Relating Gibbs Distributions to Empirically Derived Marginal Distributions for Image Analysis

D. B. Sher
Mathematics Dept
Nassau Community College
Garden City NY 11530

C. E. Wafford and D. Milun
Computer Science Dept.
SUNY Buffalo
Buffalo NY 14260

{sher,wafford,milun}@cs.buffalo.edu

Abstract

The log marginal probability ratios of a Gibbs distribution over an 8 connected neighborhood system is a linear function of its 66 clique weights. We empirically determine log marginal probability ratios of artificial noiseless images and use the pseudoinverse method to compute the closest Gibbs distribution. We compare these Gibbs distributions to *ad hoc* distributions suggested in the literature and to the empirical marginals from which they are derived.

1 Introduction

One of the popular methods for noise removal, image labeling and segmentation is finding the maximum *a posteriori* (MAP) interpretation of the image using a model with zero mean independently distributed noise and a Gibbs prior. Gibbs distributions have convenient theoretical properties: they are compact — of lower dimensionality than marginal distributions — and important optimization methods such as the iterated conditional modes (ICM) algorithm and simulated annealing are guaranteed to converge [Bes86, GG84].

This paper examines how closely Gibbs distributions model the local structures that commonly occur in a typical class of artificial imagery. Our experiments use empirical distributions over 3×3 neighborhoods from images of occluding ellipses and rotated rectangles as in figure 1. We model these marginal distributions with Gibbs distributions, experimentally determine how well the Gibbs distributions models the empirical data, and test the Gibbs model against the empirical distribution using the ICM and simulated annealing algorithms.

In D. and S. Geman's 1984 paper [GG84] the Gibbs distribution is defined as follows:

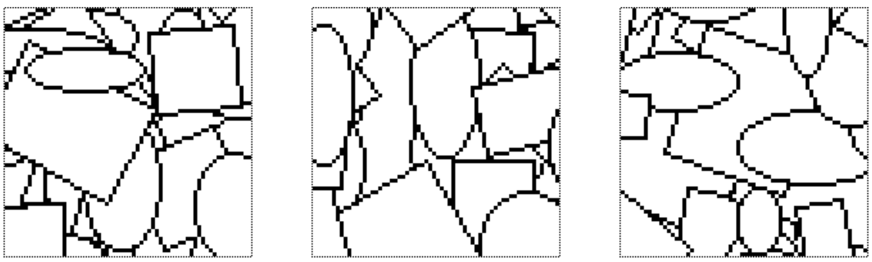

Figure 1: Some example images

A *Gibbs distribution* relative to $\{S, G\}$ is a probability measure π on Ω with the following representation:

$$\pi(\omega) = \frac{1}{Z} e^{-U(\omega)/T}$$

where Z and T are constants and U, called the *energy function,* is of the form

$$U(\omega) = \sum_{c \in C} V_c(\omega)$$

Recall that C denotes the set of cliques for G. Each V_c is a function on Ω with the property that $V_c(\omega)$ depends only on those coordinates x_s of ω for which $s \in c$.

The ICM or annealing algorithms use the probability ratio between labeling ω_{old} and ω_{new} where the labeling of ω_{old} differs from the labeling ω_{new} at the label of a single pixel, x. This ratio is:

$$\frac{\pi(\omega_{new})}{\pi(\omega_{old})} = e^{-(U(\omega_{new}) - U(\omega_{old}))/T}$$

The ratio depends linearly on the energy difference $U(\omega_{new}) - U(\omega_{old})$ which is computed from the formula below:

$$U(\omega_{new}) - U(\omega_{old}) = \sum_{c \in C_x} V_c(\omega_{new}) - V_c(\omega_{old})$$

where C_x is the set of cliques that contain x.

If we assume that spatially translating a clique does not change its weight function, then, in the graph corresponding to pixels with 8-connected neighbors, there is 1 type of clique containing 1 pixel, 4 types of cliques containing 2 pixels, 4 types of cliques containing 3 pixels, and 1 type of clique containing 4 pixels. If a clique contains n pixels there are 2^n possible binary labelings the members of the clique can possess. Thus a MRF over 8-connected neighborhoods is described by $2 * 1 + 4 * 4 + 8 * 4 + 16 * 1 = 66$ possible values of V_c.

Given the labeling of the 8 neighbors of a pixel x, the log of the ratio between the probability that x is labeled 1 and the probability that x is labeled 0 is a linear function of the 66 numbers that describe the Gibbs distribution. Thus the log marginal ratios of a Gibbs distribution is a linear function of its clique weights. Given a set of 256 empirically determined log marginal ratios the method of pseudoinverses determines the geometrically closest Gibbs distribution. This paper studies whether the Gibbs distributions derived in this manner are similar to those proposed in the literature and compares the effectiveness of the Gibbs priors to the empirical marginals from which they are derived.

2 Hancock Kittler Smoothing

One problem with empirical marginal distributions is that many local neighborhoods with small probabilities never occur in the sample. This causes many of the empirical marginals to be zero in the sampled distribution. However the probability ratios of such local neighborhoods may be necessary for image analysis using these distributions.

Hancock and Kittler[HK90] studied this problem for fields directly estimated by people (where this problem is even more acute) and suggested a solution to this problem. They add "noise" to an estimate distribution, by smoothing it. The marginal probability of a neighborhood labeling depends on how close it is, by the L_1 norm, to the high probability neighborhood labelings.

In previous work [MS91, MS93] we used a modification of this method of Hancock and Kittler, and smoothed empirical probability densities, and used these probability density functions as prior information at a simulated annealer, and in ICM algorithm. We found that using this Hancock-Kittler smoothing yielded reconstructions which were visually and statistically superior.

We investigated this property, and found that as a rough estimate, the theoretical noise should be 0.61 of the true noise level. In this paper we shall refer to this particular HK level as "optimum" HK.

3 Background

This approach of using local neighborhood characteristics is essentially that of Markov random fields (MRF), originally developed and applied to image domains by J. Besag and D. and S. Geman [Bes74, Bes86, GG84]. Many researchers have followed up on this research, and have developed various techniques for "solving" images when the local neighborhood characteristics are specified as an MRF, for example [JW90, GGGD90, TGD92, BT92, Mar85, MMP85, She91, GG91, Bla89, LD89, Swa90]. However, not much attention has been paid to the methods of deriving the probability density functions of the MRFs.

Gray, Kay and Titterington [GKT92] follows up on work by Derin and Elliott [DE87] on the estimation of MRF parameters using a method called the "logit" method, and a "histogram count", which is very similar to the sampling ideas of this paper.

Lakshmanan and Derin [LD89] and Manjunath and Chellappa [MC91] also try to estimate the MRF parameters from a single image, but they use either a simulated annealing or nearest neighbor approach to do so.

Kim and Yang [KY91] use the error backpropagation technique from neural networks to estimate the MRF parameters.

4 Derived Distributions

The Gibbs distributions suggested by various authors in the literature tend to have a very simply structure over the 66 clique potentials. In fact, many authors choose to only use the 4-connected neighborhood, and thus only cliques containing 1 or 2 pixels, and only have 10 cliques potential values.

Those that use the 8-connected neighborhood ([DE87, LD89] among others), usually use some relatively simple rule such as assigning one value β to a clique if all pixels within it are equal, and $-\beta$ otherwise. Even many people who use 4-connected neighborhoods, use a similar scheme, such as [ZZAA92].

The distributions found by the method of pseudoinverses, as described in Section 1, have only a small number of cliques with non-zero potentials. This makes our derived cliques a very compact method of storing local information. The cliques that our method generates are very different from most *ad hoc* ones. See figure 2 for the set of non-zero cliques generated in the examples discussed here.

5 Effectiveness Comparisons

One way of validating the generated Gibbs distributions is to compare them to the original marginal distribution. To do this we convert them back into marginals by generating a marginal distribution

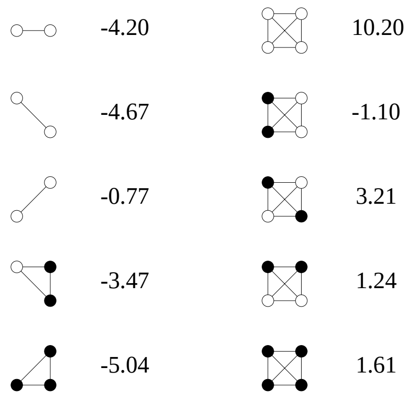

Figure 2: Cliques with non-zero potentials in our empirical Gibbs distribution.

HK level	# of switches	proportion
0.000	83	0.08229
0.024	84	0.08516
0.048	78	0.11346

Table 1: Similarity between original marginal distribution, and marginal distribution derived from generated Gibbs distribution.

defined by the Gibbs distribution. We then compare the 2 marginal distributions. Our method of comparison was modeled on the way these distributions are used: we subtract the marginal probabilities of pairs of neighborhoods (with the center set vs. center clear), in each of the two marginal distributions, and count how many of the 256 pairs change sign. This is a useful metric because this is the piece of information that the ICM algorithm effectively uses. Table 1 shows that the number of switches is not particularly high (about 80 of 256); and that the proportion of the total number of cliques that this effects is very low.

To test the actual effectiveness of the Gibbs distribution, we ran an ICM algorithm using both the empirical marginal distribution and the generated Gibbs distribution. This was done on images with 4% and 8% independent random noise added. We also ran a simulated annealer algorithm using the generated Gibbs distribution. We have found that the empirical marginals tend to do somewhat better at reconstructing the images.

We used two metrics to measure the effectiveness of the reconstructions. One is the metric of the number of bits that differ between the reconstructed image and the original noise-free image. The results from this metric are included in tables 2 and 3. The second measure we used is Pratt's Figure of Merit [Pra78]. This is a number in range 0 to 1, which represents how good an edge detector is, with 1 being perfect. It takes into account both edge position and accuracy. The results using this metric were effectively identical, and have not been included here.

		ICM		Annealer	
		Mean	Std dev	Mean	Std dev.
	HK=.000	124.2	34.8	384.0	47.4
Marginals	HK=.024	70.5	14.2	86.2	13.9
	HK=.048	168.3	10.4	139.7	10.2
	HK=.000	394.1	43.8	394.9	42.3
Gibbs	HK=.024	168.3	10.4	138.1	9.7
	HK=.048	168.3	10.4	157.8	10.4

Table 2: Number of bits incorrect when 4% noise was used.

		ICM		Annealer	
		Mean	Std dev	Mean	Std dev.
	HK=.000	217.7	33.4	392.2	43.1
Marginals	HK=.024	261.7	34.0	370.4	36.1
	HK=.048	158.4	14.6	191.7	14.2
	HK=.000	396.8	42.3	394.9	42.3
Gibbs	HK=.024	219.2	17.6	214.7	15.8
	HK=.048	219.2	17.6	247.9	16.1

Table 3: Number of bits incorrect when 8% noise was used.

6 Summary

Best results are still obtained by using the empirical marginal distribution with the optimal HK level. However, because of the compact nature of Gibbs distributions, if space is at a premium, then it is worth using HK and then applying our method to find the closest Gibbs distribution and using that. All our Gibbs distributions yield better results than marginal distributions without HK smoothing.

All our empirical Gibbs distributions (regardless of HK level) yield the same non-zero cliques. We feel that future research will clarify the reason that these particular cliques are non-zero, when describing boundary imagery.

References

[Bes74] J.E. Besag. Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal Royal Statistical Society, Series B*, 36:192–236, 1974.

[Bes86] J.E. Besag. On the statistical analysis of dirty pictures. *Journal Royal Statistical Society, Series B*, 48:259–302, 1986.

[Bla89] Andrew Blake. Comparison of the efficiency of deterministic and stochastic algorithms for visual reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(1):2–12, January 1989.

[BT92] Luigi Bedini and Anna Tonazzini. Image restoration preserving discontinuities: the Bayesian approach and neural networks. *Image and Vision Computing*, 10(2):108–118, March 1992.

[DE87] Haluk Derin and Howard Elliott. Modeling and segmentation of noisy and textured images using Gibbs random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(1):39–55, January 1987.

[GG84] Stuart Geman and Donald Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, November 1984.

[GG91] D. Geiger and F. Girosi. Parallel and deterministic algorithms from MRF's: Surface reconstructions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-13(5):401–413, May 1991.

[GGGD90] D. Geman, S. Geman, C. Graffigne, and P. Dong. Boundary detection by constrained optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):609–628, July 1990.

[GKT92] A.J Gray, J.W. Kay, and D.M. Titterington. On the estimation of noisy binary Markov random fields. *Pattern Recognition*, 25(7):749–768, 1992.

[HK90] E.R. Hancock and J. Kittler. A label error process for discrete relaxation. In *Proceedings of the 10th international conference on pattern recognition*, pages 523–527. IEEE, June 1990.

[JW90] F.C. Jeng and J.W. Woods. Simulated annealing in compound Gaussian random fields. *IEEE Transactions on Information Theory*, 36(1):94–107, January 1990.

[KY91] Il Young Kim and Hyun Seung Yang. Markov random field based image labeling with parameter estimation by error backpropagation. *IEICE Transactions on Information and Systems*, E74(10):3513–3521, October 1991.

[LD89] S. Lakshmanan and H. Derin. Simultaneous parameter estimation and segmentation of Gibbs random fields using simulated annealing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(8):799–813, August 1989.

[Mar85] Jose Luis Marroquin. Probabilistic solution of inverse problems. Technical Report TR 860, MIT Artificial Intelligence Laboratory, September 1985.

[MC91] B. S. Manjunath and R. Chellappa. Unsupervised texture segmentation using Markov random field models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-13(5):478–482, May 1991.

[MMP85] J. Marroquin, S. Mitter, and T. Poggio. Probabilistic solution of ill-posed problems in computational vision. In Lee S. Bauman, editor, *Proceedings: Image Understanding Workshop*, pages 293–309. Science Applications International Corporation, December 1985.

[MS91] Davin Milun and David Sher. Improving Markov random field reconstructions by using noisy probabilities. In V. Cantoni, M. Ferretti, S. Levialdi, R. Negrini, and R. Stefanelli, editors, *6th International Conference on Image Analysis and Processing : Progress in Image Analysis and Processing II*, pages 147–154. World Scientific, September 1991.

[MS93] Davin Milun and David Sher. Improving sampled probability distributions for Markov random fields. *Pattern Recognition Letters*, 14(10):781–788, October 1993.

[Pra78] W. K. Pratt. *Digital Image Processing.* Wiley-Interscience, New York, 1978.

[She91] David Sher. Minimizing the cost of errors with a Markov random field. *Pattern Recognition Letters*, 12(2), February 1991.

[Swa90] Michael J. Swain. Parameter learning for Markov random fields with highest confidence first estimation. Technical Report TR350, Department of Computer Science, University of Rochester, August 1990.

[TGD92] H. L. Tan, S. B. Gelfand, and E. J. Delp. A cost minimization approach to edge detection using simulated annealing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-14(1):3–19, January 1992.

[ZZAA92] Yunxin Zhau, Xinhau Zhuang, Les Atlas, and Lars Anderson. Parameter estimation and restoration of noisy images using Gibbs distributions in hidden Markov models. *CVGIP: Graphical Models and Image Processing*, 54(3):187–197, May 1992.