

Mutual Information Regularized Bayesian Framework for Multiple Image Restoration

Yunqiang Chen, Hongcheng Wang, Tong Fang and Jason Tyan
Siemens Corporate Research,
755 College Rd. East, Princeton, NJ 08540

Abstract

Bayesian methods have been extensively used in various applications. However, there are two intrinsic issues rarely addressed, namely generalization and validity. In the context of multiple image restoration, we show that traditional Bayesian methods are sensitive to model errors and cannot guarantee valid results satisfying the underlying prior knowledge, e.g. independent noise property. To improve the Bayesian framework's generalization, we propose to explicitly enforce the validity of the result. Independent noise prior is very important but largely under-utilized in previous literature. In this paper, we use mutual information (MI) to explicitly enforce the independence. Efficient approximations based on Taylor expansion are proposed to adapt MI into standard energy forms to regularize the Bayesian methods. The new regularized Bayesian framework effectively utilizes the traditional generative signal/noise models but is much more robust to various model errors, as demonstrated in experiments on some demanding imaging applications.

1. Introduction

The Bayesian framework has been widely used in many different applications, ranging from image processing (e.g. image restoration, stereo/motion estimation or segmentation [4, 19]) to machine learning (e.g. hidden Markov model [13] or graphical models [3]) and much more. In this paper, we mainly study it in the context of multiple image restoration. But the problems addressed in this paper are quite general to many traditional Bayesian methods and our solution has great potential to extend to other applications.

In image restoration, numerous methods (e.g. Wiener filter, steerable filters, Wavelet) have been proposed to improve image quality and reduce imaging noise. Recently, Bayesian restoration schemes have been extensively explored (e.g. [1, 9, 11]) to find the *maximum a posteriori* (MAP) estimation of the true signal based on the statistical noise/signal models (i.e. generative models). Very good

results have been achieved when accurate models can be trained in advance. It can be formulated as:

$$\hat{S} = \arg \max_S P(S|I) = \arg \max_S P(I|S)P(S) \quad (1)$$

where I is the observed noisy image and S is the true image we want to recover. $P(S)$ models the expected structures in the true image (e.g. smooth surfaces, step edges or corners). $P(I|S)$ is the conditional distribution of the observed image I given the true image S . It models how the observed image is generated, usually including point spread functions and noise models. The MAP estimation is obtained by finding \hat{S} that has the maximum probability: $P(\hat{S}|I)$.

Multiple images can be obtained in some cases to further improve imaging quality (e.g. ultrasound spatial/frequency compounding [14, 15] or multi-spectral remote sensing [11]). It is rather straight forward to extend the Bayesian framework to multiple images when the images are corrupted by independent noise. A basic multi-image formation model and the associated Bayesian restoration method are discussed in Section 2.

Most research has been focused on improving the signal/noise models (e.g. MRF based edge modeling [9, 11], Wavelet [1] or AQua model [12]) to better discriminate the signal and noise. Better optimization procedures (e.g. [5, 20]) are also designed. For multi-image restoration, the correlations between the signal in different images are exploited. For example, a coupled edge modeling on multi-image is proposed in [11] to achieve better edge detection and hence better edge preservation during noise reduction.

However, few have addressed the two intrinsic issues of the widely used Bayesian framework, namely *generalization* and *validity*, especially when the signal/noise models are not accurate. In real world applications, it could be very difficult to obtain accurate prior models. For example, the ultrasound speckle noise is non-stationary and changes according to ultrasound attenuation and the sub-resolution scatterers in the tissue [17]. Various types of structures (e.g. corners, edges or surfaces) are also hard to model accurately. To make things worse, over-simplified assumptions (e.g. the noise being independent of the signal) sometimes have to be made to allow tractable solutions. Under

these difficult conditions, the traditional Bayesian framework cannot provide robust results and some of the underlying prior knowledge/constraints may even be violated.

One of the important priors in multi-image restoration is that the multiple images are often corrupted by independent noise (e.g. [2, 11, 15]), which is the very basis for most Bayesian restoration methods to factorize the joint probabilities and hence allows tractable solutions. However, this prior is also the most under-utilized and often violated one when there are modeling errors. A detailed example is shown in Section 2.2, where an inaccurate weighting between noise and signal models results in significantly correlated noise. Such violations of the prior constraints indicate restoration errors and sensitivity to inaccurate models of the traditional Bayesian framework. In machine learning, Support Vector Machine (SVM) is proposed in [16] to replace the generative model based Bayesian methods for better generalization. In image restoration, however, generative models of signal and noise have been extensively studied and are essential for discriminating noise and signal. It is important not to forfeit those generative models.

To achieve better generalization, we propose to analyze the validity of the restoration results against the prior knowledge, not only the generative models $P(S)$ and $P(I_i|S)$, but also the priors behind the generative models, e.g. independent noise constraint. Independence analysis has achieved great success in many different applications (e.g. [8]), but it is rarely emphasized in multi-image restoration methods. ICA ([8]) is widely used for blind source separation. But it is not easy to extend to multiple image restoration where we have N observed images but more than N sources (noise and signal) to recover. Mutual information (MI) is a natural way to measure the (in)dependencies. Maximization of the mutual information has been extensively used to exploit the dependencies in stereopsis and image registration [7]. In our case, however, we should minimize the mutual information to enforce independence. More attention needs to be paid to the optimization methods and convergence issues because there are almost infinite solutions that are independent of a given random signal.

In this paper, we incorporate mutual information in the traditional MAP framework as an extra energy term to regularize the restoration process. Some approximations are proposed to adapt the mutual information into a standard energy term for effective optimization. In the new regularized Bayesian framework, we fully exploit both the generative signal/noise models and the independent noise constraint. The mutual information term assures the validity of the restoration results and robustness to inaccurate models while the generative models stabilize the process of minimizing mutual information. The experiments and comparisons in some difficult imaging applications demonstrate the robustness of the new regularized Bayesian framework, es-

pecially when accurate models are not readily available.

The rest of the paper is organized as follows. In Section 2, we describe our multi-image formation model and the new regularized Bayesian restoration framework. In Section 3, approximations of the mutual information term and the optimization procedure are derived. In Section 4, we report promising results and comparisons with traditional methods on some difficult medical imaging applications. Conclusions are given in Section 5.

2. Mutual information regularized multiple image restoration

2.1. Problem formulation

Multiple images or multi-channel images can greatly improve imaging quality. For example, some ultrasound B-scans can generate multiple images of the same tissue using ultrasound of different frequencies, which is also called compounding [14, 15]. The multiple images have the same underlying signal but independent speckle noise [15, 18]. Unlike the multi-image based super-resolution techniques [10], these multiple images do not provide extra information for super-resolution but can help reducing the imaging noise dramatically. Because the goal is to reduce noise instead of correcting imaging defects (e.g. out of focus or motion blur), we ignore the point spread function and use the following simple image formation model:

$$I_i = S + N_i, \quad i \in [1, M] \quad (2)$$

S is the underlying structure (i.e. the true image) and I_i is one of the M observed images, corrupted by additive noise N_i . N_i and N_j are independent when $i \neq j$. This image formation model is the simplest case in multiple image based restoration. It assumes that all images are aligned and have the same signal in them. Although simple, it is suitable for many applications and allows us to focus on the Bayesian framework and its underlying problems. For more difficult imaging processes (e.g. multi-spectral images), more complex image formation models (e.g. [11]) are necessary but the new ideas proposed in this paper are still applicable.

Image restoration based on Eq. (2) is largely overlooked and handled with simple weighted averaging (e.g. [14, 15]), followed by the traditional restoration methods on the average image. In [11], coupled edge estimation is proposed for multi-channel images to achieve more accurate edge detection and hence better restoration. Under Eq. (2), the coupled edge detection scheme is also reduced to averaging-and-filtering. As shown in next subsection, this is actually the optimal solution based on the traditional Bayesian framework. The noise is reduced by the averaging process and better edge detection and restoration can be achieved by working on the average image. However, we can easily see

that by retaining only the average image for further filtering, this scheme is obviously not optimal and fails to fully utilize the abundant information in the multiple images.

To illustrate the new multi-image restoration algorithm, we start with two images. In subsection 2.2, we extend the Bayesian method to the formation model in Eq. (2) and point out its underlying problems. In subsection 2.3, independence analysis based on mutual information is introduced for better generalization. The regularized Bayesian framework is explained in subsection 2.4.

2.2. Traditional Bayesian multi-image restoration

The traditional Bayesian framework can be easily extended to two images based on Eq. (2). With the knowledge that S , N_1 and N_2 are independent, we have:

$$\begin{aligned} P(S|I_1, I_2) &= c \cdot P(I_1, I_2|S)P(S) \\ &= c \cdot P_{N_1}(I_1 - S)P_{N_2}(I_2 - S)P_S(S) \end{aligned} \quad (3)$$

where c is a normalization constant. $P_S()$ and $P_{N_i}()$ are the prior signal/noise models respectively. For simplicity, we do not include edge modeling for the signal at this stage, which will not change the conclusion that the optimal solution is average-and-filtering anyway. Based on Gaussian noise model and smooth signal constraint, we can define the cost function $C(\hat{S})$ of the MAP estimation as:

$$\begin{aligned} \hat{S} &= \arg \min_{\hat{S}} C(\hat{S}) = \arg \min_{\hat{S}} (-\log(P(\hat{S}|I_1, I_2))) \quad (4) \\ &= \arg \min_{\hat{S}} (\lambda_1 (I_1 - \hat{S})^2 + \lambda_2 (I_2 - \hat{S})^2 + \lambda_3 (\hat{S} - \bar{S})^2) \end{aligned}$$

where \bar{S} is the average intensity of the true signal in a small neighborhood and λ_i is the weighting of each constraint. The first two terms enforce that the estimated image should look like the observed images. The 3rd term models the signal properties and prefers smooth signal. More complex models can be used to represent the signal properties (e.g. [1, 9, 11, 12]). The MAP estimation can be obtained when the derivative of $C(\hat{S})$ is equal to zero:

$$\frac{\partial C(\hat{S})}{2 \cdot \partial \hat{S}} = \lambda_1 (\hat{S} - I_1) + \lambda_2 (\hat{S} - I_2) + \lambda_3 (\hat{S} - \bar{S}) = 0 \quad (5)$$

Because the \bar{S} is unknown, $C(\hat{S})$ can be optimized in an iterative manner. At each iteration, \bar{S} is estimated based on the solution of previous iteration $\hat{S}^{(k)}$. The iterative minimization process is as follows:

$$\hat{S}^{(k+1)} = \frac{(\lambda_1 + \lambda_2)I_{avg} + \lambda_3 \bar{S}^{(k)}}{\lambda_1 + \lambda_2 + \lambda_3} \quad (6)$$

where $I_{avg} = (\lambda_1 I_1 + \lambda_2 I_2) / (\lambda_1 + \lambda_2)$. We can see, in the traditional Bayesian method, only the average image is needed for the MAP estimation and all the original images can be discarded after averaging. Apparently, it does not fully utilize all the information in the multiple images.

The problems of the Bayesian framework become apparent when there are model errors, which include oversimplified assumptions, inaccurate models, inappropriate weighting or convergence issues. For example, if the weighting of the smooth signal constraint is set too weak (e.g. $\lambda_1 = \lambda_2 \gg \lambda_3$) in Eq. (4), the MAP estimation of the signal is $\hat{S} \approx (I_1 + I_2)/2$ and the residual error (i.e. $\hat{N}_i = I_i - \hat{S}$) on each image belongs to noise, which results in $\hat{N}_1 \approx -\hat{N}_2$. We can see that, even though the independence between N_1 and N_2 is the basis in deriving the MAP framework, it is severely violated in the final results (i.e. highly correlated \hat{N}_1 and \hat{N}_2) due to weighting errors.

From this example, we can clearly see that the traditional Bayesian framework relies solely on the generative models and is sensitive to model errors. The final result might not even satisfy the underlying priors in the generative models (e.g. independent noise). How to improve the generalization performance of the traditional Bayesian framework instead of relying on perfect training processes for more accurate models is the main focus of this paper.

2.3. Generalization and independence analysis

In the image formation model in Eq. (2), the same underlying image is corrupted by independent noise, which allows us to factorize $P(I_1, I_2|S)$ for much simpler noise modeling and optimization procedures. However, the independence constraint is usually under-utilized and often violated when there are model errors as shown in previous subsection. Most previous work focuses on more accurate models $P(S)$ and $P(I|S)$ (e.g. [1, 9, 11, 12]). We address it from generalization perspective and propose to explicitly verify the restoration results against the independence constraint. Independence analysis is a very powerful tool for studying the multiple sources and has achieved surprisingly good results in many applications, e.g. [8]. But it is rarely emphasized in image restoration literature.

In the context of multiple image restoration, we can easily see the importance of the independence analysis. When we obtain an estimation of the ‘true’ signal \hat{S} , we also have the estimation of the noise components $\hat{N}_i = I_i - \hat{S}$, which should be independent of the noise on other images. When there are restoration errors (i.e. $S - \hat{S}$), it appears uniformly in the estimated noise \hat{N}_i on all images:

$$\hat{N}_i = I_i - \hat{S} = N_i + (S - \hat{S}) \quad (7)$$

Now, we can see that the generalization and validity are highly related. Inaccurate models cause restoration errors (i.e. $(S - \hat{S})$), which is common for all the estimated noise. It will increase the dependencies between the estimated noise and hence violate the independent noise constraint. Violations of the prior constraints, on the other hand, indicate restoration errors. By regularizing the restoration process explicitly with the prior constraint, we should be able

to improve the results and the generalization performance of the traditional Bayesian framework. Independence analysis can be used for this purpose. Mutual information is an elegant way to measure the independence:

$$MI(\hat{N}_1, \hat{N}_2) = \sum_{\hat{N}_1} \sum_{\hat{N}_2} p(\hat{N}_1, \hat{N}_2) \log \frac{p(\hat{N}_1, \hat{N}_2)}{p(\hat{N}_1)p(\hat{N}_2)} \quad (8)$$

It is worth noting that correlation $\rho(\hat{N}_1, \hat{N}_2)$ is a simpler way to detect the dependencies and is easier to evaluate than mutual information. But it is not suitable here because the estimation \hat{S} is usually dependent on N_i and hence $\rho(\hat{N}_1, \hat{N}_2)$ is not proportional to $\|S - \hat{S}\|^2$. A simple example could be that, for each pixel p , \hat{S}_p randomly takes the value of either I_1 or I_2 , which will guarantee $\rho(\hat{N}_1, \hat{N}_2) = 0$. But obviously \hat{N}_1 and \hat{N}_2 are not independent in this case.

2.4. MI regularized Bayesian restoration

To explicitly enforce the independent noise constraint, we propose to incorporate mutual information into the traditional Bayesian framework as a regularization term. The new regularized Bayesian restoration can be described as:

$$\hat{S} = \arg \min_{\hat{S}} C(\hat{S}) = \arg \min_{\hat{S}} (\lambda_0 (\hat{S} - \bar{S})^2 + \lambda_1 (I_1 - \hat{S})^2 + \lambda_2 (I_2 - \hat{S})^2 + \lambda_3 MI(\hat{N}_1, \hat{N}_2)) \quad (9)$$

As we can see, if the Bayesian estimation has accurate models and provides a very good restoration result, the residual errors on different images satisfy the independence constraint and the new mutual information term is at its minimal value 0 as well. But if the Bayesian estimation is wrong due to either inaccurate models or converging to a local minimum, this new regularization term will penalize the solution and make it less likely to happen.

Mutual information is very expensive to evaluate and makes the objective function extremely difficult to optimize. Some approximations similar to [7] are proposed to simplify the computation dramatically in Section 3 with an iterative optimization for the regularized Bayesian method.

When more than two images are considered, the noise components on different images are mutually independent. To enforce this constraint, we can model it by pairwise mutual information:

$$MI(\hat{N}_1, \hat{N}_2, \dots, \hat{N}_M) = \sum_{i=1}^{M-1} \sum_{j=i+1}^M MI(\hat{N}_i, \hat{N}_j) \quad (10)$$

Based on this equation, the previous two-image based restoration scheme can be easily extended to handle more images as follows:

$$\hat{S} = \arg \min_{\hat{S}} C(\hat{S}) = \arg \min_{\hat{S}} \sum_{i=1}^M \lambda_i (I_i - \hat{S})^2 + \lambda \sum_{i=1}^{M-1} \sum_{j=i+1}^M MI(\hat{N}_i, \hat{N}_j) + \lambda_0 (\hat{S} - \bar{S})^2 \quad (11)$$

3. Optimization and Implementation

To find the optimal solution of the new mutual information regularized objective function in Eq. (9), we exploit the standard iterative method. The difficulty lies in the mutual information term, which is computationally prohibitive to optimize. Some approximations of the mutual information term based on Taylor expansion are derived in this section for efficient optimization. The proposed approximations can be combined with other type of optimization procedures as well, e.g. the Graph Cut method in [7].

For the standard iterative optimization procedure, we first obtain an average of the multiple images, followed by a low-pass filtering. This low-pass filtered average image is used as the initial estimation (i.e. $\hat{S}^{(0)}$). At iteration $k+1$, we update the estimated signal for each pixel p (i.e. $\hat{S}_p^{(k+1)}$) based on the result of previous iteration (i.e. $\hat{S}^{(k)}$) so that the objective function in Eq. (9) is minimized. This process continues until converging.

The challenge in the optimization is the complexity of calculating $MI(\hat{N}_1, \hat{N}_2)$, where $\hat{N}_1 = (I_1 - \hat{S}^{(k+1)})$ and $\hat{N}_2 = (I_2 - \hat{S}^{(k+1)})$. The MI term can be expressed by the entropy $H(\hat{N}_1)$, $H(\hat{N}_2)$, and the joint entropy $H(\hat{N}_1, \hat{N}_2)$:

$$\begin{aligned} MI(\hat{N}_1, \hat{N}_2) &= H(\hat{N}_1) + H(\hat{N}_2) - H(\hat{N}_1, \hat{N}_2) \quad (12) \\ &= - \sum_z P_1(z) \log P_1(z) - \sum_z P_2(z) \log P_2(z) \\ &\quad + \sum_{z_1, z_2} P_{1,2}(z_1, z_2) \log P_{1,2}(z_1, z_2) \end{aligned}$$

where $P_1(z)$, $P_2(z)$ and $P_{1,2}(z_1, z_2)$ are the probability distribution of \hat{N}_1 , \hat{N}_2 and their joint distribution respectively. Because \hat{N}_1 and \hat{N}_2 are usually different from the true N_1 and N_2 during the iterations, these probabilities should be estimated from samples. For non-stationary noise, it is better to estimate them within a small neighborhood (e.g. 15×15 window in our experiments) centered at each pixel p . Parzen Window is used to approximate the probability by a superposition of Gaussian densities:

$$\begin{aligned} P_1(z) &= \frac{1}{|D|} \sum_p G_\psi(z - n_{1,p}) \quad (13) \\ P_2(z) &= \frac{1}{|D|} \sum_p G_\psi(z - n_{2,p}) \\ P_{1,2}(z) &= \frac{1}{|D|} \sum_p G_\psi[(z_1, z_2) - (n_{1,p}, n_{2,p})] \end{aligned}$$

where, $G_\psi(z) = \frac{1}{\sqrt{(2\pi)^2 \psi}} \exp(-\frac{1}{2} z^T \psi^{-1} z)$ and $|D|$ is the size of the samples. $n_{1,p}$ and $n_{2,p}$ are the noise components in image I_1 and I_2 at position p respectively.

By updating the signal estimation at pixel p to $\hat{S}_p^{(k+1)}$, we have the new residual errors: $n_{1,p} = I_{1,p} - \hat{S}_p^{(k+1)}$ and

$n_{2,p} = I_{2,p} - \hat{S}_p^{(k+1)}$. The $P_1(z)$, $P_2(z)$ and $P_{1,2}(z_1, z_2)$ need to be re-estimated. Finding the best update to minimize $MI(\hat{N}_1, \hat{N}_2)$ requires re-estimating these probabilities for each possible value of $\hat{S}_p^{(k+1)}$ and hence computationally prohibitive.

For efficient optimization, we use the Taylor expansion to approximate the mutual information similar to [7]. The idea is to approximate the complex mutual information with summation operations. Then the computational cost will decrease dramatically and most traditional optimization techniques can be applied as well. Unlike [7], we need to estimate mutual information locally instead of using the whole image since the signal and noise distributions may vary a lot through the whole image. Since $P_1(z)$ depends on $\hat{S}_p^{(k+1)}$, we use the Taylor expansion, $x \log x = -x_0 + (1 + \log x_0)x + O((x - x_0)^2)$, to approximate $H(\hat{N}_1)$ as:

$$\begin{aligned} H(\hat{N}_1) &= - \sum_z P_1(z) \log P_1(z) \\ &\cong \sum_z [P_1^0(z) - (1 + \log(P_1^0(z))) \cdot P_1(z)] \\ &= \sum_z P_1^0(z) - \sum_z P_1(z) - \sum_z \log(P_1^0(z)) \cdot P_1(z) \\ &= - \sum_z \log(P_1^0(z)) \cdot P_1(z) \end{aligned} \quad (14)$$

where $P_1^0(z)$ is the distribution of the estimated noise of previous iteration and $P_1(z)$ is the new distribution after update. From Eq. (13) and (14), we obtain $H(\hat{N}_i)$:

$$\begin{aligned} H(\hat{N}_i) &\approx - \frac{1}{|D|} \sum_p \sum_z \log(P_i^0(z)) \cdot G_\psi(z - n_{i,p}) \\ &= - \frac{1}{|D|} \sum_p f_i(n_{i,p}) \end{aligned} \quad (15)$$

where $f_i(z) = \log(P_i^0(z)) \otimes G_\psi(z)$ and ‘ \otimes ’ means convolution. Similarly, we can approximate the joint probability term in Eq. (12) as follows:

$$\begin{aligned} H(\hat{N}_1, \hat{N}_2) &\approx - \frac{1}{|D|} \sum_p \sum_{z_1} \sum_{z_2} \log(P_{1,2}^0(z_1, z_2)) \\ &\quad \cdot G_\psi((z_1, z_2) - (n_{1,p}, n_{2,p})) \\ &= - \frac{1}{|D|} \sum_p f_{1,2}(n_{1,p}, n_{2,p}) \end{aligned} \quad (16)$$

where $f_{1,2}(z_1, z_2) = \log(P_{1,2}^0(z_1, z_2)) \otimes G_\psi(z_1, z_2)$. Based on these approximations, we can find the cost value in Eq. (9) efficiently without re-estimating the probability density for every possible value. For iteration $k + 1$, we first estimate the $P_1^0(z)$, $P_2^0(z)$ and $P_{1,2}^0(z_1, z_2)$ based on $\hat{S}^{(k)}$. Then we convolute them with Gaussian kernels $G_\psi(z)$ and

$G_\psi(z_1, z_2)$ to obtain $f_1(z)$, $f_2(z)$ and $f_{1,2}(z_1, z_2)$. For any given update $\hat{S}_p^{(k+1)}$, we can easily calculate the cost of the MI term by looking up the corresponding value in the $f()$ and find the one maximizing the objective function:

$$\begin{aligned} \hat{S}_p^{(k+1)} &= \arg \min_{\hat{S}} (\lambda_0 (\hat{S}_p - \tilde{S}_p^{(k)})^2 + \lambda_1 (n_{1,p})^2 \\ &\quad + \lambda_2 (n_{2,p})^2 + \lambda_3 (f_{1,2}(n_{1,p}, n_{2,p}) - f_1(n_{1,p}) - f_2(n_{2,p})) \end{aligned} \quad (17)$$

The iterative optimization algorithm for two image restoration is described in Algorithm 1.

Algorithm 1: Image Restoration Using Two Images

Data: Given images I_1, I_2 with independent noise.

Result: Restored image \hat{S} .

Initialize the estimated image using the low-pass filtered average image, i.e., $\hat{S}^0 = G_\sigma \otimes (\frac{I_1 + I_2}{2})$;

while \sim stop **do**

for each pixel p **do**

 Within a neighborhood $w(p)$, centered around pixel p , estimate the $P_1^0(z)$, $P_2^0(z)$ and $P_{1,2}^0(z_1, z_2)$ based on $\hat{S}_p^{(k)}$ and obtain the corresponding $f_1(z)$, $f_2(z)$ and $f_{1,2}(z_1, z_2)$;

 For a search range $-m \leq \delta \leq m$ for each pixel, each term in Eq. (17) is computed corresponding to $\hat{S}_p^k + \delta$. The optimal update is obtained as:

$$\hat{S}_p^{k+1} = \arg \min_{-m \leq \delta \leq m} C(\hat{S}_p^k + \delta).$$

end

 If $C(\hat{S}^{k+1}) - C(\hat{S}^k) < \epsilon$, stop. Otherwise, $\hat{S}^k = \hat{S}^{k+1}$.

end

4. Experiments

To validate the efficacy and robustness of the new regularized Bayesian framework, we apply it to some difficult imaging applications, where accurate signal/noise models are difficult to obtain. The regularized objective function in Eq. (11) is used in all experiments and optimized by the steps in Algorithm 1. Please note that Eq. (11) only uses simple noise models and smooth signal constraint with no modeling of edges or corners. When regularized by the new mutual information term, it provides surprisingly sharp restoration results in the experiments.

For comparison, we use the traditional Bayesian methods. Without explicit edge modeling, traditional methods significantly blur the sharp boundaries. For reasonable re-

sults, we have to add edge modeling as in [9, 11]:

$$\hat{S} = \arg \min_{\hat{S}} \sum_p (\lambda_1 (I_{1,p} - \hat{S}_p)^2 + \lambda_2 (I_{2,p} - \hat{S}_p)^2 + \lambda_3 (1 - e_p) (\hat{S}_p - \bar{S}_p)^2 + \lambda_4 e_p) \quad (18)$$

where e_p is the edge process which is 0 or 1. When $e_p = 1$, the smoothness constraint is disabled and hence preserves the sharp edges. To prevent all the pixels being treated as edges, there is a penalty λ_4 for $e_p = 1$. Restoration quality depends heavily on the accuracy of the edge estimation.

4.1. Ultrasound frequency compounding

We first test the new algorithm in ultrasound frequency compounding. Three images are scanned using ultrasound of different frequencies. Because the signal has been taken $\log()$ before converted to display images, the multiplicative speckle noise in acoustic signal is now additive. One of the three compounding images is shown in Fig. 1 (a). The main difficulty in ultrasound imaging is the non-stationary speckle noise due to the sub-resolution scatters, ultrasound attenuation, beam forming and the fan-shape scanning. Usually, the speckle noise has larger variance and more lateral correlation at the far end of the scanning, which makes it more difficult to be discriminated from the edges. Fortunately, the independent noise property still holds [15, 18] and forms the basis of ultrasound compounding techniques.

The traditional Bayesian method is very sensitive to the inaccurate noise model. Fig. 1 (b) shows the result of the traditional method described in Eq. (18). As we can see, strong edges can be detected and preserved well by the edge modeling. But due to the stationary noise assumption in Eq. (18), at near end where noise is weaker than modeled, some weak edges that can be detected are missed and severely blurred. While in some other regions, the noise stronger than modeled is not reduced enough. We can still see some ‘worm’ patterns especially at the far end.

The regularized Bayesian method (as shown in Eq. (11)) exploits no extra information than the traditional Bayesian methods and lacks explicit edge modeling. But by regularizing the result with the independence constraint, it achieves much better result as shown in Fig. 1 (d). The importance of the MI penalty term can be clearly seen in Fig. 1 (c), where white means large penalty. At the first iteration, the restoration results (\hat{S}^0) is initialized to be the low-pass-filtered average image, which severely blurs the sharp boundaries. The MI term successfully captures those restoration errors with very few false positives in spite of the hard-to-model speckle noise. With this MI penalty term, our new algorithm successfully prevents the blurring across the edge or corners in Fig. 1 (d). Even the weakest edges are also detected and preserved much better, while the noise regions (especially the far end) are smoother.

4.2. CT image reconstruction

We also test our algorithm in an interesting extension for Computed Tomography (CT). Filtered back-projection (FBP) is widely used in most commercial CT machines for its efficiency but suffers streaking artifacts [6]. Image restoration techniques can significantly improve the FBP’s quality. We test our algorithm on the ‘Herman head’ phantom generated by CTSim software (<http://www.ctsim.org/>) as shown in Fig. 2 (a). We generate projections of the phantom and corrupt them with zero mean white Gaussian noise (variance is 1). Five different images are obtained by FBP (iradon function in Matlab). All the images have the same structure but the noise are independent. The back-projection produces streaking type of noise, which is very difficult to be distinguished from true edges in the image.

The average of the five images is shown in Fig. 2 (b). The multi-image restoration technique is applied to improve the image quality. Both traditional Bayesian method and our new algorithm significantly improve the quality. However, the traditional Bayesian method needs much more specific tuning. Simple edge modeling in Eq. 18 cannot handle the strong streaking noise. To obtain a good result, we have to tune the edge detector manually. From original phantom image, we know where are the structures and try the sobel edge detector with different size and different threshold to detect them in the average image. The 7x7 sobel filter with threshold 35 detects the structures reasonably well with least false detection (Fig. 2 (c)). The restoration result shown in Fig. 2 (d) preserves most structures but has some significant streaking noise due to the false edge detection.

The regularized Bayesian method requires no specific tuning. Initialized with the low-pass filtered average image, the MI term successfully detects the restoration errors (i.e. the blurred structures) with no false detection as shown in Fig. 2 (e). The MI term is estimated within a neighborhood and hence more blurred than the sobel edge detector in Fig. 2 (c). But it does not affect the sharpness of the restored structures because the MI term will be reduced only when the restoration error within the neighborhood is minimized. In the final result in Fig. 2 (f), all the streaking noise is removed while the structures are well preserved.

For a quantitative analysis, the results from different methods are compared with the original phantom. The average sum of squared errors per pixel is shown as follows:

Method	Average FBP	MAP	MI Regularized
Error	464.15	10.68	8.75

We can see that, even with hand tuned edge detector, the traditional Bayesian method does not handle non-stationary streaking noise as well as the new regularized Bayesian framework. In the experiments, the new regularized framework is very robust to both inaccurate signal modeling and non-stationary noise.

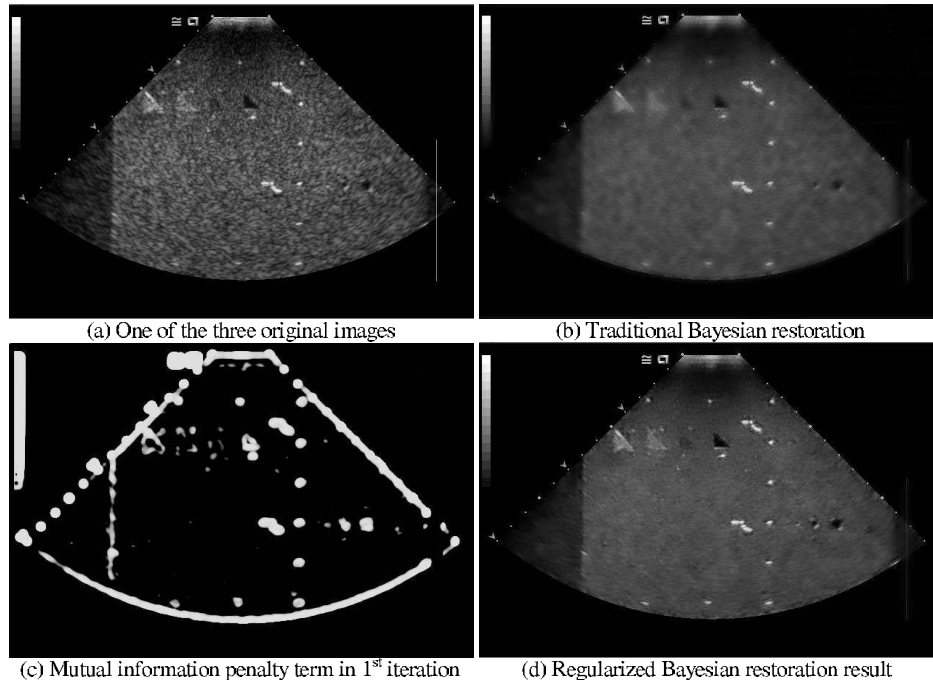


Figure 1. Ultrasound frequency compounding: (a) *The original B-Scans have non-stationary noise from the near field to the far end;* (b) *The traditional Bayesian restoration preserves the strong edges but fails on the weaker ones;* (c) *The MI penalty term (white means large penalty) captures even the weakest edges with very few false positives;* (d) *The final result of our new method preserves even the weakest structures and reduces the noise more.*

5. Conclusion

In this paper, we study the widely used Bayesian framework in the context of multiple image restoration and address two fundamental issues: the generalization ability and the validity of the MAP estimation. The traditional Bayesian framework relies solely on the generative models and cannot assure valid results in the presence of model errors. We propose to further regularize the result with the prior knowledge behind the generative models (e.g. independence between noise), which could dramatically improve the generalization ability of the Bayesian framework. Even in some very demanding imaging applications, the new regularized Bayesian method successfully recovers weak and complex structures (e.g. edge and corners) without explicit modeling of them.

We also show that the independent noise property is a powerful prior but largely neglected in previous research. It not only justifies the factorization of the joint probabilities for tractable solutions as in most previous literature, but also helps further as a powerful way to correct restoration errors. Mutual information based regularization terms are proposed to explicitly enforce the independence constraint in our new framework. Based on Taylor expansion, we adapt the mu-

tual information to standard energy terms and implement an efficient optimization procedure.

Furthermore, the new framework has great potential of extension to more complex image formation models (e.g. the multi-channel images or multi-spectral remote sensing images). There are also other types of prior knowledge/constraints (e.g. zero mean noise) which cannot be guaranteed by traditional Bayesian frameworks and worth further studying. We are also studying the possibility of integrating various types of generative signal models, e.g. Wavelet [1] or AQua model [12], into our new framework for even better results.

6. Acknowledgements

The authors would like to thank Prof. Narendra Ahuja for valuable suggestions and discussions.

References

- [1] M. Belge, M. Kilmer, and E. Miller. Wavelet domain image restoration with adaptive edge-preserving regularization. *IEEE Trans. Image Processing*, 9(4):598–608, 2000.

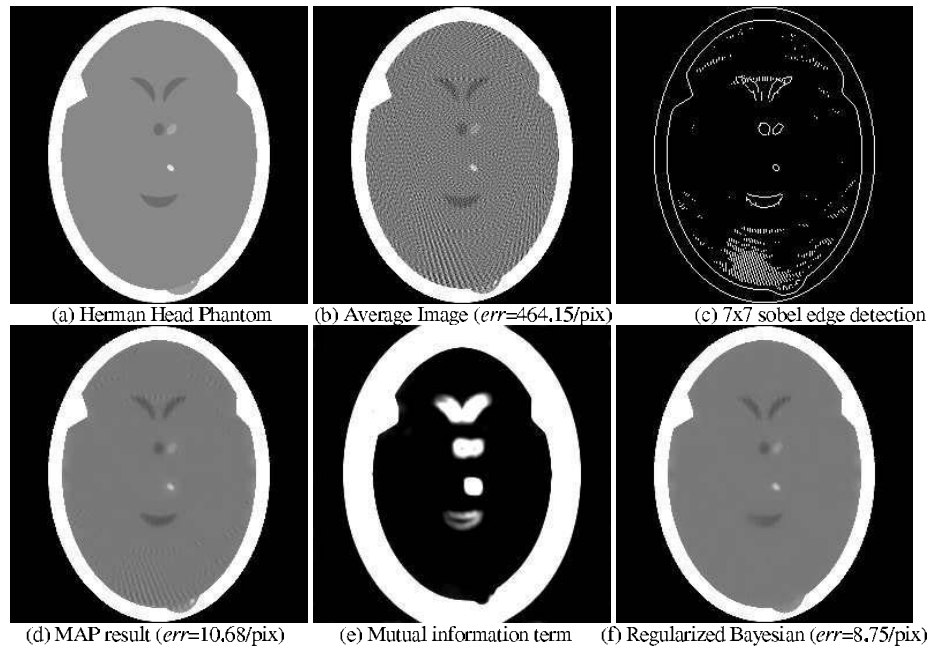


Figure 2. CT image reconstruction: (a) The original phantom; (b) Average of the 5 FBP results; (c) 7x7 sobel edge detection with threshold 35; (d) MAP restoration using edge detection result in (c); (e) The MI penalty term during the 1st iteration; (f) Regularized Bayesian framework provides the result with minimum error.

- [2] Y. Chen, H. Wang, T. Fang, and J. Tyan. Image compounding based on independent noise constraint. In *Proc. IEEE Int'l Conf. Acoustic Speech Signal Processing*, March 2005.
- [3] D. Edwards. *Introduction to Graphical Modelling, 2nd ed.* Springer-Verlag, 2000.
- [4] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intell.*, 6(6):721–741, November 1984.
- [5] H. Ishikawa. Exact optimization for markov random fields with convex priors. *IEEE Trans. Pattern Anal. Machine Intell.*, 20(10):1333–1336, October 2003.
- [6] A. C. Kak and M. Slaney. *Principles of Computerized Tomographic Imaging*. Society of Industrial and Applied Mathematics, 2001.
- [7] J. Kim, V. Kolmogorov, and R. Zabih. Visual correspondence using energy minimization and mutual information. In *Proc. IEEE Int'l Conf. on Computer Vision*, 2003.
- [8] T. W. Lee. *Independent Component Analysis: Theory and Applications*. Kluwer Academic Publishers, 1998.
- [9] S. Z. Li. *Markov Random Field Modeling in Image Analysis*. Tokyo: Springer, 2nd edition, 2001.
- [10] J. Mateos, R. Molina, and A. Katsaggelos. Bayesian high resolution image reconstruction with incomplete multisensor low resolution systems. In *IEEE Int'l Conf. on Acoustic, Speech and Signal Processing*, pages III:705–708, 2003.
- [11] R. Molina, J. Mateos, A. K. Katsaggelos, and M. Vega. Bayesian multichannel image restoration using compound Gauss-Markov random fields. *IEEE Trans. Image Processing*, 12(12):1642–1654, December 2003.
- [12] D. D. Muresan and T. W. Parks. Adaptively quadratic (AQua) image interpolation. *IEEE Trans. Image Processing*, 24:657–673, 2002.
- [13] L. R. Rabiner and B. H. Juang. An introduction to hidden Markov models. *IEEE Trans. Acoust., Speech, Signal Processing*, 3(1):4–15, January 1986.
- [14] P. M. Shankar. Speckle reduction in ultrasound b-scans using weighted averaging in spatial compounding. *IEEE Trans. on Ultrasonics, Ferroelectrics and Frequency Control*, 33(6):754–758, 1986.
- [15] G. E. Trahey, J. W. Allison, S. W. Smith, and O. T. von Ramm. A quantitative approach to speckle reduction via frequency compounding. *Ultrasonic Imaging*, 8:151–164, 1986.
- [16] V. Vapnik. *The Nature of Statistical Learning Theory*. New York, NY: Springer-Verlag, 1995.
- [17] R. F. Wagner, S. W. Smith, J. M. Sandrik, and H. Lopez. Statistics of speckle in ultrasound B-scans. *IEEE Trans. on Sonics and Ultrasonics*, 33(3):156–163, 1983.
- [18] W. F. Walker and G. E. Trahey. The application of k-space in medical ultrasound. *IEEE Trans. on Ultrasonics, Ferroelectrics and Frequency Control*, 45:541–558, 1998.
- [19] Y. Weiss. Smoothness in layers: Motion segmentation using nonparametric mixture estimation. In *Proc. IEEE Int'l Conf. on Comput. Vis. and Patt. Recog.*, pages 520–527, 1997.
- [20] J. Zhang and G. G. Hanauer. The application of mean field theory to image motion estimation. *IEEE Trans. Image Processing*, 4(1):19–33, January 1995.