100 and 1000 images respectively. In the sparse and dense model ranges, we observe again a better performance of “Bayes” than $L_1$-based methods. “Bayes PM” also shows robustness to under/over-fitting although it seems that simply computing the posterior mean does not provide sufficiently good model parameters in the median density range.

To get a more intuitive comparison about the quality of learned sparse models, we train models on 1000 images by different methods with a density of 0.2 and then run Gibbs sampling to draw 36 samples from each model. The images are shown in Figure 11. While it is hard to get good reconstruction using a model without hidden variables, the Bayesian methods produce qualitatively better images than competing methods, even though “Bayes PM” does not have higher CLL than “MLE” at this level.

A common limitation of learning Bayesian models with the MCMC inference method is that it is much slower than training a model with a point estimate. However, as shown in the experiments, the Bayesian methods are able to learn a good combination of parameters and a structure without the need to tune hyper-parameters through cross validation. Also, the Bayesian methods learn sparser models than $L_1$-based methods without sacrificing predictive performance significantly. Because the computational complexity of inference grows exponentially with the maximum clique size of MRFs, $L_1$-based models at their optimal (not so sparse) regularization level can in fact become significantly more computationally expensive than their Bayesian counterparts at prediction time. Turning up the regularization will result in sparser models but at the cost of under-fitting the data and thus sacrificing predictive accuracy.

7 Discussion

We propose Bayesian structure learning for MRFs with a spike and slab prior. An approximate MCMC method is proposed to achieve effective inference based on Langevin dynamics and reversible jump MCMC. As far as we known this is the first attempt to learn MRF structures in the fully Bayesian approach using spike and slab priors. Related work was presented in Parise and Welling (2006) with a variational method for Bayesian MRF model selection. However this method can only compare a given list of candidate models instead of searching in the exponentially large structure space.

The proposed MCMC method is shown to provide accurate posterior distributions at small step sizes. The selective shrinkage property of the spike and slab prior enables us to learn an MRF at different sparsity levels without noticeably suffering from under-fitting or over-fitting even for a small data set. Experiments with simulated data and real-world data show that the Bayesian method can learn both an accurate structure and a set of parameter values with strong predictive performance. In contrast the $L_1$-based methods could fail to accomplish both tasks with a single choice of the regularization strength. Also, the performance of our Bayesian model is largely insensitive to the choice of hyper-parameters. It provides an automated way to choose a proper sparsity level, while $L_1$ methods usually rely on cross-validation to find their optimal regularization setting.

Acknowledgement

This material is based upon work supported by the National Science Foundation under Grant No. 0914783, 0928427, 1018433.