SPARSE BAYESIAN IMAGE RESTORATION

S. Derin Babacan[1], *Rafael Molina*[2], *Aggelos K. Katsaggelos*[1]

[1]EECS Department
Northwestern University, Evanston, IL USA
sdb@northwestern.edu, aggk@eecs.northwestern.edu

[2]Departamento de Ciencias
de la Computación e I.A.
Universidad de Granada, Spain
rms@decsai.ugr.es

ABSTRACT

In this paper we propose a novel Bayesian algorithm for image restoration and parameter estimation. We utilize an image prior where Gaussian distributions are placed per pixel in the high-pass filter outputs of the image. By following the hierarchical Bayesian framework, we simultaneously estimate the unknown image and hyperparameters for both the image prior and the image degradation noise. We show that the proposed formulation is a special case of the popular l_p-norm based formulations with $p = 0$, and therefore enforces sparsity to an high extent in the filtered image coefficients. Moreover, the proposed formulation results in a convex optimization problem, and therefore does not suffer from the robustness issues common with non-convex image priors. Experimental results demonstrate that the proposed algorithm provides superior performance compared to state-of-the-art restoration algorithms although no user-supervision is required.

Index Terms— Image restoration, parameter estimation, Bayesian methods.

1. INTRODUCTION

This paper concerns with the image restoration problem where the original scene is estimated from a blurred and noisy observation. A standard formulation of the image degradation model is given in matrix-vector form by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{1}$$

where the vectors $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{n}$ represent respectively the original image, the available noisy and blurred image, and the noise with independent elements of variance β^{-1}, and $\mathbf{H}$ represents the known blurring matrix. The images are assumed to be of size $m \times n = N$ pixels, and they are lexicographically ordered into $N \times 1$ vectors. The restoration problem calls for finding an estimate of $\mathbf{x}$ given $\mathbf{y}$, $\mathbf{H}$, and knowledge about $\mathbf{n}$ and possibly $\mathbf{x}$ [1].

Currently, the state-of-the-art methods in image restoration generally incorporate image priors with heavier tails than Gaussian priors (see [2] for a recent comparison). This is in accordance with the fact that when the natural images are filtered with high-pass operators (such as derivative filters, or wavelets), the filter outputs exhibit non-Gaussian marginal statistics. Common approaches along this line incorporate Laplace priors (equivalent to utilizing l_1-norms) [3], TV-priors [4], mixture-of-Gaussian priors [5], and hyper-Laplacian distributions based on l_p-quasinorms with $0 < p < 1$ [2, 6]. The principle of modeling non-Gaussian image statistics is also coupled with the principle of sparsity (or compressibility) of filter coefficients, which is widely utilized in recent image restoration/reconstruction and compressive sensing algorithms [7].

In this paper, we propose a framework for image restoration that utilizes independent Gaussian priors on the filter output coefficients of the image. Most of the methods utilizing non-Gaussian priors result in non-convex optimization problems which are prone to robustness issues. We demonstrate that the proposed image prior, while convex, is very effective in modeling the underlying edge structure of the image. Using this approach within a hierarchical framework and incorporating a fully-Bayesian analysis, we develop a fully-automated algorithm that outperforms state-of-the-art restoration algorithms in terms of restoration performance.

2. BAYESIAN MODEL

In this work, we adopt a hierarchical Bayesian framework consisting of two stages. In the first stage, we model the observed image $\mathbf{y}$ using the conditional distribution $\mathrm{p}(\mathbf{y}|\mathbf{x}, \beta)$, and the unknown image using the prior $\mathrm{p}(\mathbf{x}|\mathbf{A})$ with $\mathbf{A} = \mathrm{diag}\,(\alpha_i)\,,\ i = 1, \ldots, N$. The parameters β and α_i of these distributions, also called *hyperparameters*, are modeled in the second stage using *hyperprior* distributions $\mathrm{p}(\alpha_i)$ and $\mathrm{p}(\beta)$.

For the observation model, we assume that the noise is zero-mean white Gaussian noise with the variance β^{-1}, such that with (1), it is represented by

$$\mathrm{p}(\mathbf{y}|\mathbf{x}, \beta) \propto \beta^{N/2} \exp\left[-\frac{\beta}{2} \parallel \mathbf{y} - \mathbf{H}\mathbf{x} \parallel^2\right]. \tag{2}$$

As the image model we use the following prior

$$\mathrm{p}(\mathbf{x}|\mathbf{A}) \propto |\sum_{k=1}^{L} \mathbf{D}_k^T \mathbf{A} \mathbf{D}_k|^{-1/2} \exp\left(-\frac{1}{2} \sum_{k=1}^{L} \mathbf{x}^T \mathbf{D}_k^T \mathbf{A} \mathbf{D}_k \mathbf{x}\right), \tag{3}$$

where $\mathbf{D}_k$, $k = 1, 2, \ldots L$ are $N \times N$ high-pass filter matrices which are used to impose smoothness constraints on the image estimate. The matrix $\mathbf{A}$ is the diagonal covariance matrix. Note that this prior is similar in spirit to the priors used in relevance vector machines [8].

As for the hyperpriors on the hyperparameters α_i and β, we choose the Gamma distributions as they are the conjugate priors for the inverse variances of the Gaussian distributions in (2) and (3).

This work was supported in part by the Comisión Nacional de Ciencia y Tecnología under contract TIC2007-65533 and the Spanish research programme Consolider Ingenio 2010: MIPRCV (CSD2007-00018).

Specifically, they are given by

$$\mathrm{p}(\alpha_i) = \Gamma(\alpha_i|a_\alpha^o, b_\alpha^o) = \frac{(b_\alpha^o)^{a_\alpha^o}}{\Gamma(a_\alpha^o)} \alpha^{a_\alpha^o - 1} \exp\left[-\alpha_i b_\alpha^o\right], \quad (4)$$

$$\mathrm{p}(\beta) = \Gamma(\beta|a_\beta^o, b_\beta^o) = \frac{(b_\beta^o)^{a_\beta^o}}{\Gamma(a_\beta^o)} \beta^{a_\beta^o - 1} \exp\left[-\beta b_\beta^o\right]. \quad (5)$$

Note that the parameters a_α^o and b_α^o are common for all hyperparameters α_i. Case of particular interest is when $a_\alpha^o = a_\beta^o = 1$ and $b_\alpha^o = b_\beta^o = 0$, which corresponds to the uniform distributions for the hyperpriors. In this case, no prior information is assumed about the hyperparameters and the observation $\mathbf{y}$ is made solely responsible for the restoration process.

Finally, using (2), (3), (4) and (5), the joint distribution $\mathrm{p}(\mathbf{y}, \mathbf{x}, \mathbf{A}, \beta)$ is defined as

$$\mathrm{p}(\mathbf{y}, \mathbf{x}, \mathbf{A}, \beta) = \mathrm{p}(\mathbf{y}|\mathbf{x}, \beta)\, \mathrm{p}(\mathbf{x}|\mathbf{A})\; \mathrm{p}(\beta) \prod_{i=1}^{N} \mathrm{p}(\alpha_i). \quad (6)$$

3. BAYESIAN INFERENCE

Using the joint distribution defined in (6), the Bayesian inference is based on the posterior distribution obtained using the Bayes' rule as follows:

$$\mathrm{p}(\mathbf{x}, \mathbf{A}, \beta|\mathbf{y}) = \frac{\mathrm{p}(\mathbf{y}, \mathbf{x}, \mathbf{A}, \beta)}{\mathrm{p}(\mathbf{y})}. \quad (7)$$

However, as in many applications, the posterior distribution cannot be obtained in closed form, since $\mathrm{p}(\mathbf{y}) = \int \mathrm{p}(\mathbf{y}, \mathbf{x}, \mathbf{A}, \beta) d\mathbf{x} d\mathbf{A} d\beta$ cannot be computed. Generally, one resorts to approximation methods such as the evidence approach, variational Bayesian methods, or maximum *a posteriori* estimates. In this work we adopt the evidence approach (also known as type-II maximum likelihood). We first decompose the posterior as

$$\mathrm{p}(\mathbf{x}, \mathbf{A}, \beta|\mathbf{y}) = \mathrm{p}(\mathbf{x}|\mathbf{y}, \mathbf{A}, \beta)\mathrm{p}(\mathbf{A}, \beta|\mathbf{y}). \quad (8)$$

The posterior distribution over the image is then given by

$$\mathrm{p}(\mathbf{x}|\mathbf{y}, \mathbf{A}, \beta) = \frac{\mathrm{p}(\mathbf{y}|\mathbf{x}, \beta)\mathrm{p}(\mathbf{x}|\mathbf{A})}{\mathrm{p}(\mathbf{y}|\mathbf{A}, \beta)}, \quad (9)$$

which is found to be a multivariate Gaussian distribution $\mathcal{N}(\mathbf{x}|\mu_\mathbf{x}, \Sigma_\mathbf{x})$ with parameters

$$\mu_\mathbf{x} = \Sigma_\mathbf{x}^{-1} \beta \mathbf{H}^T \mathbf{y}, \quad (10)$$

$$\Sigma_\mathbf{x}^{-1} = \left[\beta \mathbf{H}^T \mathbf{H} + \sum_{k=1}^{L} \mathbf{D}_k^T \mathbf{A} \mathbf{D}_k\right]. \quad (11)$$

The mean in (10) of this distribution is used as the estimate of the image. Next we consider the decomposition in (8) again to estimate the hyperparameters β and α_i. In the evidence approach, the hyperparameter posterior $\mathrm{p}(\mathbf{A}, \beta|\mathbf{y})$ in (8) is approximated by a delta-function at its mode, that is, it is represented by a degenerate distribution. The distribution $\mathrm{p}(\mathbf{A}, \beta|\mathbf{y})$ is proportional to the product of priors $\mathrm{p}(\mathbf{A})\mathrm{p}(\beta)$ and the evidence $\mathrm{p}(\mathbf{y}|\mathbf{A}, \beta)$. The evidence is found by marginalizing out the image $\mathbf{x}$, such that

$$\begin{aligned}\mathrm{p}(\mathbf{A}, \beta|\mathbf{y}) &\propto \mathrm{p}(\mathbf{A})\mathrm{p}(\beta)\mathrm{p}(\mathbf{y}|\mathbf{A}, \beta) \\ &= \mathrm{p}(\mathbf{A})\mathrm{p}(\beta) \int \mathrm{p}(\mathbf{y}|\mathbf{x}, \beta)\mathrm{p}(\mathbf{x}|\mathbf{A}) d\mathbf{x} \\ &\propto \mathrm{p}(\mathbf{A})\mathrm{p}(\beta)\mathcal{N}(\mathbf{y}|\mathbf{0}, \mathbf{C}), \end{aligned} \quad (12)$$

where

$$\mathbf{C} = \beta^{-1}\mathbf{I} + \mathbf{H}\left[\sum_{k=1}^{L} \mathbf{D}_k^T \mathbf{A} \mathbf{D}_k\right]^{-1} \mathbf{H}^T. \quad (13)$$

Using (12), we maximize the distribution $\mathrm{p}(\mathbf{A}, \beta|\mathbf{y})$ to estimate the hyperparameters β and α_i. This is equivalent to maximizing its logarithm, which is expressed as

$$\begin{aligned}\mathcal{L} &= \log \mathrm{p}(\mathbf{A}, \beta|\mathbf{y}) \\ &= -\frac{1}{2}\log|\mathbf{C}| - \frac{1}{2}\mathbf{y}^T\mathbf{C}^{-1}\mathbf{y} + \text{const} \\ &\quad + \sum_{i=1}^{N}(a_\alpha^o - 1)\log\alpha_i - b_\alpha^o\alpha_i + (a_\beta^o - 1)\log\beta - b_\alpha^o\beta. \end{aligned} \quad (14)$$

Using the Woodbury and determinant identities (see [8]), the first two terms in (14) can be rewritten as

$$\log|\mathbf{C}| = -\log|\Sigma_\mathbf{x}| - N\log\beta - \log\left|\sum_{k=1}^{L} \mathbf{D}_k^T \mathbf{A} \mathbf{D}_k\right|, \quad (15)$$

$$\mathbf{y}^T\mathbf{C}^{-1}\mathbf{y} = \beta \parallel \mathbf{y} - \mathbf{H}\mu_\mathbf{x} \parallel^2 + \sum_{k=1}^{L} (\mathbf{D}_k\mu_\mathbf{x})^T \mathbf{A} (\mathbf{D}_k\mu_\mathbf{x}). \quad (16)$$

Using these identities, the derivative of $\mathcal{L}$ with respect to $\mathbf{A}$ can be found as

$$\begin{aligned}2\frac{d\mathcal{L}}{d\mathbf{A}} &= \mathrm{Diag}\left(\left[\sum_{k=1}^{L} \mathbf{D}_k^T \mathbf{A} \mathbf{D}_k\right]^{-1} \sum_{k=1}^{L} \mathbf{D}_k^T \mathbf{D}_k\right) - \mathrm{Diag}\left(\sum_{k=1}^{L} \mathbf{D}_k^T \mathbf{D}_k \Sigma_\mathbf{x}\right) \\ &\quad - \sum_{k=1}^{L} (\mathbf{D}_k\mu_\mathbf{x})(\mathbf{D}_k\mu_\mathbf{x})^T + 2\,\mathrm{Diag}\left(\frac{(a_\alpha^o - 1)}{\alpha_i} - b_\alpha^o\right), \end{aligned} \quad (17)$$

where $\mathrm{Diag}(\cdot)$ is a diagonal matrix created with the diagonal elements of its argument. In order to solve for α_i satisfying $\frac{d\mathcal{L}}{d\alpha_i} = 0$, we utilize the approximation

$$\left[\sum_{k=1}^{L} \mathbf{D}_k^T \mathbf{A} \mathbf{D}_k\right]^{-1} \sum_{k=1}^{L} \mathbf{D}_k^T \mathbf{D}_k \approx \mathbf{A}^{-1}. \quad (18)$$

The values of α_i that maximize $\mathcal{L}$ can then be found as

$$\alpha_i = \frac{1 + 2(a_\alpha^o - 1)}{v_i + 2b_\alpha^o}, \quad (19)$$

where v_i is given by

$$v_i = \sum_{k=1}^{L} (\mathbf{D}_k\mu_\mathbf{x})_i^2 + \sum_{k=1}^{L} \left(\mathbf{D}_k^T \mathbf{D}_k \Sigma_\mathbf{x}\right)_{ii}, \quad (20)$$

where $(\cdot)_{ii}$ denotes the i^{th} diagonal element of the matrix. Similarly, the value of β that maximizes $\mathcal{L}$ can be found as

$$\beta = \frac{1 + 2(a_\beta^o - 1)}{\parallel \mathbf{y} - \mathbf{H}\mathbf{x} \parallel^2 + \mathrm{trace}\left(\mathbf{H}^T\mathbf{H}\Sigma_\mathbf{x}\right) + 2b_\beta^o}. \quad (21)$$

It can easily be shown that update equations of forms very similar to (19) and (21) are obtained if a variational Bayesian analysis is carried out.

In summary, the restoration algorithm iterates between estimating the image using (10), and estimating the hyperparameters using (19) and (21) until convergence. Note that the matrix $\Sigma_{\mathbf{x}}$ is of size $N \times N$, which with many images is probihitively large to construct explicitly. However, its explicit construction is not required since (10) can be solved very efficiently using a conjugate gradient algorithm. In (20) and (21), where its explicit construction is needed, we approximate it as a diagonal matrix by taking the reciprocals of the diagonal elements of $\Sigma_{\mathbf{x}}^{-1}$. Extensive experiments with small images showed empirically that this approximation results in very close estimates and has a minor effect in the estimation process.

We conclude this section by analyzing the sparsity enforcing property of the proposed formulation. When the shape and scale parameters in (4) and (5) are set equal to zero, i.e., $a_\alpha^o = a_\beta^o = 1$ and $b_\alpha^o = b_\beta^o = 0$, the hyperparameters α_i are estimated using $\alpha_i = 1/v_i$ (this also corresponds to the choice suggested in [8]). In locations where the filter outputs are very small (for instance in smooth regions), the values of v_i are very small, and therefore α_i values are very high. This corresponds to increased smoothness at this location due to the smoothness matrix $\sum_{k=1}^{L} \left(\mathbf{D}_k^T \mathbf{A} \mathbf{D}_k\right)$ in (11). On the other hand, in areas where the filter outputs are significantly high, the smoothness is less enforced due to low α_i values.

Let us now consider the l_p-norm based image priors, which correspond to the well-known *iteratively reweighted least squares* (IRLS) formulations [6]. Assuming that the l_p-norm is enforced on the filter outputs as is done in this work, the hyperparameter matrix is computed as [6]

$$\mathbf{A}^p = \text{diag}\left(\lambda v_i^{p/2-1}\right), \tag{22}$$

where λ is a parameter controlling the degree of smoothness, and v_i is defined in (20) (Note that [6] is a deterministic formulation so the second term in (20) related to $\Sigma_{\mathbf{x}}$ is set equal to zero). Now, in the limit $p \to 0$ we obtain

$$\lim_{p \to 0} \mathbf{A}^p = \text{diag}\left(\lambda v_i^{-1}\right), \tag{23}$$

which corresponds to a scaled version of (19). Therefore, the proposed formulation corresponds to the IRLS formulation with l_p-norms when $p \to 0$. Note, however, that as opposed to the IRLS approach, the proposed formulation results in a convex optimization problem, and therefore it is much more robust. Robustness is also an issue with methods utilizing non-convex image priors such as mixture of Gaussian priors [6].

4. EXPERIMENTAL RESULTS

In this section we present experimental results obtained by the use of the proposed algorithm and compare its performance with several state-of-the-art image restoration methods. Four algorithms are chosen for comparison: 1) The method in [9], which is based on a Gaussian image prior (denoted by *Gaussian*), 2) the method in [4] based on TV-priors (denoted by *TV*), 3) the method in [10] based on Student's t priors (denoted by *Student's t*), and 4) the method in [6] based on l_p-norm based image priors (denoted by *IRLS*). The first three methods are fully-automated Bayesian methods, i.e., they do not require parameter tuning. The results of all methods are reproduced from [10] except the *IRLS* method, which we have implemented and tuned its parameters to report its best performance. This method is based on l_p-norms where we set $p = 0.8$ as recommended by [6].

Table 1. ISNR values for the Lena, Cameraman and Shepp-Logan images degraded by a 9x9 uniform blur.

		Lena	Cameraman	Shepp-Logan
BSNR	Method	ISNR (dB)		
40dB	*Proposed*	8.28	9.10	**30.14**
	Gaussian	4.72	4.57	5.31
	TV	8.42	8.57	13.69
	Student's t	**8.49**	**9.17**	17.12
	IRLS	8.44	9.07	19.33
30dB	*Proposed*	**6.24**	**6.06**	**24.19**
	Gaussian	4.06	3.24	3.56
	TV	5.89	5.41	7.77
	Student's t	5.89	5.88	9.42
	IRLS	5.96	5.79	14.36
20dB	*Proposed*	**4.49**	**4.36**	**14.29**
	Gaussian	2.68	2.19	2.49
	TV	3.72	2.42	3.01
	Student's t	3.70	3.04	5.90
	IRLS	4.00	4.32	10.97

In the proposed algorithm, we utilized the first and second order horizontal and vertical derivative filters for $\mathbf{D}_1, \ldots, \mathbf{D}_4$, that is,

$$\begin{bmatrix} -1 & 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 & -2 & 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}, \tag{24}$$

and 45^o and -45^o derivative filters for $\mathbf{D}_5$ and $\mathbf{D}_6$, that is,

$$\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \tag{25}$$

More complex filters (such as curvelets and fan filters) will potentially lead to better reconstruction performances, whose investigation is left as future work.

We use the well-known "Lena" and "Cameraman" images and the "Shepp-Logan" phantom as our test images. Due to space limitations, we only report results with a 9×9 uniform blur point spread function, results with other kernels were very similar. White Gaussian noise is added to the blurred images to obtain blurred-signal-to-noise (BSNR) ratios of 20, 30, and 40dB. As the convergence criterion in the proposed algorithm, we used $\| \hat{\mathbf{x}}^k - \hat{\mathbf{x}}^{k-1} \|^2 / \| \hat{\mathbf{x}}^{k-1} \|^2 < 10^{-5}$, where $\hat{\mathbf{x}}^k$ is the image estimate at iteration k.

Table 1 shows the quantitative ISNR results, where ISNR is defined as $10 \log_{10}(\| \mathbf{x} - \mathbf{y} \|^2 / \| \mathbf{x} - \hat{\mathbf{x}} \|^2)$, where $\mathbf{x}$, $\mathbf{y}$ and $\hat{\mathbf{x}}$ are the original, observed, and estimated images, respectively. The first observation from Table 1 is that the method with the Gaussian image priors is outperformed in all cases by the other methods, as expected. Second, it is clear that the proposed method outperforms the other methods in 7 out of 9 cases, whereas in the other cases it provides ISNR results very close to the best-performing method *Student's t*. Note, however, that this method incorporates more complex curvelet filters than the ones incorporated in this work.

A particularly interesting result is with the Shepp-Logan phantom, where the proposed algorithm and *IRLS* provide much higher ISNR results than the other methods. This is due to the fact that the Shepp-Logan phantom can be represented with a highly sparse edge structure, and these methods enforce sparsity in the filter outputs to a much higher extent than the other methods. Moreover, it is clear that the proposed algorithm is more effective in enforcing

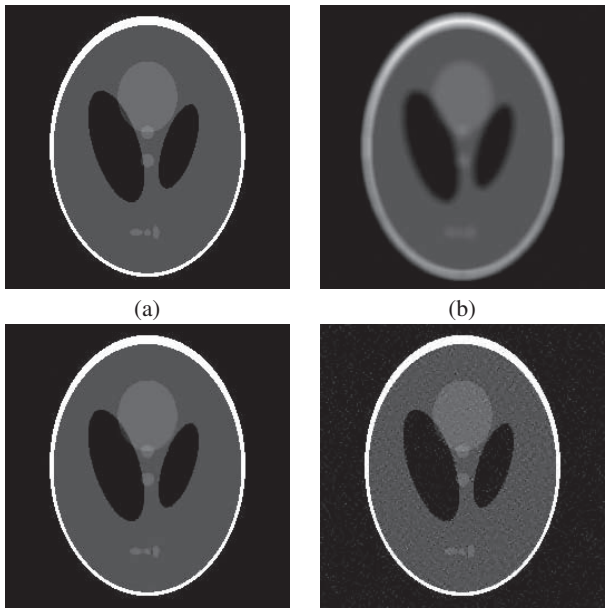

Fig. 1. (a) Original Shepp-Logan phantom, (b) Image degraded by a 9×9 uniform PSF and BSNR=30dB, (c) Restored image using the proposed algorithm (ISNR = 24.19dB), (d) Restored image using the IRLS method with $p = 0.8$ (ISNR = 14.36dB).

Fig. 2. (a) Original Lena image, (b) Image degraded by a 9×9 uniform PSF and BSNR=30dB, (c) Restored image using the proposed algorithm (ISNR = 6.24dB), (d) Restored image using the IRLS method with $p = 0.8$ (ISNR = 5.96dB).

sparsity than IRLS, as the ISNR results in all noise levels demonstrate. On the other hand, with the more realistic images "Lena" and "Cameraman", the proposed algorithm still provides higher restoration performance than IRLS.

The original, degraded, and restored images for the Shepp-Logan phantom are shown in Fig. (1), and for the Lena image in Fig. (2) both for the BSNR=30dB case. For the IRLS method, the image corresponding to the the maximum ISNR achieved during the iterations is shown. On the other hand, the parameters of the proposed method are automatically calculated and the images obtained at the convergence are shown. Despite full-automation, in both cases the proposed method provided restored images with higher visual quality.

Finally, the computational complexity of the proposed algorithm is similar to (and even lower than some of) the other algorithms. Moreover, the computational complexity of the algorithm can be significantly improved with a straightforward application of operator splitting schemes as in [2].

5. CONCLUSIONS

In this paper we presented a novel Bayesian algorithm for image restoration and parameter estimation using independent Gaussian image priors placed on high-pass filter outputs of the image. We adopt a fully-Bayesian analysis to jointly estimate the original image and the unknown hyperparameters (including the observation noise) such that no parameter tuning is necessary. We have shown that the proposed formulation is a special case of the non-convex l_p-norm based image priors with $p = 0$, but it results in a convex optimization problem. Finally, we demonstrated with experimental results that the proposed algorithm provides higher restoration performance than state-of-the-art algorithms in most cases. Future work includes utilizing more complex high-pass filters and extending the framework to blind deconvolution.

6. REFERENCES

[1] A.K. Katsaggelos, Ed., *Digital Image Restoration*, Springer-Verlag, 1991.

[2] D. Krishnan and R. Fergus, "Fast image deconvolution using hyper-Laplacian priors," in *NIPS 2009*. MIT Press.

[3] M.A.T. Figueiredo, J.M. Bioucas-Dias, and R.D. Nowak, "Majorization-minimization algorithms for wavelet-based image restoration," *IEEE Trans. Image Processing*, vol. 16, no. 12, pp. 2980–2991, Dec. 2007.

[4] S. D. Babacan, R. Molina, and A.K. Katsaggelos, "Parameter estimation in TV image restoration using variational distribution approximation," *IEEE Trans. Image Processing*, vol. 17, no. 3, pp. 326–339, March 2008.

[5] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W.T. Freeman, "Removing camera shake from a single photograph," *SIGGRAPH 2006*, vol. 25, pp. 787–794.

[6] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," in *SIGGRAPH 2007*, p. 70.

[7] M. Elad, M. A.T. Figueiredo, and Y. Ma, "On the role of sparse and redundant representation in image processing," *to appear in the Proceedings of the IEEE*, 2010.

[8] M. Tipping, "Sparse Bayesian learning and the relevance vector machine," *JMLR*, pp. 211–244, 2001.

[9] R. Molina, "On the hierarchical Bayesian approach to image restoration. Applications to Astronomical images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, no. 11, pp. 1122–1128, Nov. 1994.

[10] G. Chantas, N. Galatsanos, A. Likas, and M. Saunders, "Variational Bayesian image restoration based on a product of t -distributions image prior," *IEEE Trans. Image Processing*, vol. 17, no. 10, pp. 1795–1805, Oct. 2008.