Training Double Neighborhood Markov Random Fields by Sampling—How Much Data is Required?

Davin Milun
milun@cs.buffalo.edu

David B. Sher
sher@cs.buffalo.edu

Computer Science Department
SUNY at Buffalo
226 Bell Hall
Buffalo, NY, USA, 14260

Abstract

We have been developing edge relaxation and binary image enhancement systems using parameters derived from an ensemble of training images [25, 22, 23, 26, 27]. We tabulate the frequencies of local structures in the training ensemble and reconstruct noisy images so that they best match the local characteristics of the set of training images. This paper investigates how many such training images are required to generate a useful and consistent set of local neighborhood probabilities.

1 Introduction

We have been developing edge relaxation and binary image enhancement systems using parameters derived from an ensemble of training images and ground truth images [25, 22, 23, 26, 27]. Our scheme is based on Markov random fields, but also allows for arbitrary, correlated noise in the images. We tabulate the frequencies of local structures in the training ensemble (pairs from the noisy and ground truth images) and reconstruct noisy images so that they best match the local characteristics of the set of training images. See figure 1.

The distributions resulting from sampling, even from many images, are sparse (improbable configurations often do not occur in the sample) and inaccurate for improbable configurations, so we use a technique developed by Hancock and Kittler [13] to "smooth" the distribution. We have previously shown that their technique improves binary image reconstructions [27].

This paper investigates how many pairs of training images are required to generate a useful and consistent set of local neighborhood probabilities.

2 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 [2, 3, 10]. 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 [14, 15, 9, 11, 30, 1, 18, 21, 20, 28, 8, 4, 17, 5, 29]. However, not much attention has been paid to the methods of deriving the probability density functions of the MRFs.

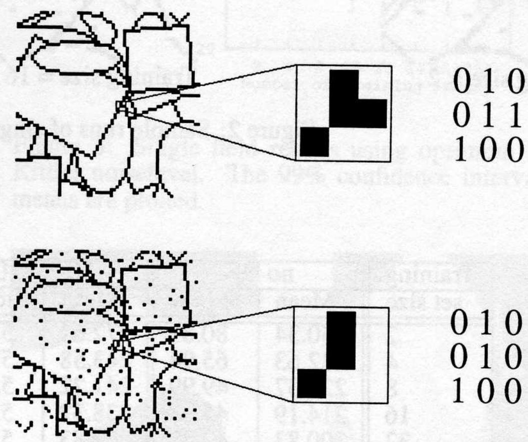

Figure 1: Example of a double neighborhood probability. The uncorrupted image is shown above, and the noisy, corrupted image is shown below. Enlarged is one neighborhood pair. The existence of this particular neighborhood pair would be recorded in the probability density function in position 010010100010011100_2 (reading the corrupted image as the high bits) which is decimal 75932.

Gray, Kay and Titterington [12] follows up on work by Derin and Elliott [6] 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 our technique. However, their methods do not make use of sets of training images and place many more restrictions on the form of the MRF than we do.

Lakshmanan and Derin [17] and Manjunath and Chellappa [19] 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. They also do not employ a training set of images, and also restrict the form of the local neighborhood interactions.

Kim and Yang [16] use the error backpropagation technique from neural networks to estimate the MRF parameters.

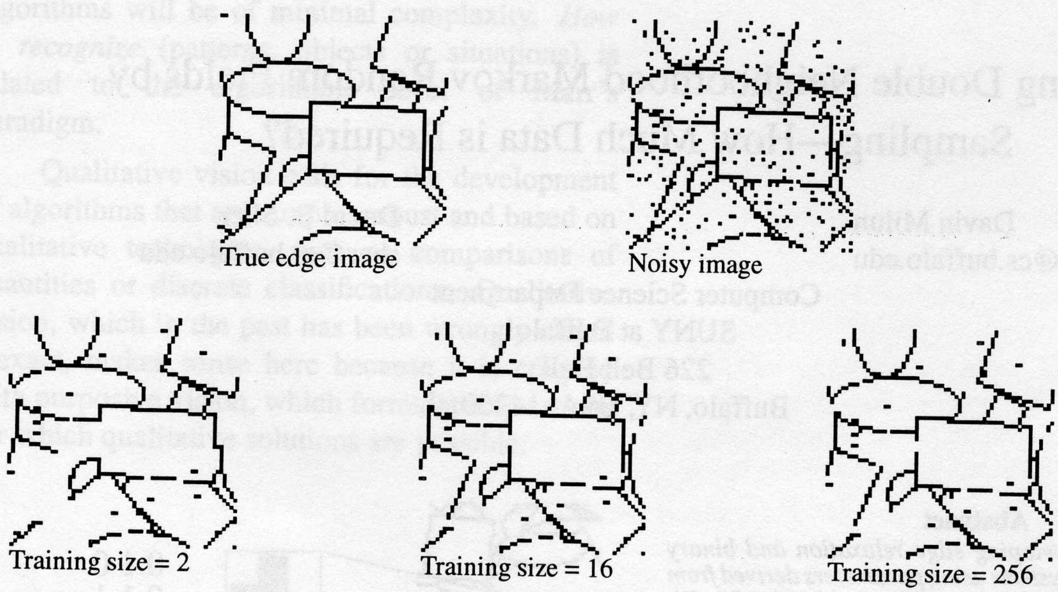

Figure 2: Sample runs of single field case with Optimal HK.

Training set size	no HK Mean	no HK Stddev	HK=.010 Mean	HK=.010 Stddev	HK=.036 Mean	HK=.036 Stddev	HK=.050 Mean	HK=.050 Stddev
2	350.34	80.81	272.64	58.58	151.70	35.53	149.95	14.99
4	282.63	65.94	243.58	52.80	134.52	26.66	148.81	14.84
8	228.37	49.99	241.22	55.99	133.40	26.17	148.93	14.82
16	214.19	45.85	238.32	52.18	123.99	17.57	148.83	14.85
32	200.83	42.32	242.83	51.68	124.03	17.59	148.90	14.82
64	196.80	41.11	241.23	53.15	124.11	17.62	148.85	14.78
128	193.25	42.28	243.72	53.11	123.94	17.59	148.79	14.82
256	194.89	40.71	240.66	52.93	124.07	17.59	148.84	14.80

Table 1: Numerical results of the single field study using the bits of error as the metric. Each row contains the results generated with a set training pairs of one size, but the results are the average over 4 different sets of training pairs of that size. All reconstructions were performed on the same set of 128 test images.

Training set size	no HK Mean	no HK Stddev	HK=.006 Mean	HK=.006 Stddev	HK=.010 Mean	HK=.010 Stddev	HK=.025 Mean	HK=.025 Stddev	HK=.050 Mean	HK=.050 Stddev
2	276.16	66.83	252.91	67.74	253.88	66.02	252.58	67.69	253.04	68.64
4	272.28	74.84	247.99	64.16	244.48	66.55	243.90	67.35	241.17	64.12
8	265.48	66.45	241.57	66.85	244.07	65.13	240.96	65.87	241.19	67.50
16	271.57	67.43	242.44	66.63	236.16	67.58	241.50	66.98	243.46	68.53
32	266.98	68.50	240.20	66.56	242.04	65.29	240.09	66.47	235.72	67.00
64	274.31	70.17	243.46	63.31	240.24	66.81	237.85	65.47	236.20	63.26
128	272.10	68.76	244.65	69.12	240.54	69.24	238.78	64.28	235.06	65.75
256	275.34	71.50	242.08	70.41	240.01	65.51	235.60	63.61	232.89	64.30

Table 2: Numerical results of the double field study using the bits of error as the metric. Each row contains the results generated with a set training pairs of one size, but the results are the average over 4 different sets of training pairs of that size. All reconstructions were performed on a random set of 256 test images.

3 Algorithm

We are doing binary image reconstruction using MRF methods. We use a method very similar to Besag's ICM method [3] to solve our fields.

In the single field case, we store the marginal probabilities of each of the 3×3 neighborhoods, and iterate through the image, changing the center pixel of neighborhoods as random positions, based on whether the changed or unchanged neighborhood is more probable to occur.

In the double field case, we use the same method, except that now we keep a set of 512 marginal probability distributions, one for each of the 512 existing neighborhoods. Thus, at each stage, we decide whether to change a pixel, based on the distribution associated with the particular 3×3 neighborhood in the noisy image at the pixel.

4 Results

We have performed two main sets of tests: in [24] with regular ("single field") MRFs (which are summarized below) and in this paper with our double neighborhood MRFs ("double field").

4.1 Single field MRFs

Our test data generator for the single field study creates 64×64 binary images with 50 randomly placed, overlapping ellipses and randomly rotated rectangles all of random size. Our training set contains 2048 such images; our test set contains 16 such images, but with 6% of their bits randomly flipped.

In order to gauge the performance of our method we used the metric of the number of bits that differ between the reconstructed image and the true image.

Table 1 shows the numerical results we achieved. The amount of training data required can be summarized as follows: For no HK, there is no statistically significant improvement after 32 images. For the case of HK = 0.010, there is no statistically significant improvement after 4 images. For the case of HK = 0.050 it does not matter how many training images were used.

For the most interesting case, when HK is optimal at 0.036, we find that there is no statistically significant improvement after 16 training images were used. Figure 2 shows some sample runs of this case, and figure 3 shows the plot of the results with 99% confidence interval error bars.

4.2 Double neighborhood MRFs

Our test data generator for the double field study creates pairs of gray scale images together with the corresponding ground truth edge images. Each pair consists of randomly placed, randomly sized, circles of random solid color, and an edge map representing those pixels through which an edge passes. Some examples of these images are shown in figure 4. Our training set contains 2048 such pairs of gray scale images. For our test set we generated random pairs of images, as they were needed during the test runs. Each image was corrupted with additive Gaussian noise of standard deviation 16 and then corrupted with JPEG compression down to 1.5 bits per pixel (bpp). We then ran a thresholded, thinned Sobel [7] edge detector on the set of doubly corrupted images. We trained the double neighborhood MRF on the pairs of true edge images and edge images of the corrupted images. An example image can be found in figure 5.

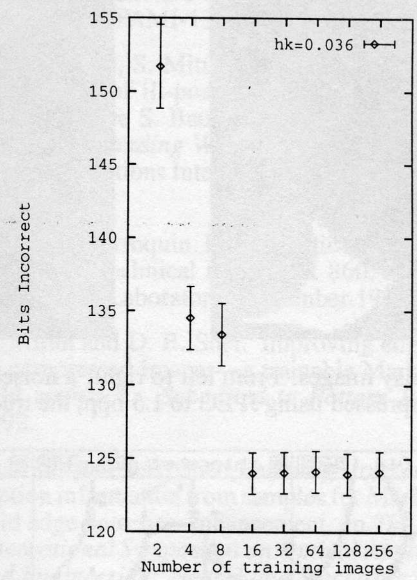

Figure 3: Single field results using optimum Hancock-Kittler noiselevel. The 99% confidence intervals of the means are plotted.

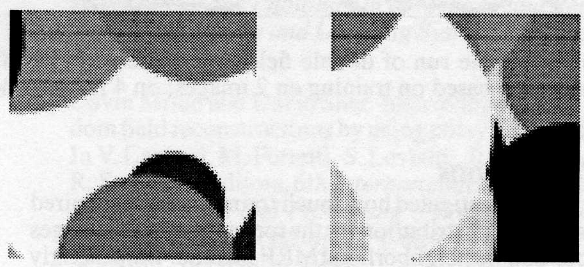

Figure 4: Some instances of generated gray scale random images.

An ICM [3] algorithm was applied to the test set at a variety of Hancock-Kittler smoothing levels. Figure 6 shows the results with HK=0.010 using distributions sampled from training sets of 3 different sizes.

In order to gauge the performance of our method we used the metric of the number of bits that differ between the reconstructed image and the true image.

Using this measure, we previously found that our technique of reconstruction shows an improvement over the thinned Sobel [22, 26]. Those results were generated by training on an ensemble of 128 images of size 64×64.

Table 2 shows the numerical results we achieved. For double fields, the amount of training data required can be summarized as follows: For no HK and HK = 0.06, there is no statistically significant difference regardless of the number of training images used. For the cases where HK = 0.010 – 0.050 there are no pairwise significant differences, however there is a statistically significant difference between the very low and very high number of sampled images. The HK = 0.050 results are displayed in figure 7.

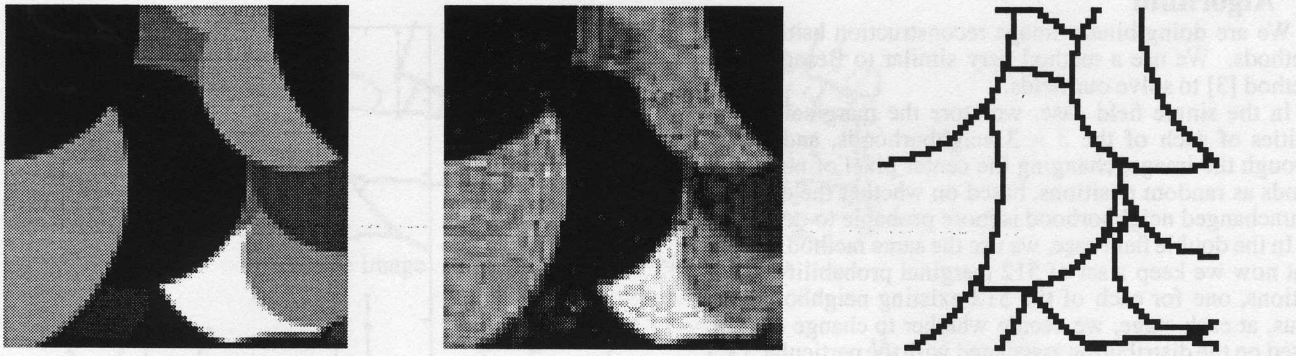

Figure 5: Noisy images. From left to right: a noiseless image; an image with standard deviation 16 Gaussian noise added and then compressed using JPEG to 1.5 bpp; the true edge in this image

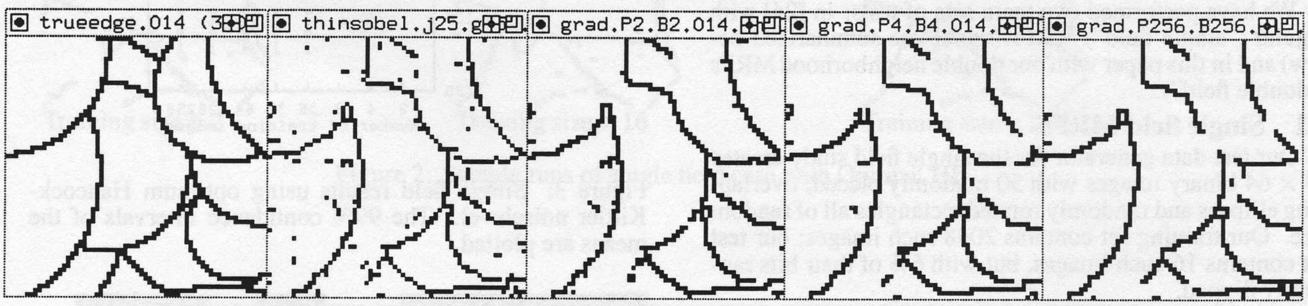

Figure 6: Sample run of double field case. Showing, in left to right order: the true edge image; the thinned Sobel; the reconstruction based on training on 2 images; on 4 images and on 256 images.

5 Conclusions

We have investigated how much training data is required to form a useful distribution for the reconstruction of images using our double neighborhood MRF method. Surprisingly we have found that only a few images are needed to obtain results as good as those which used as many as 256 training images.

References

[1] 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.

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

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

[4] 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.

[5] Paul Chou. Multi-modal segmentation using Markov random fields. In *IJCAI*, July 1987.

[6] 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.

[7] Richard O. Duda and Peter E. Hart. *Pattern Classification and Scene Analysis*. John Wiley & Sons, Inc., New York, London, Sydney, Toronto, 1973.

[8] 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.

[9] 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.

[10] 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.

[11] J. K. Goutsias. A theoretical analysis of Monte Carlo algorithms for the simulation of Gibbs random field images. *IEEE Transactions on Information Theory*, IT-37(6):1618–1628, November 1991.

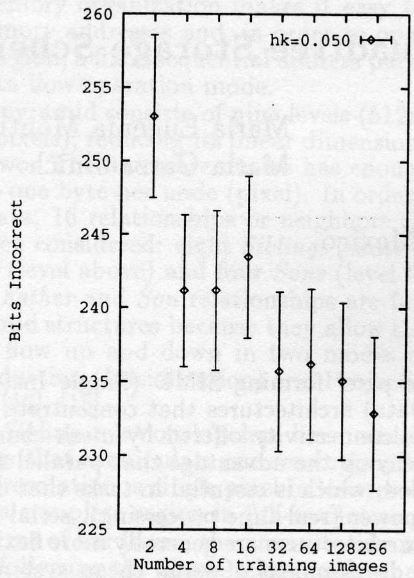

Figure 7: Double field results using Hancock-Kittler level of 0.050. The 99% confidence intervals of the means are plotted.

[12] 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.

[13] 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.

[14] 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.

[15] N. Karssemeijer. A relaxation method for image segmentation using a spatially dependent stochastic model. *Pattern Recognition Letters*, 11:13–23, 1990.

[16] 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.

[17] 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.

[18] Gabor Lugosi. Learning with an unreliable teacher. *Pattern Recognition*, 25(1):79–85, 1992.

[19] 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.

[20] 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.

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

[22] D. Milun and D. B. Sher. Improving edge detectors on compressed images—a trainable Markov random field approach. *Submitted to Pattern Recognition*, 1992.

[23] D. Milun and D.B. Sher. Learning structural and corruption information from samples for Markov random field edge detection enhancement. In *The SPIE 1992 International Symposium on Optical Applied Science and Engineering: Neural and Stochastic Methods in Image and Signal Processing*, San Diego, 1992.

[24] D. Milun and D.B. Sher. Training Markov random fields by sampling—How much data is required? In *The SPIE 1993 International Symposium on Optical Engineering and Photonics in Aerospace and Remote Sensing: Adaptive and Learning Systems II*, Orlando, April 1993.

[25] 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.

[26] Davin Milun and David Sher. Learning structural and corruption information from samples for Markov random field binary image reconstruction. In *11th ICPR: Conference on Image, Speech and Signal Analysis*, volume 3, pages 513–516, The Hague, The Netherlands, September 1992.

[27] Davin Milun and David Sher. Improving sampled probability distributions for Markov random fields. *To appear in Pattern Recognition Letters*, 1993.

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

[29] 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.

[30] 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.