpha_j) p(\alpha_j|x_j) r^{(n)}(\alpha_j) \tag{18}$$

5 Experiments

The first problem that we will consider is an image restoration problem, such as might be encountered in the context of interpreting satellite imagery. This will enable us to compare the performance of KH and SKH PR.

We consider a scene with each imaged patch (pixel) i having some label α_i. The image pixel contains some (unary) measurement x_i which we use to infer the label α_i. In our experiment x_i is taken to be a grey level between 0 and 255, and there are five possible labels $\alpha_i = 0..4$. In a remote sensing problem x_i could denote some multispectral measurements, and α_i could denote terrain type, e.g. forest, desert, city.

In our synthetic model we use a conditional probability of grey level given pixel label that is a sum of a Gaussian probability distribution and a 10% uniform dropout,

$$p(x_i|\alpha_i) = \frac{0.9}{\sqrt{2}\sigma} \exp\left\{-\frac{(x_i - (a\alpha_i))^2}{2\sigma^2}\right\} + 0.1\Theta(x_i - 254) \tag{19}$$

$\Theta(x)$ is the Heaviside (step) function, and we choose the parameters a and σ so that the Gaussians of labels differing by one overlap considerably ($a = 2\sigma$). This means that $p(\alpha_i|x_i)$ misclassifies $0.9 \times 32\% + 10\%$ of the data.

We choose a prior model completely specified by

$$\hat{p}(\alpha_i, \alpha_j, j \in I_i) = \begin{cases} 3k & \alpha_i = \alpha_j \\ k & \alpha_i \neq \alpha_j \end{cases} \tag{20}$$

The prior probability has no other structure. k is chosen to normalize the probabilities to 1, but we never actually need to know its value, so it may be ignored.

In this model the binary prior probability exerts pressure on neighboring pixels to be similar and may be used to correct for misclassifications due to the unaries. Both binary and unary terms compete, and a good solution is one which achieves a balance between their competing requirements.

We can construct a typical representative of the scene population by using a Gibbs sampler and the result is shown in figure 1. The labels of the scene are represented by grey levels. Next to an image depicting the ground truth or original image, we display an

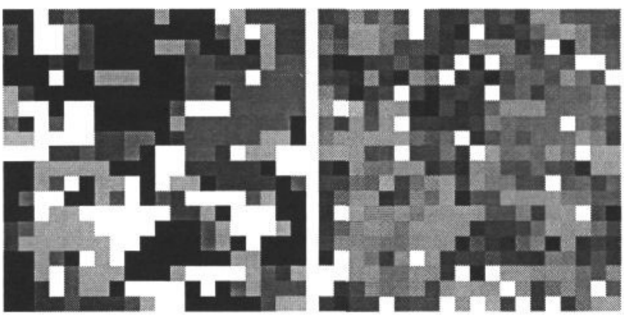


Figure 1: The "ground truth" scene, and its image

image of the measurements x_i. The white pixels in the right image are the 10% dropout.

Before discussing the results we discuss how to evaluate performance. In equation 10 we presented the relaxation cost function. For an unambiguous labelling this simplifies to the posterior probability. We denote the log posterior probability by C, which we separate into two terms, C_u the unary probability, and the binary probability C_b. We take the natural logs of the probabilities, and ignore common multiplicative factors. This means that only differences in C are significant.

$$C_u = \log\{p(x_i, i = 0..N|\vec{\alpha})\} \tag{21}$$

$$C_b = \log\{\hat{p}(\vec{\alpha})\} \tag{22}$$

$$C = C_u + C_b \tag{23}$$

We can now present results for KH and SKH relaxation in table 1 and figure 2. For reference purposes the score for the original image is also presented. A better score is larger (more likely) than a worse score. Thus we see that SKH PR gives a better result than the original image, as measured by its posterior probability C. The unary probability is much improved, and the binary term a little degraded to produce overall a good solution. SKH PR converges in between 20 and 100 iterations.

The KH method was terminated without convergence after 100 iterations, giving a very poor score. This is not surprising. After a very large number of iterations the KH method will probably converge to a single label for the whole image, which is its global maximum. It is dominated by the binary probabilities since only these occur in the update rule. Thus the C_b score is very good, but the C_u score poor. Overall it performs badly as shown by C. The resulting labellings are shown in figure 2. It is easy to see that the KH algorithm is dominated by the binary probabilities or priors. We also note that the 39% misclassification rate of the unaries alone is reduced to 19.5% by SKH and 30% by KH PR.

Table 1: Unary and binary probabilities

Method	C_u	C_b	C
original image	-1473.81	421.867	-1051.94
SKH	-1397.81	408.684	-989.13
KH, 5 iterations	-1444.27	476.798	-967.474
KH	-1922.57	606.434	-1316.14

All this is very much as would be expected based on the cost functions identified earlier in this paper. There was however one interesting surprise, concerning non-equilibrium states. If one terminates the KH algorithm after 5 iterations one obtains a very good result. However, left to converge, the results are always poor. Our cost function analysis can tell us what to expect the method to converge to, but we do not have any framework in which to study non-equilibrium performance. Naturally the cost function can tell us when to terminate the KH algorithm at a good point.

It should be made clear that KH PR has also been used in applications where the priors have more detailed structure and many local maxima. The prior for the problem studied in this section has a dominant global maximum.

5.1 CAG Matching

The second application that we will consider has both unary and binary measurements. This is a suitable application for SCKP PR.

Recently Matas, Marik and Kittler [9] have developed a new scheme for color based recognition. The scheme is based on a representation called the color adjacency graph, or CAG. A clustering algorithm in chromaticity space is applied to the image, resulting in a segmentation into sets of regions of approximately uniform chromaticity. These regions produce the nodes of the CAG. Edges are inserted between nodes (region sets) that share common boundaries. The attribute of each node is its chromaticity histogram, and the edge attribute is a histogram of the intensity ratio along the common boundary.

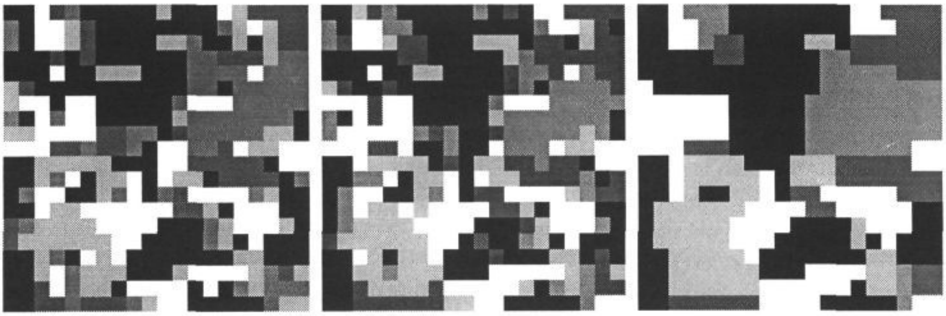
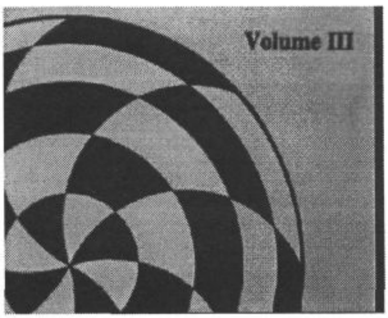


Figure 2: The original scene (left), and the SKH (center) and KH results (right).

The edge histograms are approximately invariant to orientation or object deformation. In the experiments described in [9] a constant illumination is used and the chromaticity histogram is therefore also approximately invariant to orientation or object deformation. For further details consult [9].

(a) 

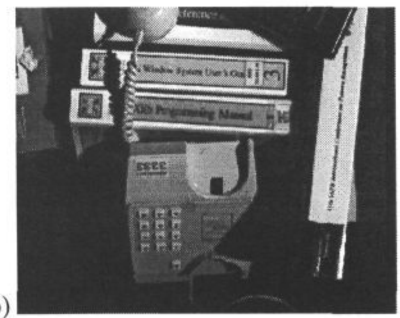
(b) 

Figure 3: The (a)model and (b) scene images

In this application the models are acquired by processing images of the object of interest, without background, in the same way as scene images are processed. In order to apply PR we must estimate the conditional probabilities of the measurements on the labels, in other words we need $p(x_i|\alpha_i)$ and $p(A_{ij}|\alpha_i, \alpha_j)$. Ideally these conditional probabilities should be based on comparison of the full scene and model histograms. So far we have adopted the expedient of estimating a covariance from the largest mode in the model histogram, and then modelled the pdf by a Gaussian. This works well in cases where the distribution is unimodal.

In figure 3(a) we show a typical model image. This is a volume of ICPR proceedings in a yellow and blue cover. In figure 3(b) we show a scene consisting of several reference manuals. The CAG for the scene is shown in figure 4.

Thus far we have tested SCKP and CKP PR on several simple problems from the Matas, Marik dataset, and both algorithms correctly identify the model in all images. The reason for this is that the unary and binary measurements give quite high discriminative power, and both algorithms usually converge in one or two iterations. The dataset is too simple to discriminate between SCKP and CKP PR.

Datasets with more ambiguity are being collected and will be tested. We expect the

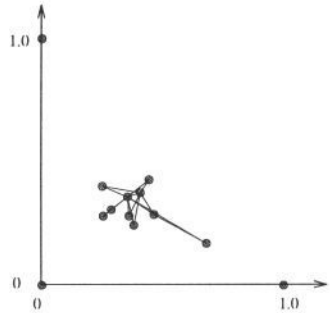


Figure 4: A CAG with nodes plotted in chromaticity space

SCKP to be better than CKP when they differ. This is because the unaries (chromaticities) are very important and we expect a binary dominated algorithm to be unsatisfactory.

Finally we make a point about computational complexity. If support is sought from all objects, i.e. the product in equation 18 is over all objects, and the sum over all labels, then the complexity of the algorithm is N^2M^2 per iteration. This means the algorithm is quadratic in number of models.

However in the implementation of the above problem we have a graph which specifies which adjacent objects are considered to support any given object. Each model is a graph, and disconnected nodes in the model graph have zero compatibility coefficient. They can therefore be ignored in the sum over labels in equation 18. If there are N_e edges in the scene graph and M_e edges in the model graph the complexity per iteration becomes approximately $NM + N_eM_e$ which is linear in the number of models.

6 Conclusion

We have presented a cost function for Probabilistic Relaxation with product support. It applies to both the algorithm presented by Kittler and Hancock, and also the algorithm presented by Christmas, Kittler and Petrou. All unambiguous labellings that are stable stationary points of the update formula are local maxima of the cost function.

Motivated by the cost function, we have proposed a new class of Probabilistic Relaxation algorithms. They appear to have considerable promise, and work is in progress to explore these new algorithms. The area of applicability of the new algorithms is extremely wide, since many recognition and reconstruction problems in vision are conveniently formulated as MAP problems.

The two applications studied in this paper have relatively little structure in their priors. It should be noted that part of the appeal of KH PR is its ability to deal with problems that have significant structure expressed in their priors. The relative performance of KH and SKH PR on such problems remains an open question.

The image restoration problem demonstrated clearly that the asymptotic performance of SKH PR is better than KH PR. A surprising result was that non-equilibrium perfomance of KH PR is best. The cost function gives us a basis on which to terminate KH PR, but we have not gained any theoretical insight into why the non-equilibrium perfomance is good.

The CAG problem as studied so far is too simple to discriminate between CKP and

SCKP PR. It is nonetheless included since it is an ideal application for both algorithms. All the assumptions made in the formulation of the posterior probability are satisfied, which is not always the case for other matching problems. In addition, because the unary and binary measurements are both important, it will be a good test bed for matching methods.

7 Acknowledgements

We would like to acknowledge G. Matas and R. Marik for their assistance with preparation of the colour matching experiment.

References

[1] J. Kittler, J. Illingworth, Relaxation labelling algorithms-a review, Image and Vision Computing 3(4), 206-216, (1985).

[2] R. A. Hummel, S. W. Zucker, On the foundations of Relaxation Labeling Processes, IEEE Pattern Analysis and Machine Intelligence 5(3), 267-287, (1983).

[3] J. Kittler, E. R. Hancock, Combining evidence in probabilistic relaxtion, Int. J. of Pattern Recognition and Artificial Intelligence 3(1), 25-51, (1989).

[4] W. J. Christmas, J. Kittler, M. Petrou, Matching in computer vision using probabilistic relaxation, *to appear in* IEEE Pattern Analysis and Machine Intelligence.
see also W. J. Christmas, J. Kittler, M. Petrou, Matching of road segments using probabilistic relaxation: a hierarchical approach, SPIE 2304, San Diego, California, USA, (1994).

[5] M. Pelillo, On the dynamics of relaxation labelling processes, IEEE Int. Conf. Neural Networks, Orlando, (1994).

[6] O.D. Faugeras, M. Berthod, Improving consistency and reducing ambiguity in stochastic labeling: An optimization approach, IEEE Pattern Analysis and Machine Intelligence 3(4), 412-424, (1981).

[7] A. J. Stoddart, M. Petrou and J. Kittler, A new algorithm for Probabilistic Relaxation based on the Baum Eagon theorem, to appear in Computer Analysis of Images and Patterns, Prague, (1995).

[8] A. J. Stoddart, M. Petrou, J. Kittler, Probabilistic Relaxation as an Optimizer, VSSP-TR-3/95.

[9] J. Matas, R. Marik, J. Kittler, On representation and matching of multi-colored objects, pp 726-732, 5th Int. Conf. Computer Vision, Boston, June (1995).