Regularized Image Restoration

Pradeepa D. Samarasinghe and Rodney A. Kennedy

Research School of Engineering, College of Engineering and Computer Science,
The Australian National University, Canberra, ACT,
Australia

1. Introduction

Image restoration or deconvolution of a blurred natural image is a mature research activity with a rich set of available techniques and algorithms, well-summarised in review articles, Banham & Katsaggelos (1997); Kundur & Hatzinakos (1996). Despite this history and volume of work, there is current research activity motivated by the desire to find yet superior methods to restore the ground truth image (GTI). Important performance metrics to assess the efficacy of restoration methods include: restoration accuracy, computational complexity and convergence speed. In this chapter we use these performance metrics in the development of restoration methods of greatest utility for real-world applications where complexity/speed is a major concern and the evaluation of image restoration needs to take into account the highly structured features of natural images and, to a lesser extent, the human visual system.

The scope of this work focusses on non-blind image restoration where the point spread function (PSF) of the blur convolutional kernel is known. Blind deconvolution is, by its nature, a more challenging problem, Haykin (1994); Kundur & Hatzinakos (1996). However with effective and efficient PSF estimation techniques, Fergus et al. (2006); Joshi et al. (2008); Krahmer et al. (2006); Nayar & Ben-Ezra (2004); Oliveira et al. (2007), the research trend has been to handling blind deconvolution in two steps, with PSF estimation as the first step and image estimation as the second step, Levin et al. (2009). This motivates us to focus on efficient algorithms for image restoration where the blur convolutional kernel is known.

In this chapter, we first analyze existing linear deterministic restoration models and develop a class of novel models with better performance. Then using regularization as the basis, we link linear deterministic and stochastic restoration models. By introducing a previously developed novel visual metric to image regularization analysis, we study the purported superior performance of stochastic prior models and demonstrate that those models are not superior to simpler linear deterministic prior models. In addition, we show that the high complexity "derivative likelihood" models under the maximum a posteriori (MAP) framework offer no significant advantage to a properly configured, efficient "normal likelihood" model.

2. Quadratic regularization in image restoration

2.1 Regularization

Image acquisition being an inverse problem can be modeled by a continuous model in an infinite dimensional space, which is categorized as a (linear) Fredholm integral equation of

the first kind, Demoment (1989). In the sense of Hardamard, Hadamard (1952), a solution to a well-posed problem satisfies the conditions of existence, uniqueness and stability. As Fredholm integral equations of the first kind do not meet the criteria for a well-posed problem, image restoration belongs into the general class of problems which are classified as ill-posed problems, Tikhonov & Arsenin (1977). The ill-posed nature of image restoration problem implies that, small bounded perturbations in the data may lead to unbounded deviations in the solution, Phillips (1962).

For images defined on a discrete set, linear algebra can be used to find solutions for ill-posed problems such as image restoration. One of the simplest methods to restore images affected by a linear distortion is the use of the pseudo inverse, Albert (1972), for which the solution fulfils the first two conditions (existence and uniqueness) of Hardamard's well-posed problem, Hadamard (1952), but fails in meeting the stability condition. This motivates or leads to regularization as one of the most widely accepted and used techniques, in which the solution fulfils all three conditions of a well-posed problem. The concept underlying regularization is to find an acceptable solution from imperfect data, for which, the problem should be stated more completely by including some extra or priori information, Miller (1970); Tikhonov & Arsenin (1977).

Regularization approaches to image restoration are classified broadly in two ways: stochastic regularization which uses the knowledge of covariance matrices of the GTI and noise; and deterministic regularization which deems that most natural images are relatively featureless with limited high-frequency activity, Banham & Katsaggelos (1997). While stochastic regularization has been used extensively in the past, with important contributions to the field such as Wiener filter, Wiener (1942), recently, much emphasis has been on the use of derivative filters with deterministic regularization, Fergus et al. (2006); Levin et al. (2007). Thus, our contribution in this chapter relates to deterministic regularization and the term regularization, henceforth, refers to deterministic regularization techniques.

Among many regularization techniques, Tikhonov, Tikhonov & Arsenin (1977) regularization is one of the first and best-known techniques for stabilization. It was proposed in Tikhonov & Arsenin (1977), that the solution for

$$b = \mathcal{K}g + n, \tag{1}$$

where b is the measured data, g is the original data (ground truth), $\mathcal{K}$ is the distortion operator or the transformation and n represents additive random noise, can be achieved by constrained minimization of a functional $\Phi(g)$, which is called the *stabilizing functional*. Under the stabilizing functional approach, the image restoration problem is formulated as determining an estimate $\widehat{g}$ of g, which minimizes the functional $\Phi(g)$ under the condition that the estimate $\widehat{g}$ satisfies

$$\|b - \mathcal{K}g\|^2 = \delta, \tag{2}$$

where δ is a positive constant and $\|\cdot\|$ denotes the Frobenius norm

$$\|A\| = \sqrt{\sum_{i=1}^{m_1} \sum_{j=1}^{m_2} a_{ij}^2} \tag{3}$$

for some matrix A and a_{ij} is the (i, j) entry. The constrained minimization problem in (2) can be solved by the method of Lagrange multipliers, which is to determine $\widehat{g}$, an estimate of ground truth g, by minimizing the functional

$$\|b - \mathcal{K}g\|^2 + \lambda\,\Phi(g), \tag{4}$$

where $\lambda > 0$ is the Lagrange multiplier and is often called as the *regularization parameter*. As the regularization parameter, λ, controls the tradeoff between the solution accuracy $\|b - \mathcal{K}g\|^2$ and its degree of regularity $\Phi(g)$, choosing a proper value for λ is important in image restoration.

The first term in (4), named as *data-fidelity term* fits to the data, while *stabilizing functional* incorporates "believed" properties of the GTI. Generally the *data-fidelity term* is a standard fixed choice. In contrast, the richness and variety of image restoration techniques comes down to different choices of the regularization term, reflecting different implicit models. As the choice of the *stabilizing functional* can take a variety of forms, in this chapter, we selected two widely used model classes for our analysis: the fast quadratic stabilizing functionals introduced in section 2.2, and Sparse and Laplacian prior methods in section 3.2. The latter model class can be developed from relating the *stabilizing functional* to the *priori* knowledge using a probabilistic viewpoint and is claimed to have better performance, Levin et al. (2007).

When the stabilizing functional $\Phi(g)$ in (4) belongs to the class of nonnegative quadratic functionals, the minimization problem can be expressed as

$$\widehat{g} = \arg\min_{g}\ \|b - \mathcal{K}g\|^2 + \lambda\,\|\mathcal{D}g\|^2, \tag{5}$$

where $\mathcal{D}$ is a bounded linear operator, Miller (1970) and is often called the *regularization operator* or *stabilizing operator*. It is shown in Hunt (1973), that the minimization problem in (5) can be formulated as a constrained least squares image restoration problem when the solution g satisfies the necessary and sufficient condition of

$$(\mathcal{K}^T\mathcal{K} + \lambda\,\mathcal{D}^T\mathcal{D})g = \mathcal{K}^T b. \tag{6}$$

This leads to the closed form solution for (5) in the form

$$\widehat{g} = \big(\mathcal{K}^T\mathcal{K} + \lambda(\mathcal{D}^T\mathcal{D})\big)^{-1}\mathcal{K}^T b. \tag{7}$$

We extend, in a trivial way, the above formulation by considering $\mathcal{D}$ as the combination of R component regularization operators, in the form of

$$\mathcal{D}g \triangleq \big(\mathcal{D}_1^T, \mathcal{D}_2^T, \ldots, \mathcal{D}_R^T\big)^T g. \tag{8}$$

With the introduction of R Lagrange multipliers, the general form of (5) can be expressed as

$$\widehat{g} = \arg\min_{g}\ \|b - \mathcal{K}g\|^2 + \sum_{r=1}^{R} \lambda_r\|\mathcal{D}_r g\|^2, \tag{9}$$

for which, the closed form solution is given by

$$\widehat{g} = \big(\mathcal{K}\mathcal{K}^T + \sum_{r=1}^{R} \lambda_r(\mathcal{D}_r^T\mathcal{D}_r)\big)^{-1}\mathcal{K}^T b. \tag{10}$$

As the images are of limited support and when the corresponding hypothesis of uniformity on image edges can be made, the matrices $\mathcal{K}$ and $\mathcal{D}$ in (10) have a special structure and are called block circulant matrices, Hunt (1971). As circulant matrices can be diagonalized by the discrete Fourier transform, the minimization in (9) can be solved extremely quickly using the Fourier domain techniques, Hunt (1973).

2.2 Regularization operators as components in quadratic stabilizing functionals

The generality of the regularization operator allows the development of a class of linear operators and the minimization in (9) will be the source of many regularizing solutions for (1) depending on the choice of the regularization operator. This choice is usually based on the known details of the image formation process and plays an important role in the regularization.

The simplest regularization operator is when $\mathcal{D}$ is an identity matrix, where $\mathcal{D}g = g$, and the regularized solution for this was referred as *minimum norm restoration*, Hunt & Andrews (1977). In general, $\mathcal{D}$ often takes the form of a sparsifying operator such as a discrete approximation of a derivative operator. Through the experiments in Zhu & Mumford (1997), it was shown that even though the statistics of natural images vary from image to image, the histograms for the response of derivative filters are relatively consistent and scale invariant across the images. Taking these factors into consideration, in this section, we discuss a class of regularization operators based on the partial derivative operators (PDO), which could be used in the quadratic stabilizing functional.

2.2.1 First order partial derivative operator

When first order partial derivative operators (FOPDO) are considered as the regularization operators, $\mathcal{D}g$ in (8) can be expressed as

$$\mathcal{D}g = \begin{pmatrix} \partial_x \\ \partial_y \end{pmatrix} g,$$

where ∂_x and ∂_y are any discrete space, spatially invariant linear operators that emulate first order derivative in x and y directions, respectively, Levin et al. (2007). This type of regularization uses two component regularization operators.

2.2.2 Second order partial derivative operator

Second order partial derivative operators (SOPDO), can be derived mainly in two forms.

1. Isotropic SOPDO – When the regularization operator takes the form

$$\mathcal{D}g = \begin{pmatrix} \partial_{xx} \\ \partial_{yy} \end{pmatrix} g,$$

it is called the isotropic SOPDO. Though the SOPDO defined above cannot be considered as a true isotropic differential operator, such as the continuous Laplacian operator, it gives the simplest possible isotropic operator with even-order derivatives, Leung & Lu (1995). Similar to FOPDO, ∂_{xx} and ∂_{yy} represent any discrete space, spatially invariant linear operators that emulate second order derivatives.

2. Non-isotropic SOPDO – The non-isotropic SOPDO is formed by

$$\mathcal{D}g = \begin{pmatrix} \partial_{xx} \\ \partial_{xy} \\ \partial_{yy} \end{pmatrix} g.$$

As the edges and lines in images may occur in any direction, when the differential operator is isotropic it would give better results than a non-isotropic differential operator, Leung & Lu (1995).

2.2.3 Mixed partial derivative operator

In general, considering only even-order derivatives, the use of directional derivatives in more than one dimensional can be expressed as

$$\mathcal{D}g = (\partial^p_{s_1^p}, \partial^p_{s_2^p}, \ldots, \partial^p_{s_m^p}) g. \tag{11}$$

where p is the order of the derivatives, m is the number of dimensions and s_1 to s_m represent the direction of the derivatives.

Using the above general model, we introduce a new regularization operator, with different combinations of higher order derivative operators. In this discussion, we limit the use of higher order derivative operators up to the second order, and the new PDO is called first and second order derivative operator (FSOPDO). With FSOPDO, $\mathcal{D}g$ in (11) takes the form

$$\mathcal{D}g = \begin{pmatrix} \partial_x \\ \partial_y \\ \partial_{xx} \\ \partial_{yy} \end{pmatrix} g.$$

These quadratic regularization functionals are compared in a new perspective with the widely used prior models which are believed to result in better performance in section 3.5.

2.3 Non-blind image restoration through SOPDO

2.3.1 Noisy image deconvolution

Although most previous image restoration algorithms have considered FOPDO as the regularization model, Levin et al. (2007), we claim that SOPDO has better performance in terms of the difference between the ground truth and the estimated data on images which are susceptible to noise. Here, we deal only with additive Gaussian noise, as it effectively models the noise in many different imaging scenarios. In this section, we study in detail a few simulation results which are used to do comparison evaluations with other existing image restoration techniques.

We take non-isotropic SOPDO as the regularization operator for image restoration through least squares restoration as given in section 2.2. In order to compare the performance of the non-isotropic SOPDO prior model, we take two regularization models, FOPDO and a sparse stabilizing functional defined in Levin et al. (2007). Relating regularization to probability, the stabilizing functional in image restoration is also referred as prior model and

Smoo-	Non-isotropic SOPDO		FOPDO		Sparse		
thing	MSE	Time	MSE	Time	Itera-	MSE	Time
weight	$\times 10^{-4}$	in seconds	$\times 10^{-4}$	in seconds	tions	$\times 10^{-4}$	in seconds
0.1	4.978	0.43	5.168	0.44	10	8.050	27
	4.646	0.42	4.730	0.44	10	6.899	27
0.05					50	6.153	117
					100	6.128	231
0.01	6.043	0.42	7.561	0.44	10	4.289	27

Table 1. Comparison of stabilizing functional model.

a detailed discussion of the Bayesian interpretation to regularization, including the Sparse prior model, is covered in section 3.2. The deconvolution with non-isotropic SOPDO and FOPDO regularization lead to closed form solutions with highly efficient computation, while the Sparse prior cannot be minimized in closed form, Levin et al. (2007). In all the simulations discussed in this section we use

$$\lambda_r = \lambda, \quad \forall\, r \in 1 \ldots R, \tag{12}$$

for the quadratic regularization functionals, where the value of R depends on the respective model such as $R = 2$ for FOPDO and $R = 3$ for non-isotropic SOPDO model.

We claim, using non-isotropic SOPDO prior gives better results for images which are susceptible to noise over the FOPDO. When comparing the non-isotropic SOPDO regularization with the Sparse prior, we found that the non-isotropic SOPDO regularization outperforms Sparse prior significantly in speed. These results are shown in Table 1.

For the experiment in Table 1, we added "Gaussian" noise to the original "Picasso" image, Shan et al. (2008) with a standard deviation of 0.0001 (relative to the image value range of 0 to 1). The original colored image was first converted to greyscale with the pixel values resulting in the range from 0 to 1 and the original image was considered to be periodic. The term MSE stands for mean square error and for a two dimensional image, MSE is defined as

$$MSE \triangleq \frac{1}{L_1 L_2} \sum_{\ell_1=1}^{L_1} \sum_{\ell_2=1}^{L_2} \left(g(\ell_1, \ell_2) - \widehat{g}(\ell_1, \ell_2) \right)^2 \tag{13}$$

where g and $\widehat{g}$ represent GTI and the estimated GTI, respectively while L_1 and L_2 represent the size of the image in x and y directions, respectively. The MSE values in Table 1 are in multiples of 10^{-4} while the time is given in seconds. The results show that the non-isotropic SOPDO outperforms FOPDO on MSE and has a significant advantage over Sparse prior on speed performance.

2.3.2 Efficiency in deconvolution

As SOPDO can use frequency domain deconvolution techniques, it can be implemented highly efficiently than most of the recent non-blind deblurring techniques. The comparison was done with the Sparse deconvolution algorithm in Levin et al. (2007) named as "Levin Sparse deconvolution" and the non blind deconvolution of, Shan et al. (2008) (distributed online) named as "Shan executable". The results in Table 2 support the claim that

Restoration Technique	Efficiency in seconds
Levin Sparse deconvolution (50 iterations)	556
Levin Sparse deconvolution (10 iterations)	124
Shan executable	39
Non-isotropic SOPDO deconvolution	2

Table 2. Efficiency in non-blind image deconvolution

Restoration Technique	Image size in pixels	Kernel size in pixels	Efficiency in seconds
Levin Sparse deconv. (50 iterations)	484×752	19×27	576
		99×99	556
	910×903	99×99	1240
	1107×1694	99×99	2429
Shan executable	484×752	19×27	40
		99×99	73
	910×903	99×99	166
	1107×1694	99×99	Error
Non-isotropic SOPDO regularization	484×752	19×27	2.34
		99×99	2.44
	910×903	99×99	5.54
	1107×1694	99×99	13.17

Table 3. Efficiency results on scaling

non-isotropic SOPDO regularization model results in the best speed performance when compared with "Levin Sparse deconvolution" and "Shan executable" methods.

Further, we tested for the robustness of non-isotropic SOPDO regularization by using different sized images with varying sized kernels. The detailed results are shown in Table 3. All the images used for this experiment are color images, having separate *rgb* (red, green, blue) channels. The image and kernel sizes are given in pixels and the efficiency was measured in seconds. The results clearly show the robustness and the efficiency of the non-isotropic SOPDO regularization model with respect to different scales of image and kernel.

2.3.3 Performance in deconvolution

Several computational experiments were carried out in order to compare non-isotropic SOPDO regularization with "Levin Sparse deconvolution" and "Shan executable". The performance of these deconvolution techniques on a naturally blurred, highly textured image, given in Shan et al. (2008), are shown in Fig. 1 and Fig. 2. The blur kernel used in this experiment was retrieved through the blind deconvolution package of, Shan et al. (2008) distributed online. Closer visual inspection of the image results show that non-isotropic SOPDO technique best shows the tree branches and leaves while the other techniques have a blurring effect still remaining on the estimated result. This fact is further discussed and evidenced by evaluating the deconvolution in a new perspective in section 3.5.

(a) 27 × 27 Blur Kernel scaled upwards   (b) Blurred Image

Fig. 1. Image results for a highly textured image

3. Comparison of sparse prior models to quadratic regularization

3.1 Key issues

While the development of regularized solutions for ill-posed problems is widely discussed in the signal processing literature, recently by looking at the ill-posed image restoration problem from a probabilistic view point, some researchers claim that the Sparse prior model, Fergus et al. (2006); Levin et al. (2007) (discussed in more detail below in section 3.2.1) outperforms quadratic regularization models (discussed in section 2.2). The analytical study in this section addresses the following problems:

1. Are sparse prior models superior to quadratic regularization models?
2. What is the source of better performance of sparse prior models?
3. Are fast quadratic regularization models good enough for image restoration?

3.2 Regularization – Bayesian interpretation

Inverse problems such as image restoration are seen as probabilistic inference problems, where lack of information is compensated by assumptions. Therefore, it is not surprising, when the nature of the regularization detailed above is taken into consideration, to see that there is a close relationship between regularization and Bayesian estimation. Applying Bayes theorem to the image restoration problem in (1), for a known blur kernel, the posterior distribution can be written as

$$p(g|b) \propto p(b|g)\, p(g), \tag{14}$$

where $p(b|g)$ represents the likelihood and $p(g)$ represents the prior for the ground truth image. The estimation of the GTI based on posterior distribution can be classified in several ways. The minimum mean-square error estimate represents the mean of the posterior density, the MAP estimate stands for the mode of the posterior density while the maximum likelihood

(a) Shan executable

(b) Levin Sparse deconvolution – 10 iterations

(c) Levin Sparse deconvolution – 50 iterations
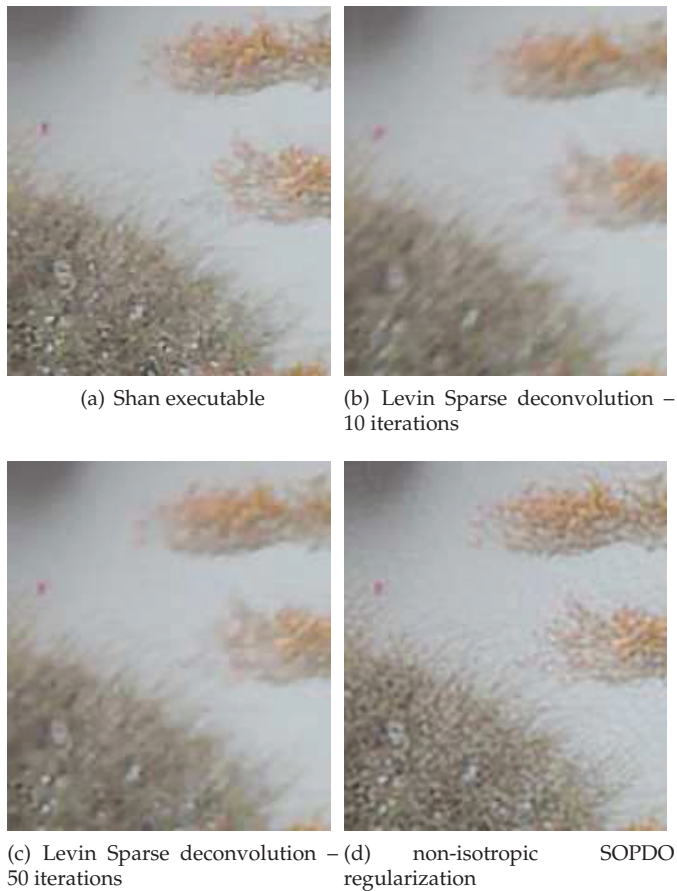
(d) non-isotropic SOPDO regularization

Fig. 2. Image results for a highly textured image

(ML) estimate may be viewed as a special case of MAP where no prior distribution is used, Hunt (1977).

Under the MAP technique, estimation of the GTI simplifies to

$$\widehat{g} = \arg \max_{g} p(g|b). \tag{15}$$

Considering the non-blind image deconvolution process, we convert (15) to an energy minimization problem, where the energy is defined as

$$E(g) \triangleq - \log \big(p(g|b) \big). \tag{16}$$

Different likelihood and prior models on the ground truth have been applied for image restoration in literature. An analysis of existing prior models can be found in Mignotte (2006).

Considering the fact that for a given g, the variation in b is due to the noise n, Hunt (1977), together with the above definitions, non-blind image restoration problem can be recast as seeking the unknown GTI, $g(i,j)$, that minimizes the functional

$$\left\| b - \mathcal{K}g \right\|^2 + \sum_{r=1}^{R} \sum_{ij} \lambda_r \, \rho\left(\mathcal{D}_r g(i,j)\right), \tag{17}$$

where $\mathcal{D}_r$ is the rth of R linear operators, i, j are pixel indices, $\lambda_r > 0$ are the regularization parameters, $\| \cdot \|$ stands for the Frobenius norm and $\rho(\cdot)$ is a scalar memoryless nonlinear mapping, generally taking the form

$$\rho(z) \triangleq |z|^{\alpha} \tag{18}$$

for judicious choice of real parameter α (not necessarily integer).

Many techniques belong to this class and differ only in: the set of linear operators $\mathcal{D}_r$, $r = 1, 2, \ldots, R$, and the nonlinear mapping $\rho(z)$ (or choice of α). Numerous image restoration techniques have been developed under this framework from the early work, Geman & Geman (1984); Greig et al. (1989) to the most recent research, Fergus et al. (2006); Levin et al. (2007; 2009); Shan et al. (2008).

3.2.1 Sparse prior model

In recent literature, it is shown that, when derivative filters are applied to natural images, the filter outputs tend to be sparse, Olshausen et al. (1996); Simoncelli (1997). That is, the histogram of the derivative filtered image peaks at zero and falls off much faster than a Gaussian distribution. These heavy tailed natural image priors are used in a number of applications in image processing literature, such as denoising, Roth & Black (2005); Simoncelli (1999), reflection separation, Levin & Weiss (2007); Weiss (2001) and deconvolution, Levin (2007); Shan et al. (2008) in which, they are implemented in various ways such as student-t distributions, Roth & Black (2005) and scale mixtures of Gaussian distributions, Fergus et al. (2006); Portilla et al. (2003).

In Levin et al. (2007), sparsity is incorporated by having $\mathcal{D}_r$ as the derivative filters and $\alpha = 0.8$ in (18) as the prior term, which results in

$$\left\| b - \mathcal{K}g \right\|^2 + \sum_{r=1}^{R} \sum_{ij} \lambda_r \left(\mathcal{D}_r g(i,j)\right)^{0.8}. \tag{19}$$

This can be solved in the spatial domain using the Conjugate Gradient algorithm, Barrett et al. (1994).

3.2.2 Laplacian prior model

Although not as close as the Sparse prior to the natural image priors, Laplacian prior with $\alpha = 1$ in (18) is expected to result in a less smooth solution than the Gaussian prior. With the Laplacian prior, the optimization becomes

$$\left\| b - \mathcal{K}g \right\|^2 + \sum_{r=1}^{R} \lambda_r \| \mathcal{D}_r g \|_1. \tag{20}$$

Recently, much attention has been paid in solving L^1 norm regularization problems through compressed sensing. in Kim et al. (2007), an efficient method for optimizing a solution to a problem similar to (20) was discussed when $\mathcal{D}_r$ are invertible.

3.2.3 Gaussian prior model

When $\alpha = 2$, minimization in (17) is called the Gaussian prior deconvolution in Levin et al. (2007) and is equivalent to the quadratic regularization problem in (9). Thus, in this chapter, we use the terms Gaussian prior and quadratic (specifically isotropic SOPDO) regularization interchangeably.

3.3 Image restoration evaluation

3.3.1 Visual metric for evaluation

For all the restoration performance analysis and comparisons in this paper, we use a recently developed visual metric called SSIM (Structured SIMilarity) index, Wang et al. (2004), which has not been used for the comparison of prior models in the image restoration literature to date. The approach of SSIM is motivated by the highly structured characteristics of the natural image, where the strong neighborhood dependencies carry important information about the structures of the objects in the visual scene, Wang & Bovik (2009).

Assuming x and y are local image patches representing the same patch in the original and estimated images, the local SSIM index measures the similarities of three elements of the image patches: the similarity $\ell(x,y)$ of the local patch luminances (brightness values), the similarity $c(x,y)$ of the local patch contrasts, and the similarity $s(x,y)$ of the local patch structures. These local similarities are expressed using simple, easily computed statistics, and combined together to form local SSIM, $S(x,y)$, Wang & Bovik (2009).

$$S(x,y) = \ell(x,y) \cdot c(x,y) \cdot s(x,y)$$
$$= \left(\frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \right) \cdot \left(\frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \right) \cdot \left(\frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \right), \tag{21}$$

where μ_x and μ_y are the local sample means of x and y, respectively, σ_x and σ_y are the local sample standard deviations of x and y, respectively, and σ_{xy} is the sample cross correlation of x and y after removing their means. The items C_1, C_2, and C_3 are small positive constants that stabilize each term, so that near zero sample means, variances or correlations do not lead to numerical instability.

Due to the fact that the underlying principle of SSIM is to extract the structural information which complies with the human visual system, SSIM maps are asserted to be a better signal fidelity measurement over MSE, Wang & Bovik (2009). In evaluating images through MSE, all image pixels are treated equally and content dependent variations in image fidelity are not accounted for. The two main indicators in SSIM evaluations, mean SSIM (MSSIM) and SSIM maps have values in the range from 0 to 1, where 1 indicates the best restoration. Although MSSIM and SSIM maps are generally used as visual fidelity metrics, we evaluate image restoration with the histogram of the SSIM map, as it provides an accurate view of the local restoration.

$$g \longrightarrow \boxed{\mathcal{K}} \longrightarrow b \longrightarrow \boxed{\mathcal{L}} \longrightarrow \hat{g}$$
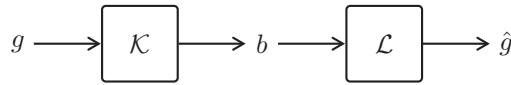
Fig. 3. Image restoration model, where g is the ground truth image, b is the distorted image, $\mathcal{K}$ is the blur operator, $\mathcal{L}$ is the deblur process and $\hat{g}$ is the estimated image.
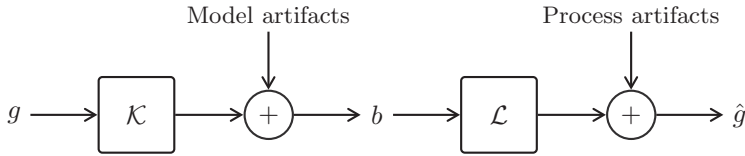
Model artifacts   Process artifacts

$$g \longrightarrow \boxed{\mathcal{K}} \longrightarrow \bigoplus \longrightarrow b \longrightarrow \boxed{\mathcal{L}} \longrightarrow \bigoplus \longrightarrow \hat{g}$$

Fig. 4. Simulated image restoration model.

3.3.2 Image restoration models

Ignoring the presence of noise in image acquisition represented by (1), general image restoration could be represented by the model shown in Fig. 3, where $\mathcal{L}$ represents the deblur process. The notation in (1) and the representation in Fig. 3 may be an over-simplification. From physical intuition, we could see that even though g is continuous by nature, image recording imposes limitations on the spatial extent of g and b, leading to artifacts which impact on the final estimate of image restoration.

As illustrated in Fig. 4, we categorize these spatial artifacts in two ways. The "Model artifacts" are those, which are not present on naturally blurred images, but introduced in blur simulations as a result of sharp intensity transitions at the boundary of a finite image. Generating a blur image from a finite GTI causes unnatural blur distortions in the vicinity of the boundary of the image. Suppression of these "Model artifacts" could be accomplished by preprocessing the observed degraded image with techniques such as truncation and reducing the size of the blurred image. On the other hand, "Process artifacts" come along with the deblur process $\mathcal{L}$ due to finite b, which affect the performance of most deconvolution algorithms.

In order to show the effect of "Process artifacts", we restore an image, originally, of size 255×255 pixels, but truncated in order to remove the "Model artifacts" introduced by a 13×13 pixels blur kernel, making the final image of size 242×242. The results of restoration with Sparse, Laplacian and Gaussian priors are shown in Fig. 5. In this experiment, deconvolution with Sparse and Laplacian priors were carried out using iterative re-weighted least squares (IRLS) method, Meer (2004), through the code available online, Levin et al. (2007), while the Gaussian prior is processed with both IRLS and fast Fourier techniques (FFT) separately. In our simulations, we processed IRLS for 150 iterations beyond which there were no further improvements. Analyzing the results of the performance of the Gaussian prior model with the FFT and IRLS techniques, we see that process artifacts are better handled by the IRLS technique than the FFT and this result is justified by the IRLS processing of Sparse and Laplacian prior models.

Both the "Model" and "Process" artifacts discussed above are not part of natural images, but are imposed artificially by the image modeling and processing techniques. Thus we claim that
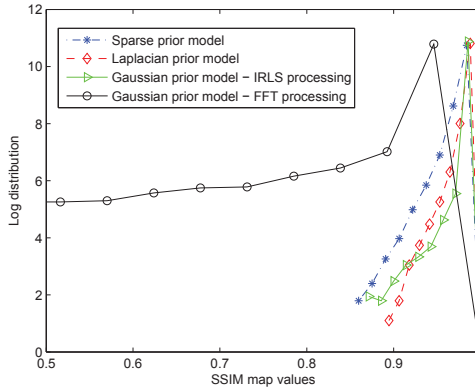
Fig. 5. Image restoration results with prior models.

Regularization operator	r	λ_r
FOPDO	1, 2	β_1
SOPDO	3, 4	β_2
FSOPDO	1, 2	β_1
	3, 4	β_2

Table 4. Choice of regularization parameter (λ) values for different quadratic regularization operators used in the simulations of Table 5

the evaluation of image restoration should be carried out excluding these artifacts to properly assess the performance of any image restoration method.

3.4 Performance of quadratic regularization operators

In order to achieve the objective of studying the performance of different operators in the quadratic regularization as detailed in section 2.2, we carried out some simulations, where we avoided the effect of "Model artifacts" by taking a boundary strip off from the blurred image. In our evaluations, we used FOPDO, SOPDO and FSOPDO models to compare the performance. From this point onwards the term SOPDO refers to isotropic-SOPDO unless stated otherwise.

The simulations, for which the results are demonstrated in Fig. 6, are executed in the same environment as the simulation for Fig. 5, but with quadratic regularization models. We evaluated the performance of the regularization models under varying regularization parameter (λ) values as discussed in section 2.2. While the choice of parameters representing λ for FOPDO, SOPDO and FSOPDO are given in Table 4, the actual values for the respective parameters are given in Table 5.

While the overall SSIM values for few of the simulation results under varying λ values are shown in Table 5, the histogram distribution representing the first line of Table 5 is shown in Fig. 7. Overall, by analyzing these results, we claim that, in the presence of "Process artifacts", a better performance could be achieved with FSOPDO over FOPDO and SOPDO
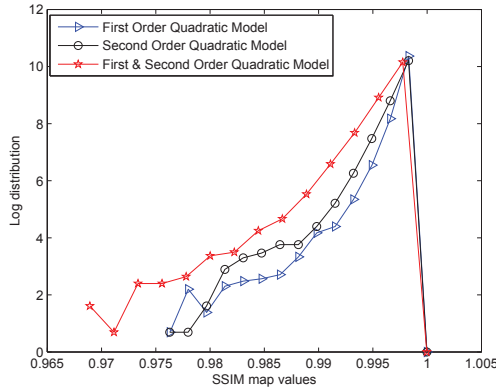
Fig. 6. Image restoration results with quadratic regularization models.

	MSSIM values for		
Regularization parameter	FOPDO	SOPDO	FSOPDO
$\beta_1 = \beta_2 = 0.001$	0.9412	0.9597	0.9626
$\beta_1 = 0.001, \beta_2 = 0.003$	0.9596	0.9597	0.9657
$\beta_1 = 0.003, \beta_2 = 0.001$	0.9412	0.9674	0.968

Table 5. Performance of quadratic regularization operators under varying regularization parameter values

(a) Result for row 2
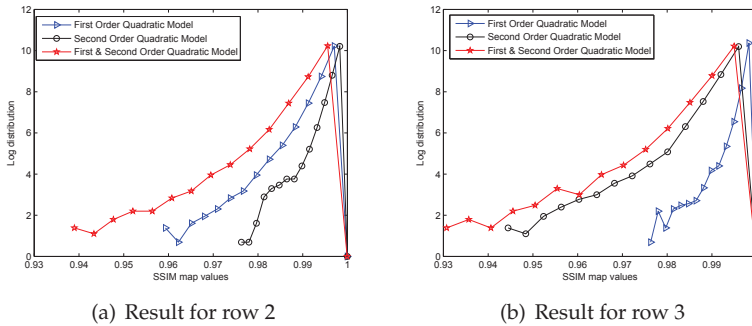

(b) Result for row 3

Fig. 7. Image restoration results for simulations in Table 5

models. In the next section we compare the performance of these quadratic regularization models by removing the "Process" and "Model" artifacts.

3.5 Regularization model performance comparison

As shown earlier in section 2.2, we modeled the regularization of image restoration based on the quadratic regularization terms (sometimes called as the least squares regularization) and in section 3.2, we discussed the existing probabilistic models under a MAP framework.
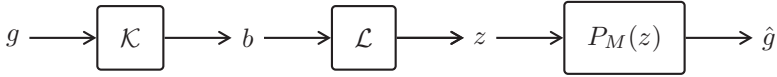
Fig. 8. Image restoration model for a naturally blurred image, where $\mathcal{K}$ is the blur process, $\mathcal{L}$ is the deblur process, g, b, z, $\hat{g}$ stand for GTI, blur image, deblurred image with artifacts, and the final estimated GTI respectively. The process $P_M(z)$ decouples "Model" and "Process" artifacts from the deblurred image .

These models form a method of regularization in image restoration. This section is devoted for the comparison of these models. The comparison in this section will guide us for making recommendations for the appropriate regularization technique and is discussed at the end of this section.

3.5.1 SSIM performance comparison

As the objective of our simulations is to evaluate the contribution of the regularization models towards image restoration, we use the restoration model shown in Fig. 8, where we decouple artifact effects from restoration by projecting the estimated image with

$$P_M(z)(i,j) = \begin{cases} z(i,j), & \text{if } i,j \in M \\ 0, & \text{otherwise} \end{cases} \tag{22}$$

where M is a region without "Model" and "Process" artifacts.

To be consistent with the SSIM map region in Fig. 5, we take a large image of support 1024×1024 and project the final image to a 242×242 region within the inner region of the estimated image, which is least affected by the artifacts. The restoration was carried out with FFT processing of the Gaussian prior and IRLS processing of Sparse and Laplacian priors. The comparison of the performance of the priors is shown in Fig. 9. In it we note that Gaussian prior with FFT processing has performed as well as or better to the Sparse and Laplacian prior models.

As the literature claims that iterative algorithms such as conjugate gradient algorithms suppress noise and perform better in noisy blur image restoration, we simulated a noisy blurred restoration under the same conditions given for Fig. 9, but with different regularization parameter values, as more weight should now be given to the prior over data. The noise added was Gaussian with zero mean and 0.01 variance. The optimal results we obtained for varying λ are shown in Fig. 10. With these results, we claim that Gaussian prior handles noisy images as better as the Sparse and Laplacian prior models.

Thus, these results pave a new path of thinking and we claim that quadratic regularization with SOPDO model, when appropriately configured and used in a realistic context, free from unnatural artifacts, is comparable to Sparse prior model in terms of image restoration performance under the SSIM criterion.

3.5.2 Efficiency comparison

As the optimization problem in least-squares regularization is convex and as the fast Fourier techniques could be applied for the computation, for an image of size $L \times L$ pixels, the restoration through least-squares regulation has a complexity of $O(L \log L)$ operations. In
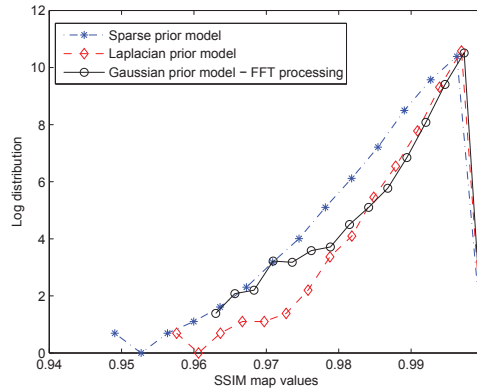
Fig. 9. Image restoration results for the system in Fig. 8 with FFT processing of Gaussian Prior and IRLS processing of Sparse and Laplacian prior models.
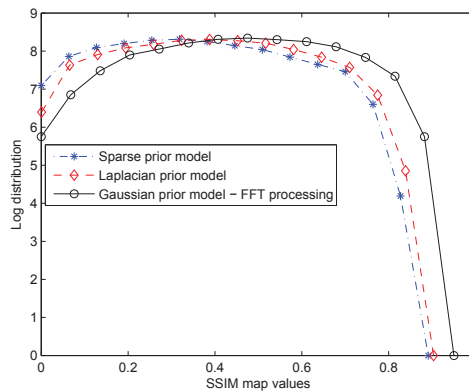

Fig. 10. Image restoration results for a noisy image under the same environment in Fig. 9.

contrast, when a Sparse prior is used, the optimization problem is no longer convex and cannot be minimized in closed form. Using the conjugate gradient method, Barrett et al. (1994), or IRLS method, the optimization can be solved in $O(L\, i_{max})$ where i_{max} represent the maximum number of iterations.

A few simulation results on efficiency are shown in Table 6, where all the values are in seconds and represent the time taken for the restoration using each of the respective model. While the quadratic regularization deconvolution was carried out using Fourier domain techniques, the Sparse deconvolution was carried out using the IRLS method. Under the IRLS algorithm, it is experienced that in order to achieve an acceptable result, the number of iterations should be at least 50 and better results could be achieved when the number of iterations are above 100. From the results shown, it is evident that when the size of the image increases, the relative efficiency of the restoration through Sparse prior model becomes extremely low.

Image size in pixels	Time for Quadratic regularization with FFT	Time for Sparse prior	
		50 iterations	100 iterations
grey-scale 255×255	0.08	30	60
grey-scale 1024×1024	1.1	303	596
colored 484×752	1.7	371	730
colored 1107×1694	9.0	2180	4292

Table 6. Efficiency of regularization operators. The times taken for restoration of grey-scale and colored images are given in seconds for each of the regularization operators.

3.5.3 Regularization recommendations

In addition to lower efficiency and not-superior performance, Sparse prior models lack in proper theoretical guidelines for selecting the best regularization parameter. In contrast, the quadratic regularization models can use well-established methods such as L-curve criterion, Hansen (1998) and the Generalized Cross Validation criterion, Hansen (1998) for choosing the value of λ. Difficulties in selecting the optimal converging point in non-convex minimization techniques such as IRLS also is an issue.

According to the theoretical and experimental details provided above, we propose that if we could decouple image restoration and "Process artifact" handling, then the use of quadratic regularization models will result in more efficient and effective image restoration in comparison to Sparse and Laplacian prior models. The decoupling of image restoration and "Process" artifact handling could be achieved through techniques such as tiling, Liu & Jia (2008), which enables the uses of the efficient least squares regularization.

Thus, coming back to our problem formulation in section 3.1, we claim that:

1. Sparse prior models are not superior to quadratic regularization models in terms of performance in image restoration.
2. In terms of efficiency, Sparse prior models are significantly inferior to quadratic regularization models.
3. The performance through Sparse prior model increases over quadratic regularization models when boundary effects are not addressed and processing artifacts are not compensated for.
4. Quadratic regularization models provide the best image restoration for large images in terms of efficiency and effectiveness while they provide a good enough solution for other images when the boundary artifacts are taken care of.

Analyzing the above items further, if the improvements of the Sparse prior model are in artifact handling, not in image restoration, we can pose the following questions.

"Do more complicated prior models such as Sparse, which are asserted be better matched to natural images, actually help image restoration in terms of restoring natural image features?"

"If those complicated prior models hold no significant advantage, is it worth the effort spend on them compared to simple and efficient prior models which restore closer or better than those prior models?"

4. Likelihood model analysis

Different likelihood models in the prior model in (14) have been studied in various ways. The fact that most of these models are not justified with proper theoretical foundations encouraged us to analyze and understand the variations and the validity and accuracy of the (implicit) underlying assumptions, which could explain the different performances.

This investigation guides our development of a new scheme for the multiple image likelihood model described in section 4.1.2. The likelihood model analysis is carried out using this new model and the theoretical analysis is corroborated by the computational experiments detailed in section 4.3.

4.1 Likelihood models in image restoration

4.1.1 Likelihood model for a single image

In image restoration literature, the likelihood for a single image is defined by modeling the image noise (n) as a set of independently and identically distributed (i.i.d.) random variables following a Gaussian distribution for all pixels, which is given by

$$p(b|g) = \prod_{\ell_1=1}^{L_1} \prod_{\ell_2=1}^{L_2} N\big(n(\ell_1,\ell_2)|0,\sigma\big), \tag{23}$$

where $N(\cdot|\mu,\sigma)$ denotes a Gaussian distribution with mean μ and variance σ^2, while L_1 and L_2 represent the image support.

4.1.2 Likelihood model for multiple images

Based on the above likelihood model for a single image, we develop a new model for the likelihood of multiple images as detailed below.

Given a set of R degraded images of a common GTI g, the posterior distribution for the GTI can be derived by extending (14), resulting in

$$p(g|b_1,b_2,\ldots,b_R) \propto p(b_1,b_2,\ldots,b_R|g)\,p(g), \tag{24}$$

where, generalizing (1),

$$b_r = \mathcal{K}_r g + n_r, \quad r = 1,2,3,\ldots,R \tag{25}$$

and $\mathcal{K}_r$ are operators representing possibly different but known blurs, and n_r are noise images. Under the assumption that n_p is independent of n_q for all $p \neq q$, the likelihood in (24) is

$$p(b_1,b_2,\ldots,b_R|g) = \prod_{r=1}^{R} N(n_r). \tag{26}$$

Thus, for a group of R images satisfying the noise independency condition in (26), the likelihood can be modeled as

$$p(b_1,b_2,\ldots,b_R|g) = \prod_{r=1}^{R} \prod_{\ell_1=1}^{L_1} \prod_{\ell_2=1}^{L_2} N\big(n_r(\ell_1,\ell_2)|0,\sigma_r\big), \tag{27}$$

where σ_r represent the standard deviation of the Gaussian distribution for n_r. This new model for the likelihood of multiple images will be used for the analysis of likelihood models in the next section.

r	∂_r	$\partial_r n$	$\omega(\partial_r)$
1	∂_1	$\partial_x n$	1
2	∂_2	$\partial_y n$	1
3	∂_3	$\partial_{xx} n$	2
4	∂_4	$\partial_{xy} n$	2
5	∂_5	$\partial_{yy} n$	2

Table 7. An example of set Θ in (30) with $R = 5$

4.1.3 Likelihood models for analysis

Out of the various likelihood models introduced in the literature of image restoration, we consider two recent approaches in Levin et al. (2007) and, Shan et al. (2008) for our analysis.

In Levin et al. (2007), the single image likelihood conforms to (23) and is explicitly given by

$$p(b|g) \propto e^{-\frac{1}{2\sigma^2}\|\mathcal{K}g - b\|^2}, \tag{28}$$

where $\|\cdot\|$ stands for the Frobenius norm.

In Shan et al. (2008), the likelihood is defined with different orders of partial derivatives, denoted by operator ∂_r, of a single degraded image. For ease of understanding, we represent their model in the form

$$p(b|g) = \prod_{\partial_r \in \Theta} \prod_{\ell_1=1}^{L_1} \prod_{\ell_2=1}^{L_2} N\big(n(\ell_1, \ell_2)|0, \sigma\big) \, N\big(\partial_r n(\ell_1, \ell_2)|0, \sigma_{\omega(\partial_r)}\big), \quad r = 1, 2, 3, \ldots, R \tag{29}$$

where Θ is a set of partial derivative operators given by

$$\Theta \triangleq \{\partial_1, \partial_2, \partial_3, \ldots, \partial_R\}. \tag{30}$$

For example, in Shan et al. (2008), the set Θ has the elements $\{\partial_x, \partial_y, \partial_{xx}, \partial_{xy}, \partial_{yy}\}$, in which, ∂_x is the first order derivative in x direction and ∂_y is the first order derivative in y direction and similar interpretations hold for higher order derivatives.

Further, Shan et al. (2008) shows that the partial derivatives of n also follow Normal distributions with standard deviation values based on the order of the partial derivative operator. The standard deviations of the partial derivatives are specified in the form

$$\sigma_{\omega(\partial_r)} = (\sqrt{2})^{\omega(\partial_r)}\sigma, \tag{31}$$

where $\omega(\partial_r)$ represents the order of the partial derivative operator ∂_r. Few example elements of the set Θ in (30) with the respective standard deviation values are given in Table 7.

As there was no analysis presented behind using the higher order partial derivatives of noise in Shan et al. (2008) leading to (29), we provide an interpretation of formula, based on our new general likelihood model for a group of degraded images of a common ground truth g in (27).

Guided by the likelihood expression (27), we can define a virtual group of images for the likelihood model in (29) as

$$b_r = \partial_r b, \quad r = 1, 2, 3, \ldots, R \tag{32}$$

and in order to align with model (25), define

$$\mathcal{K}_r g \triangleq \partial_r(g * k), \quad r = 1, 2, 3, \dots, R \tag{33}$$

$$n_r \triangleq \partial_r n, \quad r = 1, 2, 3, \dots, R \tag{34}$$

where $*$ stands for the convolution operator and k is the blur kernel.

From this, we infer that the likelihood (29) implicitly assumes $\partial_p n$ is independent of $\partial_q n$ for all $p \neq q$. Since all virtual images are derived from a single degraded image, we can infer this is a strong assumption made to simplify the likelihood expression. In principle, it should be possible to formulate a model without recourse to the derivative images which add limited new information. We corroborate this claim in section 4.3 with experiments.

4.2 Frequency domain deconvolution

In this section we approach image deconvolution with FOPDO regularization and with different likelihood models discussed above. For our analysis, we consider the likelihood models of (23) and (29) using terminology "normal likelihood" and "derivative likelihood" with the notation using subscripts "n" and "d", respectively. With our experiments, we limit the set Θ in (30), going up to second order partial derivative operators in (29) and we take elements of Θ from the following values

$$\Theta = \{\partial_x, \partial_y, \partial_{xx}, \partial_{xy}, \partial_{yy}\}. \tag{35}$$

4.2.1 Normal likelihood deconvolution

Under FOPDO regularization as detailed in section 2.2, the stabilizing functional $\Phi(g)$ takes the form

$$\Phi(g) \triangleq \|\partial_x g\|^2 + \|\partial_y g\|^2. \tag{36}$$

Applying this stabilizing functional to the MAP framework detailed in section 4, the energy functional under "normal likelihood", can be derived as

$$E_n(g) = \|g * k - b\|^2 + \lambda \, \Phi(g). \tag{37}$$

According to the convolution theorem, the convolution operation in the spatial domain becomes an element-wise product in the frequency domain making $\mathcal{F}(g * k) = G \star K$ where $\mathcal{F}(\cdot)$ stands for discrete Fourier transform, G for $\mathcal{F}(g)$, K for $\mathcal{F}(k)$ and "$\star$" denotes element-wise product. Based on the above property, transforming (37) into frequency domain and applying Plancherels theorem, Bracewell & Kahn (1966), we derive the energy in the frequency domain for (37) as follows.

$$\mathcal{F}(E_n(g)) = \|G \star K - B\|^2 + \lambda \, \mathcal{F}(\Phi(g)), \tag{38}$$

where

$$\mathcal{F}(\Phi(g)) = \|\mathcal{F}(\partial_x) \star G\|^2 + \|\mathcal{F}(\partial_y) \star G\|^2,$$

B stands for $\mathcal{F}(b)$ and given ∂_x takes the form of a (convolution) matrix, then $\mathcal{F}(\partial_x)$ denotes its Fourier transform.

Minimizing the energy in (38) and solving for estimated G denoted as $\widehat{G}$ results in

$$\widehat{G_n} = \frac{B \star \overline{K}}{K \star \overline{K} + \lambda \, \Psi},$$ (39)

where

$$\Psi = \mathcal{F}(\partial_x) \star \overline{\mathcal{F}(\partial_x)} + \mathcal{F}(\partial_y) \star \overline{\mathcal{F}(\partial_y)},$$

$\widehat{G_n}$ is the Fourier transform of the estimated GTI under "normal likelihood", $\overline{(\cdot)}$ stands for the complex conjugate and the division is performed element-wise. The estimated ground truth image $\widehat{g_n}$ can be derived by taking the inverse Fourier transform of $\widehat{G_n}$.

With the above derivations, it is evident that Fourier domain expression used to estimate the GTI is:

1. simple and leads to a closed form solution and
2. amenable to Fast Fourier Techniques leading to a highly efficient solution.

4.2.2 Derivative likelihood deconvolution

We now give an analogous derivation for the "derivative likelihood".

The energy functional in this case is derived similar to (37),

$$E_d(g) = \sum_{\partial_r \in \Theta} \frac{1}{2^{\omega(\partial_r)}} \|\partial_r (g * k) - \partial_r b\|^2 + \lambda \, \Phi(g).$$ (40)

Transforming (40) into the frequency domain results in

$$\mathcal{F}(E_d(g)) = \sum_{\partial_r \in \Theta} \frac{1}{2^{\omega(\partial_r)}} \left(\|\mathcal{F}(\partial_r) \star G \star K - \mathcal{F}(\partial_r) \star B\|^2 \right) + \lambda \, \mathcal{F}\big(\Phi(g)\big),$$ (41)

where, ∂_r is a matrix convolution operator representing a partial order derivative operator and $\mathcal{F}(\partial_r)$ denotes its Fourier transformation.

By minimizing the energy (41), we compute the estimated G,

$$\widehat{G_d} = \frac{B \star \overline{K} \star \Omega}{K \star \overline{K} \star \Omega + \lambda \, \Psi},$$ (42)

where

$$\Omega \triangleq \sum_{\partial_r \in \Theta} \frac{1}{2^{\omega(\partial_r)}} \mathcal{F}(\partial_r) \star \overline{\mathcal{F}(\partial_r)}.$$

By taking the inverse Fourier transforms of (42), we could get the estimated GTI, $\widehat{g_d}$, under "derivative likelihood" model similar to "normal likelihood" model.

λ	MSE	MSE
	Normal likelihood	Derivative likelihood
$\times 10^{-5}$	$\times 10^{-4}$	$\times 10^{-4}$
100	2.8448	4.4371
5	2.4148	**4.0862**
0.25	**2.2292**	4.1852

Table 8. Comparison of likelihood models

4.3 Likelihood model analysis

In order to come up with the most effective and efficient restoration algorithm, we investigate the contribution of each of the likelihood models for estimating the GTI: (39) corresponding to "normal likelihood" and (42) corresponding to "derivative likelihood", respectively.

We used the same "Picasso" image which was used in Shan et al. (2008) for experiments using the likelihood model in (29). The ground truth images are estimated using the Fourier domain techniques, specifically applying (39) and (42) for the "Normal" and "Derivative" likelihood models respectively. The experiment results are given in Table 8. In order to eliminate the "Model" and "Process" artifacts as discussed in section 3.3.2, in all our simulations, the blurring was carried out under the assumption that the images are periodic.

The MSE values in the table are given as multiples of 10^{-4}, while the value of λ is given in multiples of 10^{-5}. The values in bold in Table 8 refer to the optimal MSE values the respective likelihood model could reach for varying λ. As the results show clearly, the "normal likelihood" model has a better estimate for the GTI than the "derivative likelihood" model, we claim that applying "normal likelihood" in the image restoration algorithm results in a better restoration.

Our investigation was further extended to analyze whether higher order derivatives of noise contribute to the spatial randomness of noise as claimed in Shan et al. (2008). The noise maps given in Fig. 11 are computed for different values of λ in (39) and (42).

As per the results Fig. 11(c) and Fig. 11(d), when the effect of the prior becomes smaller (i.e., the weight on the data fitting term or the likelihood becomes larger), the noise estimate is more spatially random, but with the increase in the weight of the prior, the noise estimate becomes structured (signal dependant), see Fig. 11(e) and Fig. 11(f). We experienced these results regardless of the likelihood model we used. Based on the above results, we claim that using higher order partial derivatives in the likelihood model for non-blind deconvolution does not result in a better noise map estimation while the same noise map estimation can be achieved through the normal likelihood model with the appropriate Lagrange multiplier.

Hence, through the likelihood model analysis based on benchmark image, we conclude that higher order derivatives in the likelihood model are not required for better performance whereas applying single image likelihood model with appropriate regularization results in a more effective non-blind image restoration.

(a) Ground Truth Image

(b) 19×27 Blur kernel scaled upwards

(c) Normal likelihood ($\lambda = 0.001$)

(d) Derivative likelihood ($\lambda = 0.001$)

(e) Normal likelihood ($\lambda = 0.8$)
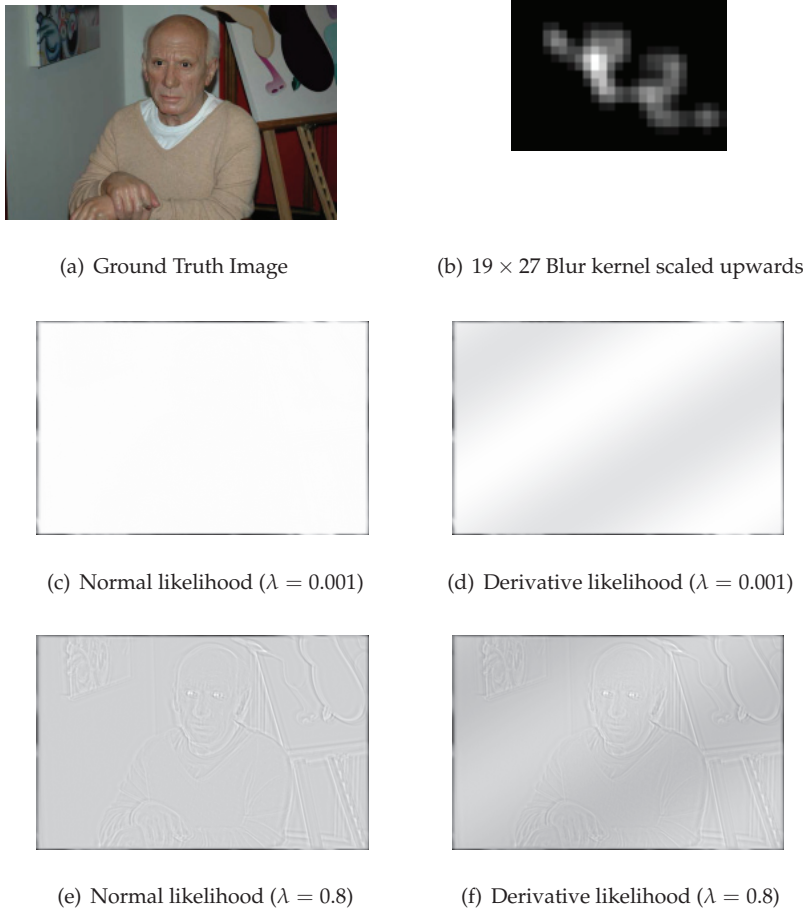

(f) Derivative likelihood ($\lambda = 0.8$)

Fig. 11. Noise maps for Likelihood models

5. Contributions

In this chapter, we have contributed to regularization based image restoration techniques in the following:

1. We have developed a general class of quadratic regularization models based on partial derivative operators (PDO), section 2.2. Out of those models, we have shown that the Second Order Partial Derivative Operator (SOPDO) model performs better than First Order Partial Derivative Operator (FOPDO) model for images susceptible to noise, while the novel First and Second Order Partial Derivative Operator (FSOPDO) model performs better than both FOPDO and SOPDO models.

2. We have used the Structured Similarity index (SSIM) map, Mean SSIM (MSSIM) value and histograms of SSIM maps as novel visual metrics for comparison and evaluation of regularization models in image restoration, section 3.3.2.

3. We have critically evaluated Sparse and Laplacian prior models against Quadratic regularization models using the novel visual metrics discussed in section 3.5. By eliminating the effects of processing and modeling artifacts, not present when capturing actual blurred natural images, we have shown that Sparse and Laplacian derivative prior models, which are claimed to be consistent with natural images, do not significantly contribute in restoring natural image features and have significantly slower relative restoration performance.

4. Finally, we have analyzed and evaluated multiple derivative operator based restoration methods under MAP/ML framework with a novel model to represent the likelihood based on multiple images, section 4.1.2. By using this novel model, we demonstrate that complex higher order derivative likelihood models are not required for better performance in image restoration.

6. References

Albert, A. E. (1972). *Regression and the Moore-Penrose Pseudoinverse*, Mathematics in science and engineering, New York : Academic Press.

Banham, M. R. & Katsaggelos, A. K. (1997). Digital image restoration, *IEEE Signal Processing Magazine* 14(2): 24–41.

Barrett, R., Berry, M., Chan, T. F., Demmel, J., Donato, J. M., Dongarra, J., Eijkhout, V., Romine, R. P. C. & Vorst, H. V. D. (1994). *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*, Society for Industrial Mathematics.

Bracewell, R. & Kahn, P. B. (1966). The Fourier transform and its applications, *American Journal of Physics* .

Demoment, G. (1989). Image reconstruction and restoration: overview of common estimation structures and problems, *IEEE Transactions on Acoustics, Speech and Signal Processing* 37(12): 2024–2036.

Fergus, R., Singh, B., Hertzmann, A., Roweis, S. T. & Freeman, W. T. (2006). Removing camera shake from a single photograph, *ACM Trans. Graph.* 25(3): 787–794.

Geman, S. & Geman, D. (1984). Stochastic relaxation, gibbs distributions and the bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6(6): 721–741.
    URL: *http://dx.doi.org/10.1080/02664769300000058*

Greig, D. M., Porteous, B. T. & Seheult, A. H. (1989). Exact maximum a posteriori estimation for binary images, *Journal of the Royal Statistical Society. Series B (Methodological)* 51(2): 271–279.

Hadamard, J. (1952). *Lectures on Cauchy's problem in linear partial differential equations*, New Haven, CT: Yale University Press.

Hansen, P. C. (1998). *Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion*, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.

Haykin, S. (1994). *Blind Deconvolution*, Prentice-Hall: Englewood Cliffs, NJ.

Hunt, B. R. (1971). Biased estimation for nonparametric identification of linear systems, *Mathematical Biosciences* 10(3-4): 215–237.

Hunt, B. R. (1973). The application of constrained least squares estimation to image restoration by digital computer, *IEEE Transactions on Computers* 22(9): 805–812.

Hunt, B. R. (1977). Bayesian methods in nonlinear digital image restoration, *IEEE Transactions on Computers* 26: 219–229.

Hunt, B. R. & Andrews, H. C. (1977). *Digital Image Restoration*, Prentice Hall Professional Technical Reference.

Joshi, N., Szeliski, R. & Kriegman, D. J. (2008). PSF estimation using sharp edge prediction, *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2008*, pp. 1–8.

Kim, S. J., Koh, K., Lustig, M. & Boyd, S. (2007). An efficient method for compressed sensing, *International Conference on Image Processing, ICIP 2007*, Vol. 3, pp. 117–120.

Krahmer, F., Lin, Y., McAdoo, B., Ott, K., Wang, J. & Widemann, D. (2006). Blind image deconvolution: motion blur estimation, *University of Minnesota, URL: http://www. ima. umn. edu* .

Kundur, D. & Hatzinakos, D. (1996). Blind image deconvolution, *IEEE Signal Processing Magazine* 13(3): 43–64.

Leung, C. M. & Lu, W. S. (1995). On the use of discrete Laplacian operators in image restoration, *IEEE Pacific Rim Conference on Communications, Computers, and Signal Processing*, pp. 411–415.

Levin, A. (2007). Blind motion deblurring using image statistics, *Advances in Neural Information Processing Systems* 19: 841–848.
URL: *http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.110.2995*

Levin, A., Fergus, R., Durand, F. & Freeman, W. T. (2007). Image and depth from a conventional camera with a coded aperture, *ACM Transactions on Graphics (Proceedings of SIGGRAPH)* 26(3): 70–79.

Levin, A. & Weiss, Y. (2007). User assisted separation of reflections from a single image using a sparsity prior, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29(9): 1647–1654.

Levin, A., Weiss, Y., Durand, F. & Freeman, W. T. (2009). Understanding and evaluating blind deconvolution algorithms, *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009*, pp. 1964–1971.

Liu, R. & Jia, J. Y. (2008). Reducing boundary artifacts in image deconvolution, *IEEE International Conference on Image Processing, ICIP'08*, Vol. 1-5, pp. 505–508.

Meer, P. (2004). Robust techniques for computer vision, *Emerging topics in computer vision* pp. 107–190.

Mignotte, M. (2006). A segmentation-based regularization term for image deconvolution, *IEEE Transactions on Image Processing* 15(7): 1973–1984.

Miller, K. (1970). Least squares methods for ill-posed problems with a prescribed bound, *SIAM Journal on Mathematical Analysis* 1(1): 52–74.
URL: *http://link.aip.org/link/?SJM/1/52/1*

Nayar, S. K. & Ben-Ezra, M. (2004). Motion-based motion deblurring, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26(6): 689–698.

Oliveira, J. P., Figueiredo, M. A. & Bioucas-Dias, J. M. (2007). Blind estimation of motion blur parameters for image deconvolution, *IbPRIA '07: Proceedings of the 3rd Iberian conference on Pattern Recognition and Image Analysis, Part II*, Springer-Verlag, Berlin, Heidelberg, pp. 604–611.

Olshausen, B. A. et al. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images, *Nature* 381(6583): 607–609.

Phillips, D. L. (1962). A technique for the numerical solution of certain integral equations of the first kind, *Journal of the ACM (JACM)* 9(1): 84–97.

Portilla, J., Strela, V., Wainwright, M. J. & Simoncelli, E. P. (2003). Image denoising using scale mixtures of Gaussians in the wavelet domain, *IEEE Transactions on Image processing* 12(11): 1338–1351.

Roth, S. & Black, M. (2005). Fields of experts: A framework for learning image priors, *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2*, IEEE Computer Society, Washington, DC, USA, pp. 860–867.

Shan, Q., Jia, J. & Agarwala, A. (2008). High-quality motion deblurring from a single image, *International Conference on Computer Graphics and Interactive Techniques SIGGRAPH 2008*, Vol. 27, ACM, New York, NY, USA, pp. 1–10.

Simoncelli, E. P. (1997). Statistical models for images: compression, restoration and synthesis, *Conference Record of the Thirty-First Asilomar Conference on Signals, Systems and Computers*, Vol. 1, pp. 673–678.

Simoncelli, E. P. (1999). Bayesian denoising of visual images in the wavelet domain, *Lecture Notes in Statistics* 141: 291–308.

Tikhonov, A. N. & Arsenin, V. Y. (1977). *Solutions of ill-posed problems*, John Wiley, New York.

Wang, Z. & Bovik, A. C. (2009). Mean squared error: love it or leave it? - A new look at signal fidelity measures, *IEEE Signal Processing Magazine* 26(1): 98–117.

Wang, Z., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity, *IEEE Transactions on Image Processing* 13(4): 600–612.

Weiss, Y. (2001). Deriving intrinsic images from image sequences, *9th International Conference on Computer Vision, ICCV*, IEEE Computer Society, pp. 68–75.

Wiener, N. (1942). Extrapolation, Interpolation, and Smoothing of Stationary Time Series.

Zhu, S. C. & Mumford, D. (1997). Prior learning and Gibbs reaction-diffusion, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19(11): 1236–1250.