Higher-order MRFs based image super resolution:
MMSE or MAP?
Yunjin Chen

Abstract—A trained filter-based higher-order Markov Random Fields (MRFs) model - the so called Fields of Experts (FoE), has proved a highly effective image prior model for many classic image restoration problems. Generally, two options are available to incorporate the learned FoE prior in the inference procedure: (1) sampling-based minimum mean square error (MMSE) estimate, and (2) energy minimization-based maximum a posteriori (MAP) estimate. It is well-known the sampling-based MMSE estimate is very time consuming, but the MAP inference has a remarkable advantage of high computational efficiency. In a recent paper, the FoE prior model was exploited for the single image super resolution (SR) task by using the MMSE inference, based on a seemingly correct conclusion that the MAP inference of the FoE prior based model, which leads a non-convex optimization problem, is prone to getting stuck in some bad local minima. However, in this letter, we demonstrate that this simpler inference criterion - the MAP estimate, works equally well compared to the complicated MMSE estimate with exactly the same prior model. Moreover, with our discriminatively trained FoE prior model, the MAP inference can even lead to further improvements. Consequently, we argue that for higher-order natural image prior based SR problem, it is not necessary to employ the time consuming MMSE estimation.

Index Terms—Bayesian minimum mean square error, a maximum a posteriori, Fields of Experts, single image super resolution

I. INTRODUCTION

Markov Random Fields (MRFs) based models have a long history in low-level computer vision problems, which treat the image as a random field \[5\]. It is now well-known that MRFs are particularly effective for image prior modeling in image processing. In a MRF-based image prior model, the probability of a whole image is defined based on the potential (or energy) of the overlapping local cliques.

An elegant MRF-based image prior model, called Fields of Experts (FoE) was recently proposed by Roth and Black \[8\]. The proposed FoE model is defined by (1) a heavy-tailed potential function, which is derived from the observation that the filter response of natural images exhibit heavy-tailed distribution when applying derivative filters onto them, (2) a set of linear filters, which are trained from image samples.

Due to its effectiveness of the FoE image prior model for many image restoration problems, many works have been devoted to the FoE-based image restoration problems, such as image denoising, inpainting, deblurring, etc. Usually, there are two options to investigate the learned FoE prior model for specific image restoration problems: (1) using sampling-based MMSE estimation, such as \[10\], \[3\], \[13\], \[14\], or (2) energy minimization based MAP estimation, such as \[8\], \[9\], \[11\], \[2\].

It is \[10\] that for the first time claimed that the MMSE estimation can lead to improved performance compared to the MAP estimation for the image denoising task with their learned FoE image prior model. After that, many works follow their suggestion to make use of the MMSE estimation for FoE related models, such as image deblurring \[11\], \[15\], image denoising \[3\], depth estimation \[12\], \[4\], image separation \[13\] and single image super resolution \[14\]. As we know, the sampling based MMSE estimation is very slow, making the resulting higher-order MRF based model hardly appealing for practical applications.

In a recent paper \[14\], the FoE prior model was exploited in the context of image super resolution. The authors also proposed to employ the MMSE estimate in the inference procedure, other than the MAP estimate. They claimed that: “the MAP estimation could not exploit the full potential offered by the probabilistic modeling, as only the posterior mode is sought for solution. Even worse, it is prone to getting stuck in local minima when the prior model is a heavy-tailed distribution, which actually leads to a non-convex optimization problem, thus there is no guarantee of obtaining even the posterior mode.” With the MMSE estimate, the FoE-based SR model demonstrates a state-of-the-art SR algorithm. However, it is well known that the sampling based approach is very time consuming, alluding to the fact that the FoE-based SR model is not appealing for practical applications.

A. Arguments in MAP and MMSE estimate for FoE-based denoising

These researchers made the choice to utilize the MMSE estimation based on a seemingly correct conclusion that the MMSE estimate performances better than the MAP estimate for the FoE based models.

However, in our previous works \[11\], \[2\], we demonstrate that at least for Gaussian denoising, the performance of the MAP inference has been significantly underestimated in the previous work \[9\] because of an imperfect training scheme. We show that with our learned FoE prior obtained from a refined training algorithm, the performance of the MAP inference is dramatically boosted. As a consequence, the MAP-based denoising model with our trained FoE prior leads to the best-performing MRF model among the MRF-based systems. Therefore, we argue that MAP-based denoising model does not perform well in previous works, e.g., \[10\], \[9\] just because
they have not obtained a good FoE prior well-suited for the MAP inference. It is not true to blame the MAP inference and the the crux of the matter is the image prior model itself. With proper FoE prior models, the MAP inference can do a better job, and therefore there is no need to use the seemingly advanced but time-consuming MMSE estimation.

B. Our contributions

In this letter, we demonstrate that the MAP inference of the FoE-based SR model has been underestimated in the recent work [14]. We show that even with exactly the same image prior model exploited in the MMSE estimation, the MAP inference can achieve equivalent performance in terms of both quantitative measurements (PSNR and SSIM values) and visual perception quality. In additional, the MAP inference can achieve equivalent performance in terms of both quantitative measurements (PSNR and SSIM values) and visual perception quality. In additional, the MAP inference can obtain further improvements with our discriminatively trained FoE image prior of the same model capacity. It is clear that the MAP inference has a obvious advantage of efficiency, and this advantage is even more pronounced with an effective quasi-Newton’s optimization algorithms - L-BFGS [6], or with our recently proposed non-convex optimization algorithm - iPiano [7].

Therefore, to sum up, we think that it is not necessary to use the time-consuming estimation MMSE for solving FoE prior-based image super resolution problem.

This conclusion obtained based on the image super resolution problem investigated in this letter strengthens our arguments drawn based on the Gaussian denoising problem. Consequently, we think it is better to employ the MAP inference when incorporating the FoE prior model for image restoration problems, because (1) there is no performance loss by using this simpler inference criterion, and (2) the MAP inference has an obvious advantage of highly efficiency.

II. MAP INFEERENCE OF FOE IMAGE PRIOR BASED SR

In a typical image super resolution task, the low-resolution (LR) image is generated from a high-resolution (HR) image using the following formulation

\[ y = DBx + \varepsilon, \]

where \( x \in \mathbb{R}^n \) and \( y \in \mathbb{R}^m \) is the HR and LR image, respectively. \( B \in \mathbb{R}^{n \times n} \) is the matrix corresponding to the blurring operation and \( D \in \mathbb{R}^{m \times n} \) (\( m < n \)) signifies the down-sampling operation. \( \varepsilon \in \mathbb{R}^m \) is the noise (typically assumed to be Gaussian white noise with level \( \sigma \)).

The FoE image prior based SR model is formulated by the following Bayesian probabilistic model

\[ p(x|y) \propto \exp(-\frac{||y - DBx||^2}{\sigma^2})p(x), \tag{II.1} \]

where \( p(x) \) is the probability density of an image \( x \) under the FoE framework, written as

\[ p(x) = \frac{1}{Z} \Pi_{c \in C} \Pi_{i=1}^N \phi((k_i \ast x)_c; \alpha_i) \]

where \( C \) is the maximal cliques, \( N \) is the number of the filters, \( (k_i \ast x)_c \) refers to the \( c \)-th pixel in the filtered image by \( k_i \), \( \phi \) is the potential function with associated weights \( \alpha \). In [14], the potential function is given by the Gaussian scale mixtures (GSMs) as

\[ \phi(z; \alpha) = \sum_{j=1}^{J} \alpha_{c,j} \mathcal{N}(z; 0, \eta_j^2 / s_j), \tag{II.2} \]

where \( \alpha_{c,j} \) are the normalized weights of the Gaussian component with scale \( s_j \) and base variance \( \eta_j^2 \).

According to the posterior (II.1), [14] used the sampling-based MMSE estimation to recover the underlying HR image \( x \). In this letter, we consider the MAP estimate. With the MAP estimation, the FoE-based SR task is formulated as the following energy minimization problem

\[ \hat{x} = \arg \min_x E(x) = \sum_{i=1}^{N} \rho(k_i \ast x) + \frac{\lambda}{2} \| DBx - y \|^2_2, \tag{III.3} \]

where \( \rho(k_i \ast x) = \sum_{c \in C} \rho((k_i \ast x)_c) \) with penalty function \( \rho = -\log \phi \) defined in (II.2).

Gradient-based algorithms are applicable to solve the minimization problem (II.3). First, we need to calculate the gradient \( \nabla \mathcal{E} \), which is given as

\[ \nabla_x \mathcal{E} = \sum_{i=1}^{N} K_i^\top \rho'(K_i x) + \lambda (DB)^\top (DBx - y), \tag{III.4} \]

where \( K_i \in \mathbb{R}^{n \times n} \) a highly sparse matrix, implemented as 2D convolution of the image \( x \) with filter kernel \( k_i \), i.e., \( K_i x \Leftrightarrow k_i \ast x, \rho'(K_i x) = (\rho'((K_i x)_1), \cdots, \rho'((K_i x)_n))^\top \in \mathbb{R}^n \), with \( \rho'(z) = \frac{z}{\|z\|^2} \sum_{j=1}^{J} \frac{\eta_j^2}{\eta_j^2 / s_j} \mathcal{N}(z; 0, \eta_j^2 / s_j). \)

In our work, we consider a quasi-Newton’s method - L-BFGS [6] and a newly developed non-convex optimization - iPiano [7] to solve the above minimization problem, instead of the commonly used conjugate gradient (CG) algorithm. We find that L-BFGS and iPiano works equally well for this problem, while both are extremely faster than CG. We refer the interested readers to [7] for more details about the iPiano algorithm.

III. EXPERIMENTAL RESULTS

As the purpose of this letter is not to propose a new state-of-the-art SR approach, but to compare the performance of the MAP inference and the MMSE inference for the FoE prior based SR model, we do not carry out comparisons with other state-of-the-art SR approaches, but only conduct a rigorous comparison with the MMSE estimate. The implementation of the MMSE estimate is provided by the authors of [14], and we use it as is.

In order to conduct a definitely fair comparison with the MMSE estimation, we first considered the MAP estimation with exactly the same image prior model exploited in [14] (8 filters of size 3 x 3 with GSMs potential), show in Figure 1 (quoted from [14]). We repeated the experiments presented in the TABLE I of [14], where eight noise-free images were upscaled with a zooming factor of 3. The results of the MMSE and MAP estimates are shown in Table I. One can see that the MAP estimate using the same image prior model
performs equally well when compared to the MMSE estimate, in terms of PSNR and SSIM index.

In order to investigate the potential performance of the MAP-based SR model, we incorporated a discriminatively trained FoE prior model into the SR model (II.3), which is directly optimized based on the MAP estimate in the context of Gaussian denoising. We employed the Student-t based FoE model trained in our previous work [2], which is defined as

\[ \rho(z) = \log(1 + z^2), \]

where the penalty function is given as the Lorentzian function

\[ E_{\text{FoE},\sigma}(x) = \sum_{i=1}^{N} \theta_i \rho(k_i * x), \quad \text{(III.1)} \]

derived from the Student-t distribution shown in Figure 1(b), and \( \theta_i \) is the weight of the corresponding filter \( k_i \). The employed filters are shown in Figure 2(a).

The results of the MAP-based SR model with the discriminatively trained FoE prior (III.1) are also shown in Table I. One can see that the MAP inference with our discriminatively trained FoE model can further improve the PSNR and SSIM results. An illustrative example is presented in Figure 3.

For the cases of mild Gaussian noise, the results of the MAP inference with two different FoE image prior models are shown in Table II. Again, one can see that the MAP estimate with the same FoE model (i.e., (II.2)) works equally well, and with our discriminatively trained FoE prior (III.1), it obtains better results.

For the MAP estimate based SR model (II.3), we need to search an optimal \( \lambda \) for each case. For the noise-free image SR task, we use a relative larger \( \lambda = 200 \), and for the SR tasks with Gaussian noise, we find the following empirical choice (1) \( \lambda = 3, \) if \( \sigma = 1, \) (2) \( \lambda = 1, \) if \( \sigma = 2, \) and (3) \( \lambda = 0.5, \) if \( \sigma = 3, \) generally works well.

Run time: As we know, the sampling-based MMSE estimate is very time-consuming. Now let us have a comparison of the run time. We run the inference algorithms on a server with CPUs: Inter(R) Xeon(R) CPU E5-2680 v2 @ 2.80GHz. For the SR task of upsampling an image of size 255 × 255, the average computation time per iteration of the MMSE-based algorithm is 87s. Typically, the MMSE estimate takes 100 iterations, and therefore for this SR task, it requires about \( \sim 2.4h, \) making this approach hardly appealing for practical application.

In contrast, the MAP inference is much faster. The average computation time per iteration of the MAP inference is (1) \( 0.094s \) for the GSM based FoE prior (II.2) and (2) \( 0.039s \) for Student-t based FoE prior (III.1). The GSM-based model is slower because of the computation of multiple scales. Typically, it takes 300~400 iterations to solve the resulting non-convex minimization problem incorporating the GSM-based FoE model. However, for the minimization problem having

| \( \sigma \) | Methods | House | Peppers | Cameraman |
|---|---|---|---|---|
| 1 | MMSE with prior (II.2) | 31.26/87.74 | 25.69/88.83 | 26.13/82.80 |
| MAP with prior (III.1) | 32.03/87.75 | 26.23/89.24 | 26.17/81.96 |
| 2 | MMSE with prior (II.2) | 30.47/85.80 | 25.23/86.00 | 25.6/80.20 |
| MAP with prior (III.1) | 30.84/85.62 | 25.38/85.75 | 25.24/78.82 |
| 3 | MMSE with prior (II.2) | 29.33/83.21 | 24.54/82.32 | 24.94/77.04 |
| MAP with prior (III.1) | 30.59/84.63 | 25.10/85.19 | 25.26/77.97 |

Table I

**Noisy image SR \((x \times 3)\) result comparison between the MMSE and MAP estimate (PSNR/100×SSIM). The better results between the first two rows are colored with blue. The best results are highlighted in bold.**

Table II

![Figure 1](image1.png)

![Figure 2](image2.png)
the Student-t based FoE, 150~200 iterations is generally sufficient, because this problem is smoother due to the smoother Lorentzian function, thus easier to solve. As a consequence, the MAP inference with the Student-t based FoE prior is able to accomplish the same SR task in 7s, which is dramatically faster than the MMSE inference (~2.4h), while even with better performance. Implementation will be available at our homepage (www.GPU4Vision.org) after acceptance.

In addition, as demonstrated in our previous works [1], [2], the MAP inference of the FoE prior based models can be easily implemented on GPU for parallel computation, which can generally obtain a speedup factor about 40×.

Given the superiority of the performance, as well as the run time of the MAP inference based SR model, it is clear that it is indeed not necessary to make use of the seemingly advanced MMSE estimate.

IV. DISCUSSION AND CONCLUSION

It is generally correct that the MAP estimate, which only seeks for the posterior mode, could not generally exploit the full potential offered by the probabilistic modeling, while the MMSE estimate, which directly draw samples from the probability model, should be more powerful.

However, in the context of higher-order MRFs based modeling, this general argument does not hold always. Concerning the higher-order MRFs based SR task, in this letter, we have shown unexpected results to the previous work, that we do not get stuck in bad local minimia. It turns out that the MAP estimate can work equally well, despite of the non-convexity of the resulting optimization problem. With our discriminatively trained higher-order MRF prior model, it can even perform significantly better.

As we know, the MMSE estimate suffers from the major drawback that it is very time-consuming, clearly preventing the higher-order MRF image prior based models from being an attractive approach of practical usage. In contrast, the MAP inference has an obvious advantage of high efficiency, making the resulting model highly appealing for practical usage.

The results of this letter, obtained based on the SR task, will strengthen our arguments drawn based on the Gaussian denoising problem in our previous work [1]. In summary, we hold the opinion that for higher-order MRF image prior based modeling, it is not necessary to exploit the time-consuming MMSE estimate, because there is no performance loss by using a simpler and faster inference criterion - the MAP estimate.

REFERENCES

[1] Y. Chen, T. Pock, R. Ranftl, and H. Bischof. Revisiting loss-specific training of filter-based mrf models for image restoration. In GCPR, pages 271–281, 2013.
[2] Y. Chen, R. Ranftl, and T. Pock. Insights into analysis operator learning: From patch-based sparse models to higher order MRFs. IEEE Transactions on Image Processing, 23(3):1060–1072, 2014.
[3] Q. Gao and S. Roth. How well do filter-based MRFs model natural images? In DAGM/OAGM Symposium, pages 62–72, 2012.
[4] C. D. Herrera, J. Kannala, P. Sturm, and J. Heikkila. A learned joint depth and intensity prior using markov random fields. In 3DTV-Conference, 2013 International Conference on, pages 17–24. IEEE, 2013.
[5] S. Z. Li. Markov random field modeling in computer vision. Springer-Verlag New York, Inc., 1995.
[6] D. C. Liu and J. Nocedal. On the limited memory BFGS method for large scale optimization. Mathematical Programming, 45(1):503–528, 1989.
[7] P. Ochs, Y. Chen, T. Brox, and T. Pock. iPiano: Inertial Proximal Algorithm for Nonconvex Optimization. SIAM Journal on Imaging Sciences, 7(2):1388–1419, 2014.
[8] S. Roth and M. J. Black. Fields of experts. International Journal of Computer Vision, 82(2):205–229, 2009.
[9] K. G. G. Samuel and M. Tappen. Learning optimized map estimates in continuously-valued MRF models. In CVPR, 2009.
[10] U. Schmidt, Q. Guo, and S. Roth. A generative perspective on MRFs in low-level vision. In CVPR, pages 1751–1758, 2010.
[11] U. Schmidt, K. Schelten, and S. Roth. Bayesian deblurring with integrated noise estimation. In Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, pages 2625–2632. IEEE, 2011.
[12] X. Wang, C. Hou, L. Pu, and Y. Hou. A depth estimating method from a single image using foe crf. Multimedia Tools and Applications, pages 1–16, 2014.
[13] H. Zhang and Y. Zhang. Bayesian image separation with natural image prior. In Image Processing (ICIP), 2012 19th IEEE International Conference on, pages 2097–2100. IEEE, 2012.

[14] H. Zhang, Y. Zhang, H. Li, and T. S. Huang. Generative bayesian image super resolution with natural image prior. Image Processing, IEEE Transactions on, 21(9):4054–4067, 2012.

[15] B. Zhao, W. Zhang, H. Ding, and H. Wang. Non-blind image deblurring from a single image. Cognitive Computation, 5(1):3–12, 2013.