

# Recent Progress in Image Deblurring

Ruxin Wang      Dacheng Tao

Centre for Quantum Computation & Intelligent Systems

Faculty of Engineering & Information Technology

University of Technology Sydney

81-115 Broadway, Ultimo, NSW

Australia

Email: Ruxin.Wang@student.uts.edu.au, dacheng.tao@uts.edu.au

11 August 2014

## Abstract

This paper comprehensively reviews the recent development of image deblurring, including non-blind/blind, spatially invariant/variant deblurring techniques. Indeed, these techniques share the same objective of inferring a latent sharp image from one or several corresponding blurry images, while the blind deblurring techniques are also required to derive an accurate blur kernel. Considering the critical role of image restoration in modern imaging systems to provide high-quality images under complex environments such as motion, undesirable lighting conditions, and imperfect system components, image deblurring has attracted growing attention in recent years. From the viewpoint of how to handle the ill-posedness which is a crucial issue in deblurring tasks, existing methods can be grouped into five categories: Bayesian inference framework, variational methods, sparse representation-based methods, homography-based modeling, and region-based methods. In spite of achieving a certain level of development, image deblurring, especially the blind case, is limited in its success by complex application conditions which make the blur kernel hard to obtain and be spatially variant. We provide a holistic understanding and deep insight into image deblurring in this review. An analysis of the empirical evidence for representative methods, practical issues, as well as a discussion of promising future directions are also presented.

## 1 Introduction

Modern imaging sciences, such as consumer photography, astronomical imaging, medical imaging, and microscopy, have been well developed in recent years and a large number of progressive techniques have emerged. These developments have enabled the acquisition of images that are of both higher speed and higher resolution (also referred to as *high-definition*). However, intrinsic or extrinsic factors behind such fast and high-resolution techniques may lead to degradation in the quality of the acquired image, of which blur is one example and is the focus of this paper. From an artistic viewpoint, image blur in photography is sometimes intentional, but in the most common imaging situations, the blur effect corrupts valuable image information and produces visually unattractive images. For example,

- 1) A fast moving car captured by a surveillance system might exhibit significant blurriness in the video or image, leading to difficulty in the license-plate recognition (*motion blur*).
- 2) It is often difficult for photographers to stabilize hand-held cameras for a long period, especially when they are taking pictures in dim lighting conditions which require long exposure times, resulting in blurred images (*camera shake blur*).

- 3) Since most imaging systems have only one focus, the resultant images usually have at most one region that is focused while the others remain blurred (*defocus blur*).
- 4) When capturing a long distance scene, atmospheric particulate matter sometimes prevents photons from moving directly to the sensor, which produces a blurred image (*atmospheric turbulence blur*).
- 5) When a lens has a different refractive index for different wavelengths of light, the lens can fail to focus all colors to the same convergence point which also results in a blurred image (*intrinsic physical blur*).

These blurry image data are a nuisance in a variety of high-quality image-based applications, e.g. image content recognition (Nishiyama et al 2011), medical diagnosis and surgery (Tzeng et al 2010), surveillance monitoring, remote sensing (Ma and Le Dimet 2009), and astronomy. Therefore, reducing such blur, which is known as *image deblurring* or *image deconvolution*, is a crucial step in improving the resolution and contrast of high-quality images.

Image deblurring is a traditional inverse problem whose aim is to recover a sharp image from the corresponding degraded (blurry and/or noisy) image. Over the years, numerous methods have been proposed to tackle the *non-blind deblurring problems* or the *blind deblurring problems*, in which classic and well-known classification schemes are employed. The former case, non-blind deblurring, indicates that the blur kernel is assumed to be known and a sharp image can be induced from both the blurry image and the kernel. Typical methods include the Richardson-Lucy method (Richardson 1972; Lucy 1974) and Wiener filter (Wiener 1949). By contrast, the blind deblurring problem, which is more practical, means that the blur kernel is unknown, and the task therefore becomes one of estimating the sharp image and/or the kernel from the degraded image. This kind of method dates back to the 1970s (Stockham Jr et al 1975; Cannon 1976). In recent years, many novel approaches have been presented to handle both the non-blind and blind deblurring problem, driven by a variety of motivations. The above simple classification scheme is not sufficiently competent to discover the connections and details of modern image deblurring methods, thus we intend in this survey to manage the organization of this domain through an analysis of ill-posedness. Ill-posedness is the most severe problem in image deblurring. In the non-blind case, the observed blurry image does not uniquely and stably determine the sharp image due to the ill-conditioned nature of the blur operator (Bertolo and Boccacci 1998). This means that if the assumed/given blur kernel and the true kernel are slightly mismatched, or if the blurry image is also corrupted by noise, the recovered image may be much worse than the underlying sharp image. In the blind case, even if the blur operator is not ill-conditioned, the deblurring problem will still be inherently ill-posed since there is an infinite set of image-blur pairs that can synthesize the observed blurry image.

Contemporary researchers have been devoted to developing new models and new prototypes, or improving the efficiency of optimization methods, to deal with ill-posedness. From the model construction perspective, most methods can be grouped into the following five categories: *Bayesian inference framework*, *variational methods*, *sparse representation-based methods*, *homography-based modeling*, and *region-based methods*. In the Bayesian inference framework, priors are introduced to impose uncertainty attributes on either the unknown sharp image or the unknown blur kernel, or both. This operation is intended to reduce the volume of the search space, where the problem’s solution lies, to suppress the ill-posedness. Variational methods render the solution unique and stable through the incorporation of regularization techniques whose role is similar to the prior’s in Bayesian inference, i.e. regularizing the solution into a constraint space. Sparse representation, a progressive topic in recent years, benefits from the fact that natural images are intrinsically sparse in some domains. The sparse property of the natural images in these domains is well suited to tackle the ill-posedness of the deblurring problem. The homography-based modeling and region-based methods are intentionally proposed for spatially variant deblurring (see Section 2). Since spatially variant blur cannot be modeled by a single blur kernel with limited support, researchers usually approximate the blurry effect by using a union of multiple kernels or homographies. Due to the growing number of unknowns, this type of method lead to a worse situation in the inverse problem. To overcome this issue, homography-based methods derive the spatially variant kernel as an integration of "temporal" homographies whose continuity is further constrained according to the properties of the blur effects in the image. Meanwhile, region-based methods focus on the restriction of spatial variations of the blur kernels. Other methods which do not fall into the above frameworks include Projection-based method, kernel regression, stochastic deconvolution, and spectral analysis.

Besides the above mathematical viewpoint, a progressive direction in deblurring concerns the development of hardware prototype systems. Traditional camera systems are generally equipped with a single lens, single

focus, and single consecutive exposure. Using these systems, a blurred image is the only achievable resource, and a large amount of useful information is lost during imaging (Wehrwein and Scharstein 2010). However, by using hardware modifications, additional observations or principles can be easily accessed to help the derivation of either the sharp image or the blur kernel, resulting in reduced ill-posedness.

In reviewing the literatures on image deblurring, we have been attracted by a number of interesting discoveries. The one such discovery is that more and more approaches have been proposed to use multiple images to jointly deblur, or at least to assist the deblurring of a target image. These images may be either correlated or uncorrelated. Multiple image deblurring is possible because a set of common patterns exist behind natural scenes which could be used to generate all kinds of natural structures. One option to incorporate multiple images is to learn the deblurring functions or subspaces by using the acquirable sharp-blurry image pairs from additional datasets. Another option is hardware modification, as noted earlier. In terms of single image deblurring, another concern claims that the global solutions of certain models do not necessarily correspond to the true solution of the problem, and a more appropriate model should be discovered. Considering these findings, the possible future directions in the image deblurring field may be summarized as learning-based methods and hardware modifications. In fact, research in these areas has started and is flourishing. As well, new single image deblurring models are necessary yet challenging to be exploited to complement the drawbacks of existing models.

The rest of this paper is organized as follows: The formulation of image deblurring is first introduced in Section 2. Five classes of modeling methods are described and comprehensively analyzed in Sections 3-7 and other methods are listed in Section 8. Section 9 discusses several issues usually encountered in designing deblurring methods. Insights to promising future directions in this field are given in Section 10. The experimental analyses and evaluations are given in Section 11, and concluding remarks are made in Section 12.

## 2 Image Blur Formulation

The blur kernel, also known as (aka) *point spread function* (PSF), causes an image pixel to record light photons from multiple scene points. Many factors can extrinsically or intrinsically cause image blur. As briefed above, blur is generally one of five types: object motion blur, camera shake blur, defocus blur, atmospheric turbulence blur, and intrinsic physical blur.<sup>1</sup> These types of blur degrade an image in different ways. An accurate estimation of the sharp image and the blur kernel requires an appropriate modeling of the digital image formation process. Hence before introducing the blur types, we first focus on analyzing the image formation model.

Recall that image formation encompasses the radiometric and geometric processes by which a 3D world scene is projected onto a 2D focal plane. In a typical camera system<sup>2</sup>, light rays passing through a lens’s finite aperture are concentrated onto the focal point. This process can be modeled as a concatenation of the perspective projection and the geometric distortion. Due to the non-linear sensor response, the photons are transformed into an analog image, from which the final digital image is formed by discretization (Delbracio et al 2012).

Mathematically, the above process can be formulated as

$$y = S(f(D(P(s) * h_{ex}) * h_{in})) + n, \quad (2.1)$$

where  $y$  is the observed blurry image plane,  $s$  is the real planar scene,  $P(\cdot)$  denotes the perspective projection,  $D(\cdot)$  is the geometric distortion operator,  $f(\cdot)$  denotes the nonlinear camera response function (CRF) that maps the scene irradiance to intensity,  $h_{ex}$  is the extrinsic blur kernels caused by external factors such as motion,  $h_{in}$  denotes the blur kernels determined by intrinsic elements such as imperfect focusing,  $*$  is the mathematical operation of convolution,  $S(\cdot)$  denotes the sampling operator due to the sensor array, and  $n$  models the noise.

<sup>1</sup>Strictly speaking, both the object motion blur and the camera shake blur belong to motion blur, while the defocus blur is an instance of the intrinsic physical blur. Here we separate them to highlight the focus of different approaches.

<sup>2</sup>Different imaging systems correspond to different image formulation models. Here, we start with the camera system and end with the most common formulation for other systems: see equation (2.3).

The above process explicitly describes the mechanism of blur generation. However, what we are interested here is the recovery of a sharp image having no blur effect, rather than the geometry of the real scene. Hence, focusing on the image plane and ignoring the sampling errors, we obtain

$$y \approx f(x * h) + n, \quad (2.2)$$

where  $x$  is the latent sharp image induced from  $D(P(s))$ , and  $h$  is an approximated blur kernel combining  $h_{ex}$  and  $h_{in}$ , as assumed by most approaches. The effect of the CRF in this formulation, which will be discussed in Section 2.2, will have a significant influence on the deblurring process if it is not appropriately addressed. For simplification, most researchers neglect the effect of the CRF, or explore it as a preprocessing step. Let us remove the effect of the CRF to obtain a further simplified formulation:

$$y = x * h + n. \quad (2.3)$$

This equation is the most commonly-used formulation in image deblurring.

Given the above formulation, the general objective is to recover an accurate  $x$  (non-blind deblurring), or to recover  $x$  and  $h$  (blind deblurring), from the observation  $y$ , while simultaneously removing the effects of noises  $n$ . Taking into account a whole image, equation (2.3) is often represented as a matrix-vector form:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \quad (2.4)$$

where  $\mathbf{y}$ ,  $\mathbf{x}$  and  $\mathbf{n}$  are lexicographically ordered column vectors representing  $y$ ,  $x$  and  $n$ , respectively.  $\mathbf{H}$  is a Block Toeplitz with Toeplitz Blocks (BTTB) matrix derived from  $h$ .

While the noise may originate during image acquisition, processing, or transmission, and is dependent on the imaging system, term  $n$  is often modeled as Gaussian noise (Levin et al 2011b; Delbracio et al 2012), Poisson noise (Kenig et al 2010; Carlván and Blanc-Féraud 2012; Lefkimmatis and Unser 2013; Ma et al 2013) or impulse noise (Chan et al 2010; Cai et al 2010). Equation (2.3) is not suitable for describing these noises since it only characterizes the plus case and the signal-uncorrelated case. On the other hand, Poisson and impulse noise are usually signal-correlated. A more detailed summary of the specific degradation models with respect to different noise assumptions can be found in Table 1. In this paper, we prefer to consider Gaussian noise.

Table 1: Summary on different noise assumption

Noise Assumption	Image Degradation Model	Noise Distribution
Gaussian noise	$y = x * h + n$	$p(y x, h) = \prod_i \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y_i - (x*h)_i)^2}{2\sigma^2}\right)$
Poisson noise	$y = n(x * h)$	$p(y x, h) = \prod_i \frac{((x*h)_i)^{y_i} \exp(-(x*h)_i)}{y_i!}$
Impulse noise	$y = n(x * h)$	$p(y_i (x*h)_i) = \begin{cases} (x*h)_i, & \text{with probability } 1-r \\ n_{max}, & \text{with probability } r/2 \\ n_{min}, & \text{with probability } r/2 \end{cases}$
		$p(y_i (x*h)_i) = \begin{cases} (x*h)_i, & \text{with probability } 1-r \\ n_i, & \text{with probability } r \end{cases}$

Notes: All noises are assumed *i.i.d.* with respect to the image location  $\{i\}$ . In the Gaussian case, the noise is assumed to be of mean zero and variance  $\sigma^2$ . Impulse noise includes two types: salt-and-pepper noise (the upper model) and random-valued impulse noise (the lower model). In salt-and-pepper case,  $n_{max}$  and  $n_{min}$  denote the maximum and minimum of the intensity range, while in random-valued case  $n_i$  is sampled from a uniform distribution in  $[n_{min}, n_{max}]$ .

Traditionally, the blur kernel in image deblurring methods is usually assumed to be *spatially invariant* (aka *uniform* blur), which means that the blurry image is the convolution of a sharp image and a single kernel (Lucy 1974; Wiener 1949; Fergus et al 2006; Kenig et al 2010). However in practice, it has been noted that the invariance is violated by complex motion or other factors. Thus *spatially variant* blur (aka *non-uniform* blur) is more practical (Levin et al 2011b), but is hard to address. In this case, the matrix  $\mathbf{H}$  in equation (2.4) is no longer a BTTB matrix since different pixels in the image correspond to different kernels.



Figure 1: Object Motion Blur

The number of unknown variables in the blind deblurring problem is therefore significantly increased, but fortunately,  $\mathbf{H}$  can be characterized by the specific properties of the blur types. If we are given the specific motions or factors which cause the blur, the problem can be effectively constrained by prior knowledge. Next, we will introduce the attributes of the five blur types mentioned above.

## 2.1 Blur Types

### 2.1.1 Object Motion Blur

*Object motion blur* is caused by the relative motion between an object in the scene and the camera system during the exposure time. This type of blur generally occurs in capturing a fast-moving object or when a long exposure time is needed. If the motion is very fast relative to the exposure period, we may approximate the resultant blur effect as a linear motion blur, which is represented as a 1D local averaging of neighboring pixels and given by

$$h(i, j; L, \theta) = \begin{cases} \frac{1}{L}, & \text{if } \sqrt{i^2 + j^2} \leq \frac{L}{2} \text{ and } \frac{i}{j} = -\tan \theta, \\ 0, & \text{otherwise,} \end{cases} \quad (2.5)$$

where  $(i, j)$  is the coordinate originating from the center of  $h$ ,  $L$  the moving distance and  $\theta$  the moving direction. Fig. 1a gives a simulated example of a Lena standard test image corrupted by  $30^\circ$ -directional motion. In practice, however, real motions are extremely complex and cannot be approximated by such a simple parametric model. An appropriate way to handle this issue is to use the non-parametric model, i.e. no explicit shape constraint is imposed on the blur kernel, and the only assumption is that the kernel needs to follow the motion path. A more serious issue is in an image where only the region of moving objects is disturbed by the blur kernel, while other regions remain clear. As shown in Fig. 1b, the bus was moving fast while the surroundings were static when the picture was taken. In this case, we cannot uniformly process the whole image by a single kernel even if the kernel accurately represents the true motion (Levin 2006). Since moving objects (e.g. the bus) occupy parts of the image, the most commonly-used approach to handle this problem is to segment the blurry regions from the clear background (Chakrabarti et al 2010), which will be detailed in Section 7, or to simulate the motion as a sequence of homographies (Tai et al 2010b), which will be discussed in Section 6.

### 2.1.2 Camera Shake Blur

*Camera shake blur* is induced by camera motion during the exposure period. This is particularly common in handheld photography and in low light situations, e.g. inside buildings or at night. This blur, like object motion blur, can be very complex, since the hands may move in an irregular direction when photographing, causing the camera translation or rotation in-plane or out-of-plane (Whyte et al 2010, 2012). An ideal situation exists if the camera is only slightly translated when capturing a long distance scene and the resultant blur is approximately spatially invariant, and can be modeled as linear motion blur in equation (2.5). However this invariance will be violated when the camera undergoes significant translation or rotation during the exposure. In Fig. 2a, for example, the flowers are at different distances from the focal plane,

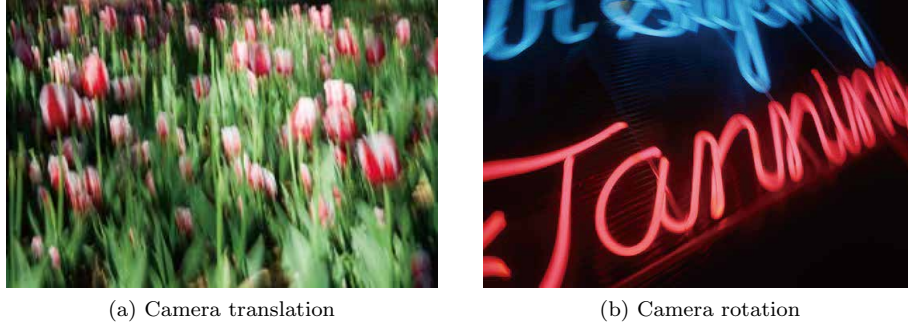


Figure 2: Camera Shaken Blur

implying that during the camera’s translation, the nearer flowers undergo a large shift with respect to the focal plane, while the distant flowers experience a slight shift. Camera rotation is a more complicated case which includes in-plane rotation and out-of-plane rotation in terms of focal plane. In the case of in-plane rotation, the blur kernel varies significantly across the image, especially for the regions far from the axis along which the camera is rotated, as shown in Fig. 2b. For out-of-plane rotation, the degree of spatial variance across the image is dependent on the focal length of the camera (for more detail see (Whyte et al 2012)). Clearly, the spatially variant blur is more suitable to explain all cases mentioned above than the spatially invariant blur. Fortunately, as Harmeling et al (2010) point out, it is reasonable to expect that the blur kernel will vary smoothly across the whole image if the depth of the scene varies smoothly, in spite of the fact that this is not true when the objects have very different distances from the focal plane.

### 2.1.3 Defocus Blur

As a result of imperfect focusing by the imaging system or different depths of scene, the fields outside the focus field are defocused, giving rise to *defocus blur*, or *out of focus blur*. This blur is familiar in our everyday photos. For example, it is often hard for a photography beginner to focus on the target object by hand. Also, when a camera is equipped with only one lens, scenes outside the *depth of field* (DOF, which is the range covered by all objects in a scene that appear acceptably sharp in an image) are all blurred in the resultant image, e.g. Fig. 3b. Traditionally, a crude approximation of a defocus blur is made as a uniform circular model:

$$h(i, j) = \begin{cases} \frac{1}{\pi R^2}, & \text{if } \sqrt{i^2 + j^2} \leq R, \\ 0, & \text{otherwise,} \end{cases} \quad (2.6)$$

where  $R$  is the radius of the circle. This is valid if the depth of scene does not have significant variation and  $R$  is properly selected. A simulated instance is shown in Fig. 3a. Practically, focusing on a target object is not difficult for modern consumer cameras since most of them are equipped with an auto-focus function; however, due to the limited DOF, it is not always possible to make the entire image sharp, as illustrated in Fig. 3b. To recover a full-focused image, the *focus sweep* technique is usually utilized to sweep the plane of focus through a desired depth range during exposure, so that the depth of field is enlarged (Bando et al 2013). An alternative way of handling this limit is to use coded aperture pairs (Zhou et al 2011), by which the depth of the scene can be also recovered.

### 2.1.4 Atmospheric Turbulence Blur

*Atmospheric turbulence blur* generally happens in long-distance imaging systems such as remote sensing and aerial imaging. This is mainly because of the randomly varying refractive index along the optical transmission path. For long-term exposure through the atmosphere, the blur kernel can be described by a fixed Gaussian model (Zhu and Milanfar 2010), i.e.

$$h(i, j) = Z \cdot \exp\left(-\frac{i^2 + j^2}{2\sigma^2}\right), \quad (2.7)$$



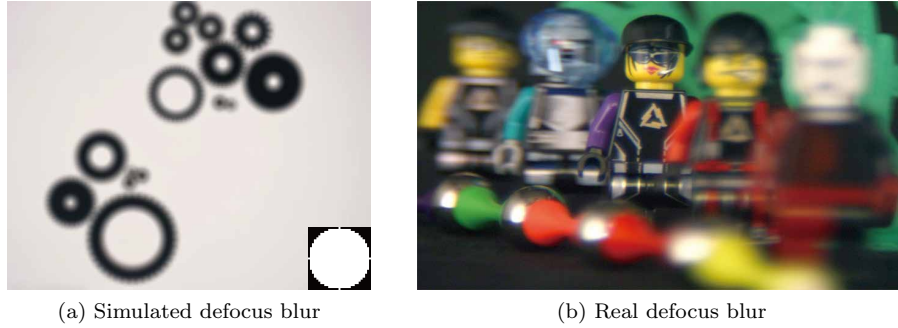


Figure 3: Defocus Blur

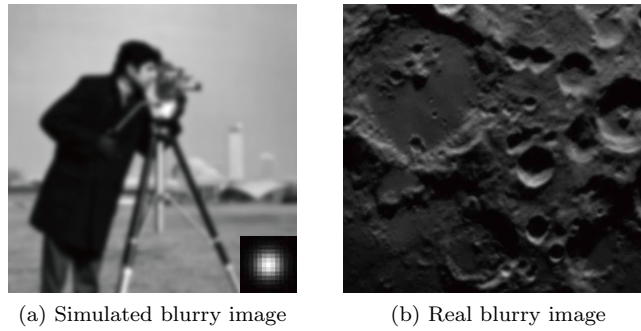


Figure 4: Atmospheric Turbulence Blur

where  $\sigma$  encodes the size of the kernel and  $Z$  is the normalizing constant ensuring that the blur has a unit volume. It has been noted that the above equation is impracticable and in fact, this blur is a mixture of multiple degradations like geometric distortion, spatially and temporally variant defocus blur, and possibly motion blur (Hirsch et al 2010; Zhu and Milanfar 2011, 2013). Fig. 4 illustrates a blurry cameraman image degraded by the kernel in (2.7), as well as a real image of Moon Surface from (Zhu and Milanfar 2013).

### 2.1.5 Intrinsic Physical Blur

*Intrinsic physical blur* is inherent in a number of imaging systems, due largely to intrinsic factors such as light diffraction, lens aberration, sensor resolution, and anti-aliasing filters. An important instance of this type is the optical aberration. In an ideal optical system, all rays of light from a point in the real world will converge to the same point on the focal plane, generating a sharp image. In reality, however, any departure of an optical system from this principle is called an optical aberration. For a real system with spherical optics, it is unrealistic that the light rays from a point source are all parallel with the optical axis, leading to monochromatic aberration, which is a branch of optical aberration. Another branch is chromatic aberration. Since lenses have different refractive indexes for different wavelengths of light, it is difficult to focus all colors on the same convergence point. Generally, the acquired image is corrupted by various optical aberrations, and the induced blur is spatially variant (Schuler et al 2012). See Fig. 5 for an example. To address this type of blur, additional calibration techniques are introduced to assist the estimation of the blur, such as the utilization of Poisson noise pattern (Delbracio et al 2012), or the checkerboard test chart (Kee et al 2011). Schuler et al (2011) proposed that all optical aberrations could be corrected by digital image processing, and Tzeng et al (2010) exploited the specific property of the fluidic lens camera system in which one of three color planes remains sharp in the imaging process.



Figure 5: Optical Aberration

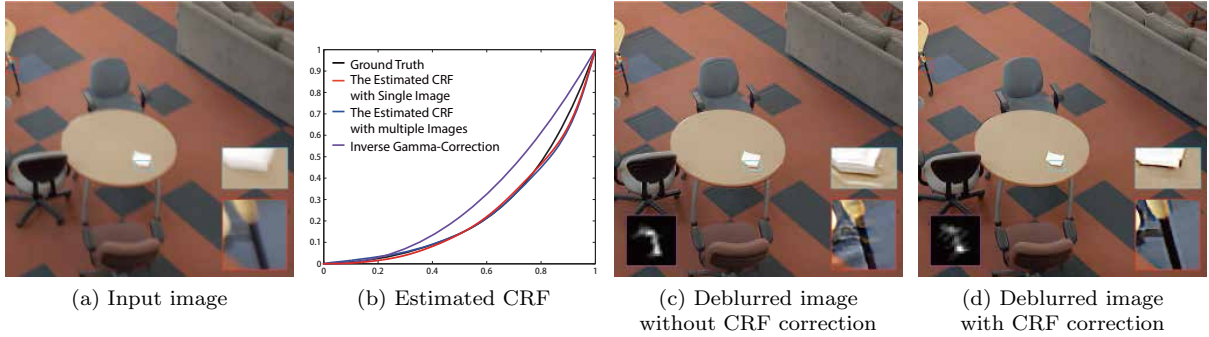


Figure 6: Effects of the nonlinear CRF on deblurring results (Kim et al 2012b)

## 2.2 Effects of the CRF $f(\cdot)$

Recall that the *camera response function*  $f(\cdot)$  in equation (2.2) is a nonlinear model mapping the scene irradiance to image intensity. The purpose of this design is to mimic the response of the human visual system or to compress the dynamic range of scenes, or for aesthetic consideration (Grossberg and Nayar 2004; Lin and Zhang 2005). However, inappropriate processing of CRF leads to severe ringing artifacts in deblurring problems. Chen et al (2012) have theoretically pointed out three kinds of CRF effects on blur inconsistency  $\Delta = f(\tilde{x} * h) - x * h$ . Here  $\tilde{x}$  is the sharp irradiance image corresponding to  $x$ . These effects are:

- 1) The nonlinear CRF has no influence on the intensity of the uniform regions, i.e.  $\Delta = 0$ ;
- 2) In low-frequency regions, the irradiance is approximately equal to its corresponding intensity, i.e.  $\Delta \approx 0$ , under the condition of small kernel  $h$  and smooth function  $f(\cdot)$ ;
- 3) In high-frequency high contrast regions, the nonlinearity of  $f(\cdot)$  can significantly damage the blur consistency, particularly when the local minimum and maximum pixel intensities are very different.

From these claims, we readily find that even if a spatially invariant blur  $h$  is assumed to be in irradiance,  $f(\cdot)$  could turn  $h$  into a spatially variant case in the intensity domain (Kim et al 2012a). To estimate the function, given a blurry/sharp image pair, Chen et al (2012) exploited Generalized Gamma Curve Model (GGCM) to fit  $f(\cdot)$ . Kim et al (2012a) and Tai et al (2013) addressed the estimation in two ways, one of which is based on a least-squares formulation when the blur kernel is known, while the other is solved via rank minimization without the need for a known kernel. Fig. 6 illustrates a result of (Kim et al 2012a,b), indicating that large number of artifacts remain in the selected regions of the deblurred image without the CRF correction.

In summary, each of the five blur types discussed in this section has specific properties, inspiring researchers to develop effective and efficient algorithms for deblurring. Needless to say, a good deblurring model should be suitable for all blur types which, however, is difficult to be explored. Fortunately, in the deblurring community, there have been many progressive methods proposed in recent years. In the following sections, we will discuss the five categories of existing methods from the modeling perspective in detail. First let us focus on Bayesian inference framework.



### 3 Bayesian Inference Framework

In statistics, Bayesian inference updates the states of a probability hypothesis by exploiting additional evidence. Bayes' rule is the critical foundation of Bayesian inference and can be expressed as

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)}, \quad (3.1)$$

where  $A$  stands for the hypothesis set and  $B$  corresponds to the evidence set. This rule states that the true posterior probability  $p(A|B)$  is based on our prior knowledge of the problem, i.e.  $p(A)$ , and is updated according to the compatibility of the evidence and the given hypothesis, i.e. the likelihood  $p(B|A)$ . In our scenario for the non-blind deblurring problem,  $A$  is then the underlying sharp image  $x$  to be estimated, while  $B$  denotes the blurry observation  $y$ . For the blind case, a slight difference is that  $A$  means the pair of  $(x, h)$  since  $h$  is also a hypothesis in which we are interested. Explicitly (3.1) can be written for both cases as

$$\text{Non-blind:} \quad p(x|y, h) = \frac{p(y|x, h)p(x)}{p(y)}, \quad (3.2)$$

$$\text{Blind:} \quad p(x, h|y) = \frac{p(y|x, h)p(x)p(h)}{p(y)}. \quad (3.3)$$

Note that either  $x$  or  $y$  and  $h$  are usually assumed to be uncorrelated. Irrespective of case, the likelihood  $p(y|x, h)$  is dependent on the noise assumption, as listed in Table 1. How to further infer the equation (3.1) in the literatures inspires us to explore three directions: maximum a posteriori, minimum mean square error, and variational Bayesian methods.

#### 3.1 Maximum a Posteriori

The most commonly-used estimator in a Bayesian inference framework is the *maximum a posteriori* (MAP). This strategy attempts to find the optimal solution  $A^*$  which maximizes the distribution of the hypothesis set  $A$  given the evidence set  $B$  in (3.1). In the blind case,

$$\begin{aligned} (x^*, h^*) &= \operatorname{argmax}_{x, h} p(x, h|y) \\ &= \operatorname{argmax}_{x, h} p(y|x, h)p(x)p(h), \end{aligned} \quad (3.4)$$

while in the non-blind scenario, the term  $p(h)$  is discarded according to equation (3.2). Fig. 7 shows a diagram of this type of method.

##### 3.1.1 Maximum Likelihood Estimation

We start with the introduction of a classic non-blind algorithm, Richardson-Lucy (RL) deconvolution (Richardson 1972; Lucy 1974), which is widely used in astronomical imaging and medical imaging.<sup>3</sup> Assuming that the prior  $p(x)$  takes the form of a uniform distribution, the MAP estimator then becomes a maximum likelihood estimator (MLE), i.e.

$$\begin{aligned} x^* &= \operatorname{argmax}_x p(y|x, h) \\ &= \operatorname{argmin}_x -\log p(y|x, h), \end{aligned} \quad (3.5)$$

where in the second equation the minimized equation is known as negative log-likelihood. Under the noise assumption of Poisson distribution (Shepp and Vardi 1982), (3.5) is derived as

$$x^* = \operatorname{argmin}_x \sum_i ((x * h)_i - y_i \log(x * h)_i). \quad (3.6)$$

---

<sup>3</sup>In the standard RL, no specific noise assumption is assumed. Here, we introduce RL from the MLE perspective under Poisson noise assumption to facilitate the ongoing analysis.

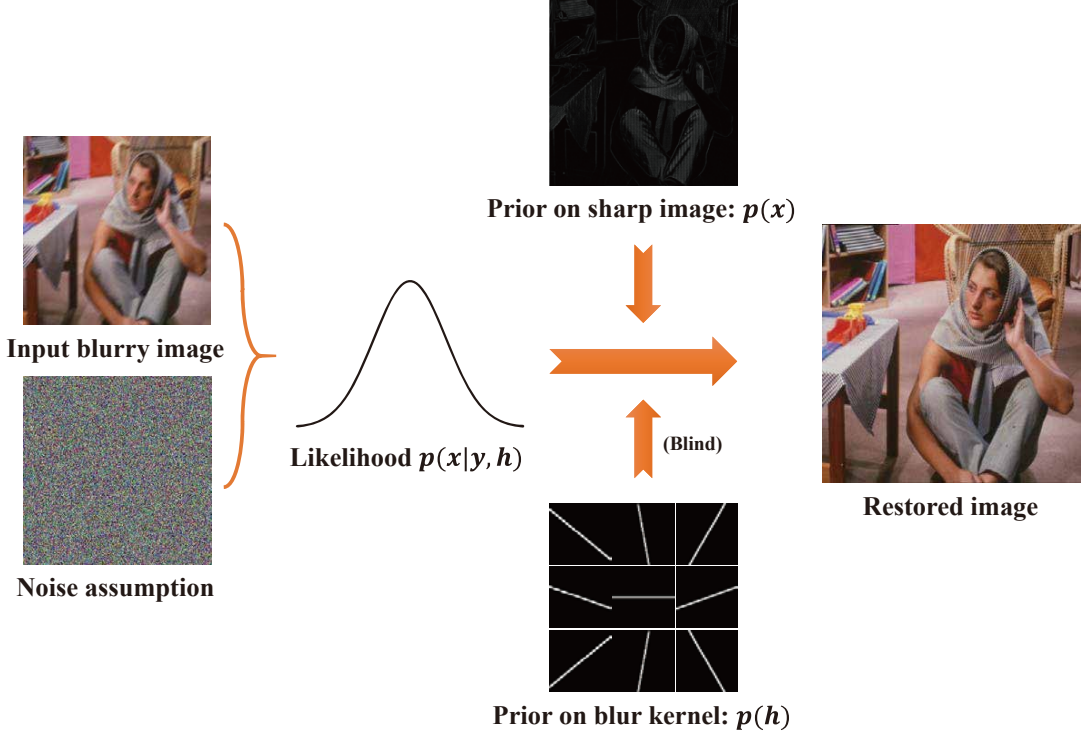


Figure 7: MAP framework for image deblurring

Taking the derivative with respect to  $x$  and setting it to zero, we can get

$$\left(\mathbf{1} - \frac{y}{x * h}\right) * h_- = \mathbf{0}, \quad (3.7)$$

where  $\frac{y}{x * h}$  denotes the point-wise division,  $\mathbf{1}$  is the spatially invariant function which is one everywhere, and  $h_-$  is the symmetrical reflection of  $h$ , meaning  $h_-(i, j) = h(-i, -j)$ . From the unit property of  $h_-$ , i.e.,  $\mathbf{1} * h_- = \mathbf{1}$ , (3.7) can be written as

$$\frac{y}{x * h} * h_- = \mathbf{1}. \quad (3.8)$$

Multiplying both sides by  $x$  and utilizing the *Banach fixed-point* theorem yields

$$x^{t+1} = x^t \odot \left[ \left( \frac{y}{x^t * h} \right) * h_- \right], \quad (3.9)$$

where  $\odot$  is point-wise multiplication. The above RL deconvolution procedure produces a sequence of estimations  $x^t$  and eventually converges to the optimal solution  $x^*$ . If we take the blind case into consideration, (3.9) should be incorporated with an additional step for the estimation of  $h$  in each iteration:

$$x^{t+1} = x^t \odot \left[ \left( \frac{y}{x^t * h^t} \right) * h_-^t \right], \quad (3.10)$$

$$h^{t+1} = \frac{h^t}{\mathbf{1} * x_-^t} \odot \left[ \left( \frac{y}{x^t * h^t} \right) * x_-^t \right]. \quad (3.11)$$

Note that this iterative process is not guaranteed to converge to the global solution since  $\min_{(x, h)} -\log p(y|x, h)$  is not convex, and a small error in  $h$  will lead to significant artifacts in the resultant images. To handle this ill-posedness, the penalized MLE has been introduced:

$$(x^*, h^*) = \underset{x, h}{\operatorname{argmin}} \quad \sum_i ((x * h)_i - y_i \log(x * h)_i) + \lambda_x \Psi_1(x) + \lambda_h \Psi_2(h), \quad (3.12)$$

where  $\Psi_1(\cdot)$  and  $\Psi_2(\cdot)$  are the penalty functions on  $x$  and  $h$ , respectively. Direct derivation leads to the iterations for the penalized version, i.e.,

$$x^{t+1} = x^t \odot \frac{\left(\frac{y}{x^t * h^t}\right) * h_-^t}{\mathbf{1} + \lambda_x \frac{\partial \Psi_1(x^t)}{\partial x^t}}, \quad (3.13)$$

$$h^{t+1} = h^t \odot \frac{\left(\frac{y}{x^t * h^t}\right) * x_-^t}{\mathbf{1} * x_-^t + \lambda_h \frac{\partial \Psi_2(h^t)}{\partial h^t}}. \quad (3.14)$$

Under this formula, [Temerinac-Ott et al \(2012\)](#) utilized the total variation (TV) regularization for  $\Psi_1(x)$  and derived a multi-view RL algorithm. In their method, the underlying sharp 3D image is composed of images observed from multiple viewpoints, each of which is corrupted by a different blur kernel. The multiple blurry images are integrated in a unified formulation from which the sharp image is jointly deduced. [Lefkimmiatis and Unser \(2013\)](#) employed the Schatten norms of Hessian matrix on each image pixel as the regularization on  $x$ . This kind of regularizer is based on the second-order derivatives instead of the first-order derivatives, thus favoring piecewise-smooth solutions, as opposed to TV which produces piecewise-constant solutions.

Regarding  $\Psi_2(h)$ , [Keuper et al \(2013\)](#) discovered a specific property of the wide-field fluorescence microscopy (WFFM) system, which suggests that the *optical transfer function* (OTF), i.e. the Fourier transform of the kernel  $h$ , is well localized and smooth. They proposed imposing the TV constraint on both  $x$  to preserve edges, and OTF to ensure the smoothness of the kernel. [Kenig et al \(2010\)](#) developed a novel approach to restrict  $h$  to a kernel subspace which was produced by either linear or kernel *principal component analysis* (PCA) based on the general forms of the WFFM PSF. This is easily integrated into the iterative RL deconvolution procedure. Additionally, the residual denoising operation is employed to avoid over-smoothing the useful high-frequency details, which is also used in ([Keuper et al 2013](#)).

Most of the above methods are based on the assumption of a spatially invariant kernel. [Tai et al \(2011\)](#) developed the projective motion RL algorithm to tackle the spatially variant case. Under the proposed projective motion blur model, the operations of convolution and correlation in the conventional RL algorithm are replaced by a sequence of forward projective motions and their inverses via *homographies*, which will be discussed in Section 6. Various regularizers are applicable in their framework.

### 3.1.2 Priors for MAP

The penalized MLE in equation (3.12) is exactly the MAP since the penalty functions act as the priors in that  $p(x) = \exp(-\lambda_x \Psi_1(x))$  and  $p(h) = \exp(-\lambda_h \Psi_2(h))$ . Various MAP based methods focus on the development of priors to obtain attractive deblurring results.

It is commonly agreed that for a natural image, the gradients of the sharp image tend to obey a heavy-tailed distribution, meaning that the distribution of gradients has most of its mass on small values but assigns significantly more probability to large values than a Gaussian distribution ([Field 1994](#); [Fergus et al 2006](#)). Thus natural images often contain large regions of constant intensity or gentle intensity whose gradients are interrupted by occasional large changes at the edges or occlusion boundaries. To express such a prior, various approximations of the heavy-tailed distribution are used to describe the gradients. One classic method is to inform a Laplace distribution on the magnitude of the gradients:

$$p^{Lap}(\nabla x) = \prod_i \frac{1}{2b} \exp\left(-\frac{\|\nabla x_i\|_1}{b}\right), \quad (3.15)$$

where  $\nabla$  is the gradient operator,  $\|\cdot\|_1$  is the  $\ell_1$  norm, and  $b$  denotes the scale parameter. Additionally, for computation efficiency, a generalized Gaussian model is used by [Levin et al \(2007\)](#) in their design of an optimal aperture filter, and a Gaussian prior is imposed on the gradient patches instead of gradient pixels by [Hu et al \(2012\)](#). The autocorrelation function of the blur kernel is shown in relation to the covariance matrix of the Gaussian distribution. Through the Fourier transform of the autocorrelation matrix and an additional phase retrieval stage, the blur kernel can be easily recovered. Unlike the MAP methods involving repeated reconstructions of the sharp image, this approach directly relies on basic statistics of the blurry image and is therefore efficient.

Due to the imperfect fitness of the above priors to the real gradient distribution of sharp natural images, these methods tend to remove mid-frequency textures, even though structures such as edges can be preserved. To improve the fitness to the heavy-tailed distribution, [Fergus et al \(2006\)](#) proposed using a *Gaussian mixture model* (GMM) having finite mixture numbers. [Chakrabarti et al \(2010\)](#) extended this to Gaussian scale mixture (GSM) which is a mixture of infinite Gaussian models with a continuous range of variances. The form of the corresponding zero-mean GMM and GSM are respectively

$$p^{GMM}(\nabla x) = \prod_i \sum_{c=1}^C \mathcal{N}(\nabla x_i | 0, \xi_c), \quad (3.16)$$

$$p^{GSM}(\nabla x) = \prod_i \int_{\xi} \mathcal{N}(\nabla x_i | 0, \xi) p(\xi) d\xi, \quad (3.17)$$

where  $\xi$  is the standard derivation of the Gaussian distribution,  $c$  and  $C$  in GMM is the index and the total number of mixtures, and  $p(\xi)$  in GSM is a probability distribution on  $\xi$ . In terms of GSM, a critical issue is the infinite selection of  $\xi$ , which makes it computationally expensive. According to the theoretical analysis by [Palmer et al \(2005\)](#), equation (3.17) can be derived from the variational perspective as

$$p(\nabla x) = \prod_i \sup_{\xi_i > 0} \mathcal{N}(\nabla x_i | 0, \xi_i) p(\xi_i), \quad (3.18)$$

where  $p(\xi)$  takes the form of  $\exp(f(\frac{\xi}{2}))$ . This theoretical convenience is directly utilized by ([Chakrabarti et al 2010](#)) and further investigated by [Zhang et al \(2013a\)](#); [Zhang and Wipf \(2013\)](#). In ([Zhang et al 2013a](#)), authors proposed an adaptive sparse prior based on (3.18). They coupled multiple blurry observations in a joint deblurring procedure by assuming that the parameters  $\{\xi_i\}$  in (3.18) are shared across all images. In their subsequent work ([Zhang and Wipf 2013](#)), the idea of the shared parameter is extended to the single image deblurring problem by additionally assuming the homography expression of blur kernels. In both works, the sparse regularizer is adaptively adjusted according to the noise level estimated in each iteration.

Another question raises: even though the assumed prior coincides perfectly with the gradient distribution of the sharp image, is the recovered gradient distribution able to fit our prior as we expected? A recent research by [Cho et al \(2012c\)](#) responded in the negative. According to their analysis, the above heavy-tailed priors are generally independently forced on each pixel or each local patch, failing to capture the global statistics of gradients. Thus the reconstructed image favors flat regions, resulting in the gradient distribution of the MAP estimates failing to match that of the original sharp image. To address this issue, the authors proposed penalizing the *Kullback-Leibler* (KL) divergence between the gradient distribution of the estimated image  $p_e(\nabla x)$  and a reference distribution  $p_r(\nabla x)$ ,

$$KL(p_e || p_r) = \int_{\nabla x} p_e(\nabla x) \ln \left( \frac{p_e(\nabla x)}{p_r(\nabla x)} \right) d\nabla x, \quad (3.19)$$

which induces a constraint and can be combined with (3.4). The reference distribution is characterized as the generalized Gaussian model and immediately estimated from the blurry image by using the approach of ([Cho et al 2010a](#)). Taking the same idea, [Zhuo et al \(2010\)](#) presented a more direct method which forces the gradients of the deblurred image close to those of the reference image. It is direct because: 1) a *Lorentzian* error norm is imposed on the difference of the two gradients rather than the difference of their distributions, i.e.

$$\rho(\nabla x_e, \nabla x_r) = \log \left( 1 + \frac{1}{2} \left( \frac{\nabla x_e - \nabla x_r}{\varrho} \right)^2 \right), \quad (3.20)$$

where  $\varrho$  is a predefined constant; and 2) the reference image is a flash image of the same scene which is sharp but corrupted by a degraded illumination.

Besides the priors on image gradients, the knowledge on image intensities or transformed domains is extremely helpful in specific applications. For example, [Chen et al \(2011\)](#) developed a content-aware prior for document image deblurring in which the histogram of the whole image is the weighted sum of the histogram of the foreground (dark pixels) and of the background (white pixels). Additionally, a local upper-bound constraint together with TV is imposed to restrain the artifacts in the resultant images. Similarly,

Cho et al (2012a) excavated the specific properties of text images which were then used in the deblurring process. In both methods, the domain-specific knowledge is incorporated into the optimization by employing the variable splitting techniques. Shaked and Michailovich (2011) saddled the generalized Gaussian prior on the representing coefficients in a linear transformed domain and then derived an efficient algorithm to solve the corresponding MAP problem.

As with the PCA technique used in RL deconvolution (Kenig et al 2010), subspace techniques have recently attracted researchers’ attention. The above-mentioned priors express general aspects of human knowledge about natural images. As noted, however, natural images are composed of repetitive local patterns, and some classes of images, such as faces, even lie on a subspace. Discovering this information can help to develop novel priors under the MAP framework. Joshi et al (2010b) proposed a specific restoration problem for personal photo images. Under the MAP framework, they constrained the target face image close to the space expressed by eigen-faces and mean-face generated from the photos of the same person. Benefitting from this space constraint and a sparse gradient constraint, more facial details can be recovered in the resultant photo, while the artifacts are well suppressed. A more general example is for natural images, where the extracted patches can be constrained to be close to a low dimensional manifold (Peyré 2009). Based on this observation, Ni et al (2011) derived a non-blind deblurring method based on a manifold prior learnt from databases. The benefit of these subspace-based priors is the incorporation of more collaboration between the sharp and blurry images, rather than only imposing human heuristic assumptions. This idea is also the key in hardware modification-based methods, as well as learning-based methods, which will be discussed in Section 10.

### 3.1.3 Edge Emphasizing Operation

In the MAP framework, an auxiliary operation is usually employed to produce promising deblurring results; that is, the *edge emphasizing operation*. Typically, the aim of the edge emphasizing operation is to detect and restore the large-scale step edges which generally occur when the blurred edges drift far away from the latent sharp edges, meaning that the blur kernel is large. These operations include the shock filter (Osher and Rudin 1990), the fuzzy operator (Russo and Ramponi 1992), the morphological filtering (Schavemaker et al 2000), the forward-and-backward diffusion process (Gilboa et al 2002), and the recently proposed edge prediction techniques (Joshi et al 2008; Cho et al 2011b). However, these methods often fail when narrow edges or highly textured regions appear in the image, since these patterns exhibit a wide spread of edges in the blurring process. Edge emphasizing operations manipulates over local structures in an image, and thus occasional noise can significantly influence the performance of these operations (Wang et al 2012a, 2013; Faramarzi et al 2013).

To handle this issue, Wang et al (2012a) conducted both theoretical and empirical analyses on the relationship between edge emphasizing operations and MAP estimators. They showed that the advantages of MAP could compensate for the drawbacks of edge emphasizing operations because MAP depends on image statistics which cannot be affected by local structure variations and noises. Nevertheless, the MAP estimator is based on specific blur models and lacks generalization with respect to different models. Fortunately, edge emphasizing operations can address various blur models without any adaptation. Therefore, incorporating the edge emphasizing operation into the iterative MAP estimation can remedy the limitations of both types of method, resulting in improved performance (Cho and Lee 2009). In Wang et al.’s work, the large-scale step edges are recovered by pre-smoothing and shock filter. The narrow edges are then restored by proposing a strongness-aware prior for the MAP scheme, measuring the strongness of local structures. The deblurring procedure is iterated between the large-scale step edge sharpening and MAP estimation. Faramarzi et al (2013) also proved that the edge-emphasizing smoothing operation is beneficial for an accurate estimation of blur kernels. They employed the method of Xu et al (2011) to simultaneously sharpen the edges and remove large numbers of low-amplitude structures via the  $\ell_0$  gradient minimization. Alternatively, Cho et al (2011b) showed that the Radon projections of a blur kernel can be derived from the edges detected in the blurry image, and the kernel can be recovered by enough Radon projections. As a result, a constraint on the blur kernel expressed by Radon projections is added to the likelihood  $p(y|x, h)$ , forming the so-called RadonMAP. Almeida and Almeida (2010) instead described the edge emphasizing operation as a prior modeled as a response of a set of edge detectors in different directions. They imposed a sparse assumption on this prior, favoring piece-wise constant image estimates. To avoid over-smoothing, the operation of gradually

decreasing the regularization parameter was used to produce a promising result. [Xu and Jia \(2010\)](#) pointed out that not all strong edges are profitable for kernel estimation. They proposed a new metric to measure the usefulness of the image edges. The selected edges according to this metric are then used to induce the kernel formation. What follows is an adaptive kernel refinement procedure which is to ensure the sparseness of the estimated kernel without damaging its large-valued elements.

### 3.1.4 Marginalization Techniques

Even though appropriate priors or suitable edge emphasizing operations have been utilized in the MAP framework and have resulted in improved performance, intrinsic problems remain. A recent outstanding research by [Levin et al \(2009, 2011b\)](#) comprehensively analyzes the failure of the MAP scheme and points out how to make the MAP estimation successfully recover the true blur kernel. The  $\text{MAP}_{x,h}$  scheme for blind deconvolution in (3.4) can be rewritten as

$$(x^*, h^*) = \underset{x,h}{\operatorname{argmin}} \|y - x * h\|^2 + \lambda \left( \sum_i |\nabla_h x_i|^\alpha + |\nabla_v x_i|^\alpha \right) \quad (3.21)$$

under the Gaussian noise assumption and a sparse derivative prior or heavy-tailed prior on image gradients.  $\nabla_h$  and  $\nabla_v$  are horizontal and vertical derivatives of the image. According to Levin et al.'s conclusions, the solution of (3.21) under the sparse prior usually favors a blurry result rather than a sharp result, even though  $y$  is generated up to infinitely large image samples from the perfect prior. This observation is also known as the *no-blur explanation* or *delta effect* of MAP, which means that the solution of (3.21) has a higher probability of being

$$\begin{cases} x^* &= y \\ h^* &= \delta \end{cases} \quad (3.22)$$

than the true solution. From the estimation theory perspective, it is evident that we cannot gather sufficient measurements for the  $\text{MAP}_{x,h}$  problem since the number of unknown variables grows with the image size, even to infinity.

As noted by [Levin et al \(2009, 2011b\)](#), the strong asymmetry between the dimensionality of  $x$  and  $h$  provides a favorable property for handling the blind deconvolution. This means that while the dimensionality of  $x$  increases with the image size, the support of  $h$  remains fixed and is small relative to the image size. From this viewpoint,  $h$  can achieve an increased number of measurements when the image size becomes large. Thus, estimation theory tells us through sufficient measurements on  $h$  that the recovered blur kernel under  $\text{MAP}_h$  can be arbitrarily close to the true kernel. Mathematically, the  $\text{MAP}_h$  is

$$\begin{aligned} h^* &= \underset{h}{\operatorname{argmax}} p(h|y) \\ &= \underset{h}{\operatorname{argmax}} \int p(x, h|y) dx, \end{aligned} \quad (3.23)$$

where  $h^*$  is the true kernel, as stated in Claim 3 of ([Levin et al 2011b](#)). Once the kernel is estimated,  $x$  can then be solved in a non-blind deblurring scheme. In their subsequent work ([Levin et al 2011a](#)), Levin et al. noted that  $\text{MAP}_h$  is generally complex and hard to directly compute because the marginalization in (3.23) involves all possible  $x$  explanations, which is computationally intractable. An approximation method was proposed to derive the  $\text{MAP}_h$ . To estimate the blur kernel, they assumed the *i.i.d.* Gaussian imaging noise and the GMM prior on image derivatives, as well as a uniform distribution on  $h$ . Equation (3.23) can then be written as

$$\begin{aligned} h^* &= \underset{h}{\operatorname{argmax}} p(y|h) \\ &= \underset{h}{\operatorname{argmax}} \int p(x, y|h) dx. \end{aligned} \quad (3.24)$$

The above problem is solved by the *expectation-maximization* (EM) framework that alternates between the estimation of  $p(x|y, h)$  which is still a Gaussian (E-step), and the computation of  $h$  under the minimum



mean square error (M-step). In the E-step, however, calculating the mean and covariance of  $p(x|y, h)$  under a sparse prior is generally hard, so the authors proposed approximating the conditional distribution by using variational inference, which will be discussed in Section 3.3.

Wang et al (2013) have discovered several intrinsic issues between edge emphasizing operations and image statistics through a large number of experiments on ImageNet (Deng et al 2009) composed of 1.2 million images in total. Their research points out that the limited number of large scale step edges within a natural image cannot ensure a robust estimation of the blur kernel. Additionally, due to the diversity of natural images, the sparse derivative priors are not consistent across them and it is almost impossible to find a robust measurement that favors sharp explanations for all of them. Different from their previous work (Wang et al 2012a) which uses  $\text{MAP}_{x,h}$ , they adopted the marginalization scheme in (3.23) and developed an adaptive sparse prior composed of two components to ensure robustness. The first component is the commonly-used sparse derivative prior, while the second encodes the edge emphasizing operation.

### 3.2 Minimum Mean Square Error

The Bayesian framework aims to estimate  $x$  or  $h$  from the posterior  $p(x, h|y)$  and a loss function  $L((x^*, h^*), (x, h))$ . The expected loss is computed over all unknown variables,

$$\tilde{L}((x^*, h^*), (x, h)|y) = \int L((x^*, h^*), (x, h))p(x, h|y)dx dh, \quad (3.25)$$

which is called Bayesian expected loss (Brainard and Freeman 1997). The optimal solution  $(x^*, h^*)$  is then chosen to minimize  $\tilde{L}((x^*, h^*), (x, h))$ . If we take  $L((x^*, h^*), (x, h))$  as a Dirac delta loss function, i.e.  $L((x^*, h^*), (x, h)) = 1 - \delta((x^*, h^*) - (x, h))$ , (3.25) becomes the  $\text{MAP}_{x,h}$  problem. Alternatively, if  $L((x^*, h^*), (x, h))$  is the square error loss, we can obtain the *minimum mean square error* (MMSE) formulation:

$$\begin{aligned} (x^*, h^*) &= \underset{\hat{x}, \hat{h}}{\operatorname{argmin}} \int \|\hat{x} - x\|^2 \|\hat{h} - h\|^2 p(x, h|y) dx dh \\ &= \underset{\hat{x}, \hat{h}}{\operatorname{argmin}} \mathbb{E}\{x, h|y\}. \end{aligned} \quad (3.26)$$

For the non-blind deblurring case, the above equation turns into

$$\begin{aligned} x^* &= \underset{\hat{x}}{\operatorname{argmin}} \int \|\hat{x} - x\|^2 p(x|y, h) dx \\ &= \underset{\hat{x}}{\operatorname{argmin}} \mathbb{E}\{x|y, h\}. \end{aligned} \quad (3.27)$$

The equivalence between MMSE and MAP has been proved by Levin et al (2011b), that is *if  $p(h|y)$  has a unique maxima, then for large images, the  $\text{MAP}_h$  estimator followed by an  $\text{MMSE}_x$  image estimation is equivalent to a simultaneous  $\text{MMSE}_{x,h}$  estimation of both  $x$  and  $h$ .*

In spite of this, the empirical demonstrations on image denoising have shown the advantages of MMSE over MAP approaches (Schmidt et al 2010). MAP solutions usually exhibit piecewise constant regions and result in incorrect statistics of the output image, as previously noted, whereas MMSE can achieve the desired statistics by exploiting the uncertainty of the model. Furthermore, the image restoration performance of the MMSE estimator is highly correlated with the generative quality of the model. This observation is particularly useful since MMSE benefits from a powerful learnt generative model even without any regularization weight. As is well known, the regularization parameter is related to the noise level of the degraded image. By taking the above superiority of MMSE, Schmidt et al (2011) integrated the noise estimation process into the MMSE framework by treating the noise standard deviation as a variable of the posterior, i.e., given  $h$  and  $y$ ,

$$p(x|y, h) = \int p(x, \sigma|y, h) d\sigma. \quad (3.28)$$

Compared with MAP, one problem in MMSE remains. Due to the lack of complete knowledge on the joint distribution ( $p(x, h|y)$  and  $p(x|y, h)$ ) in real applications, it is difficult to take expectation in (3.26) and

(3.27) over all possible explanations. To handle this problem, Schmidt et al (2010, 2011) proposed to use the Gibbs sampling method to alternatively generate the sequence of the variable samples, e.g. in deblurring  $\{(x^1, z^1, \sigma^1), \dots, (x^T, z^T, \sigma^T)\}$  where  $z$  is a latent variable. Another way to address the above issue is to abandon the full optimality requirements and use a particular class of estimators to approximate MMSE, such as linear MMSE:

$$\mathbb{E}\{A|B\} = WB + v, \quad (3.29)$$

where  $A$  and  $B$  are as those in (3.1), and  $W$  and  $v$  are the parameters encoding the deblurring process. Minimizing  $\mathbb{E}\{\|A - WB - v\|^2\}$  with respect to  $W$  and  $v$  can result in the optimal  $W$  and  $v$  expressed as

$$\begin{cases} W &= \mathbb{C}_{AB}\mathbb{C}_B^{-1}, \\ v &= \mathbb{E}\{A\} - W\mathbb{E}\{B\}, \end{cases} \quad (3.30)$$

where  $\mathbb{C}_{AB}$  denotes the cross-covariance matrix between  $A$  and  $B$ ,  $\mathbb{C}_B$  is the auto-covariance matrix of  $B$ . Thus the linear MMSE estimator is given by

$$A^* = \mathbb{C}_{AB}\mathbb{C}_B^{-1}(B - \mathbb{E}\{B\}) + \mathbb{E}\{A\}. \quad (3.31)$$

Clearly, the linear MMSE estimator is dependent on the first- and second-order moments of  $A$  and  $B$ .

In a multi-image deblurring setting, the underlying sharp image is generally interrupted by different degradation operations, yielding multiple observations. A typical case is a pair of blurry/noisy images. These two images are correlated with each other and can be restored simultaneously. To be clear, the noisy image can first be denoised to produce a nearly sharp image which can then be used as a guideline or a constraint in the deblurring process. A recent approach by Michaeli et al (2012) has exploited this strategy by proposing the partially linear MMSE (PLMMSE) estimator. Denoting the denoised image  $C$  as a constraint on the estimation of  $W$  and  $v$  in (3.29), PLMMSE is

$$\mathbb{E}\{A|B, C\} = W(C)B + v(C), \quad (3.32)$$

where  $W(C)$  and  $v(C)$  are functions of  $C$ . However, the above estimator is not applicable since it needs the knowledge of the conditional covariance  $\mathbb{C}_{AB|C}$  which is difficult to acquire (Michaeli et al 2012). Therefore a relaxation of the restriction is conducted, resulting in a separable partially linear MMSE:

$$\mathbb{E}\{A|B, C\} = WB + v(C). \quad (3.33)$$

In this case, the PLMMSE estimator is given by

$$A^* = \mathbb{C}_{A\tilde{B}}\mathbb{C}_{\tilde{B}}^{-1}\tilde{B} + \mathbb{E}\{A|C\}, \quad (3.34)$$

where

$$\tilde{B} = B - \mathbb{E}\{B|C\}. \quad (3.35)$$

For more details of the derivation, please see the Appendix in (Michaeli et al 2012). The superiority of PLMMSE over LMMSE lies in the fact that PLMMSE only requires the knowledge of the second-order statistics of  $A$  and  $B$ , and this estimator can reach the lowest worst-case MSE among all estimators which depend solely on the second-order statistics of  $A$  and  $B$ .

### 3.3 Variational Bayesian methods

*Variational Bayesian methods* approximate the intractable integrals arising in a Bayesian inference framework. This situation occurs when there are unknown variables and latent variables over which we want to marginalize without explicitly computing them, such as the marginalization over  $x$  in the above  $\text{MAP}_h$  problem. This type of method, on one hand, provides an analytical approximation of the posterior probability of the unobserved variables to deduce the statistical properties of these variables (Fergus et al 2006). On the other hand, it derives a lower bound for the marginal likelihood of the observed data, which can be used to model selection (Levin et al 2011a) and high-order statistics analysis on the likelihood of the unobserved variables (Zhang et al 2010).

Variational Bayesian methods have recently been applied to image deblurring. [Levin et al \(2011b\)](#) have made the important comment that the variational Bayes approximation experimentally outperforms all existing methods with a different estimation strategy in motion deblurring tasks.

The most frequently-used type of variational Bayes is known as *mean-field variational Bayes*. Recall that the posterior distribution  $p(A|B)$  is over the set of unobserved variables  $A$  given the observed data  $B$ . Here we try to approximate  $p(A|B)$  by finding a variational distribution  $q(A)$  that is restricted to a family of distributions with a simpler form than  $p(A|B)$ , i.e.  $q(A) \approx p(A|B)$ . The mean-field variational Bayes utilizes the KL divergence to measure the difference between the two distributions,

$$\mathcal{L}(q) := D_{KL}(q||p) = \int q(A) \log \frac{q(A)}{p(A|B)} dA. \quad (3.36)$$

To apply this strategy to the deblurring task, [Miskin and MacKay \(2000\)](#) developed an ensemble learning strategy. In their method, the unobserved set  $A$  is an ensemble of the sharp image  $x$ , the blur kernel  $h$  and the noise variance  $\sigma^2$  if the Gaussian noise is assumed. The addition of  $\sigma^2$  frees the user from tuning this parameter, as well as adaptively balancing the data-fidelity term and the prior term. For simplification, a separable factorization of the  $q$  distribution is employed, which corresponds to the *mean field theory* ([Bishop 2006](#)),

$$q(x, h, \sigma^2) = q(x)q(h)q(\sigma^2). \quad (3.37)$$

This leads to

$$\begin{aligned} \mathcal{L}(q) &= \mathbb{E}_{q(x)} \left\{ \log \frac{q(x)}{p(x|y)} \right\} + \mathbb{E}_{q(h)} \left\{ \log \frac{q(h)}{p(h|y)} \right\} \\ &\quad + \mathbb{E}_{q(\sigma^2)} \left\{ \log \frac{q(\sigma^2)}{p(\sigma^2|y)} \right\}. \end{aligned} \quad (3.38)$$

The variable  $B$  in (3.36) is the observed blurry image  $y$ . Our goal is to minimize  $\mathcal{L}(q)$  with respect to  $q(x, h, \sigma^2)$ . By taking the convenience of the above factorization, we can take an alternating update procedure of coordinate descent to solve (3.38), that is: minimizing one factor (for example  $q(x)$ ) while marginalizing out the other factors ( $q(h)$  and  $q(\sigma^2)$ ). The updates are performed by computing the closed-form optimal parameter updates, and performing line-search along the direction of these updated values. According to ([Bishop 2006](#)), the optimal factors obtained from the corresponding sequential updates are given by the expectation of the joint distribution with respect to all unobserved variables except the one of interest, and thus

$$\begin{aligned} \log q^*(x) &= \mathbb{E}_{q(h), q(\sigma^2)} \{ \log p(x, h, \sigma^2, y) \} + \text{const} \\ &= \mathbb{E}_{q(h)} \{ \log p(y|x, h) \} + \log p(x) + \text{const}, \end{aligned} \quad (3.39)$$

$$\begin{aligned} \log q^*(h) &= \mathbb{E}_{q(x), q(\sigma^2)} \{ \log p(x, h, \sigma^2, y) \} + \text{const} \\ &= \mathbb{E}_{q(x)} \{ \log p(y|x, h) \} + \log p(h) + \text{const}, \end{aligned} \quad (3.40)$$

$$\begin{aligned} \log q^*(\sigma^2) &= \mathbb{E}_{q(x), q(h)} \{ \log p(x, h, \sigma^2, y) \} + \text{const} \\ &= \log p(\sigma^2) + \text{const}, \end{aligned} \quad (3.41)$$

Following the calculation of the above equations, the final estimates of the sharp image  $x$ , the blur kernel  $h$  and the noise variance  $\sigma^2$  are taken as the mean values of the distributions  $q^*(x)$ ,  $q^*(h)$  and  $q^*(\sigma^2)$ , respectively.

Using this framework, [Fergus et al \(2006\)](#) and [Whyte et al \(2010, 2012\)](#) operated the variational formulation on the gradient domain, i.e.,  $x$  in the above equations is replaced by the image gradients  $\nabla x$ , to facilitate the statistical assumption on model priors, such as sparse gradients. [Babacan et al \(2012\)](#) proposed a super-Gaussian prior over the image intensities and treated the parameter of this prior as a latent variable in the variational inference. The optimal solutions are similar to equations (3.39-3.41).

Levin et al (2011a) employed the variational inference to handle the marginalization problem in  $\text{MAP}_h$ . By re-arranging the equation (3.36), a free energy is defined as

$$\begin{aligned}\mathcal{F}(q) &= \int q(x, z) \log q(x, z) dz dx \\ &\quad - \int q(x, z) \log p(y, x, z|h) dz dx \\ &= D_{KL}(q(x, z)||p(x, z|y, h)) - \log p(y|h),\end{aligned}\tag{3.42}$$

where  $z$  is an auxiliary latent variable specifying the prior distribution. Due to the non-negativeness of the KL-divergence, minimizing  $\mathcal{F}(q)$  is equivalent to maximizing the lower bound of the likelihood  $p(y|h)$  which is the goal of  $\text{MAP}_h$ . Following an iterative optimization procedure, the optimal distribution can be solved. Levin et al.'s method differs from Fergus et al (2006) and Whyte et al (2010, 2012) in that the target distribution to be approximated is  $p(x|y, h)$  rather than  $p(x, h|y)$ . However, Fergus et al.'s approach selects  $h^*$  from the estimated  $q(x, h)$  distribution by marginalizing out all possible  $x$ 's, and thus belongs to the  $\text{MAP}_h$  approach.

Rather than applying variational inference for deblurring tasks, Zhang et al (2010) explored a different application that compares the reliability of two restoration tasks in the camera shake situation: one is to estimate from a blurry image (deblurring) and the other addresses a sequence of noisy images (denoising). The posterior probabilities of these two tasks are  $p_b = p(x|y_b, \sigma_b^2)$  and  $p_n = p(x|\{y_n^k\}_{k=1}^N, \sigma_n^2)$ , respectively, where  $y_b$  is the blurry image,  $y_n^k$  is the  $k$ -th noisy image,  $\sigma_b^2$  and  $\sigma_n^2$  are the noise variance in two types of image. Since the Hessian matrices of  $\log p_b$  and  $\log p_n$  with respect to  $x$  provide an uncertainty measurement associated with the estimation of  $x$ , the authors applied the variational distributions over the hidden variables (here, the motion paths) to approximate the posterior, i.e.  $\mathcal{L}_b(q) = D_{KL}(q||p_b)$  and  $\mathcal{L}_n(q) = D_{KL}(q||p_n)$ . The comparison result between the Hessian matrix of  $\mathcal{L}_b(q)$  and that of  $\mathcal{L}_n(q)$  reveals that restoration by denoising multiple images is generally more reliable than deblurring a single image.

## 4 Variational Methods

*Variational methods* stem from the calculus of variations and are typically used as approximation methods in a wide variety of settings, such as quantum mechanics, classical mechanics, finite element analysis, and statistics. Such approximations are engaged to convert an ill-posed problem into a well-posed problem which is characterized by exploring additional constraints to reduce the size of the solution space of the unknown variables (Jordan et al 1999). To approximate the problem, a typical setting involves the extremum (maximum or minimum) of a functional composing a function and the associated constraints:

$$\min_A \Phi(A; B) + \lambda \Psi(A),\tag{4.1}$$

where  $A$  is the undetermined variables and  $B$  is the observations. In variational principle,  $\Phi(A; B)$  is called the data-fidelity function,  $\Psi(A)$  is the regularization function, and  $\lambda$  denotes the regularization parameter. Under this formulation, the non-blind image deblurring problem can be written as

$$\min_x \Phi(x; y, h) + \lambda_x \Psi_x(x),\tag{4.2}$$

while the blind case is

$$\min_{x, h} \Phi(x, h; y) + \lambda_x \Psi_x(x) + \lambda_h \Psi_h(h).\tag{4.3}$$

The term  $\Phi$  is determined according to the noise assumptions listed in Table 1. In this section, we generally assume the Gaussian noise model, and the corresponding  $\Phi$  is given by

$$\Phi = \|y - x * h\|_2^2.\tag{4.4}$$

The discussion of the variational methods is organized from three aspects: regularization, optimization, and analysis/synthesis reformulation.

## 4.1 Regularization Techniques

Similar to the character of the prior in the Bayesian inference framework, the regularizers in a variational framework express human knowledge on the interested blurry images. Such knowledge can constrain the solution space such that the deblurred images are favored by human sense.

### 4.1.1 Regularization in Single-image Deblurring

Classically, to stabilize the deblurring result, it is expected the solution to have a small norm, and thus the Tikhonov-Miller regularizer (Tikhonov 1963) is imposed on the sharp image:

$$\Psi_x(x) = \|x\|^2. \quad (4.5)$$

However, this choice is rarely used in modern deblurring tasks because of the property whereby the resultant images will have over-smoothed edges. With this regard, the development of the first-order regularizers which maximally preserve the significant details is more frequently adopted. A typical example is the *total variation* (TV) proposed by Rudin et al (1992)

$$TV_i(x) = \|\sqrt{|\nabla_h x|^2 + |\nabla_v x|^2}\|_1, \quad (4.6)$$

where the subscript  $i$  means it is the *isotropic* version. Complementarily, the *anisotropic* TV is

$$TV_a(x) = \| |\nabla_h x| + |\nabla_v x| \|_1. \quad (4.7)$$

Both  $TV_i$  and  $TV_a$  enhance the visualization of edges in the resultant images. They mainly differ from each other in their sensitivity to edge directions. From the formulations, we can see that  $TV_i$  enforces the same strength on the edges with different directions, whereas  $TV_a$  favors certain directions. Both methods have proven to be useful in numerous applications, such as image denoising, decomposition, super-resolution, inpainting, and non-blind deblurring. Nevertheless, when applied to blind deblurring problems, some failures occur.

Note that TV is intrinsically an  $\ell_1$  norm of the image gradients, and thus induces sparsity over image gradients. According to the delta-effect of MAP mentioned in Section 3, simultaneously estimating  $x$  and  $h$  will result in a blurry image. Another perspective from the  $\ell_1$  properties can assist the understanding of TV failure. For a sharp image of natural scenes, the gradient magnitude is typically sparse, meaning that most values are either zero or very small, but may occasionally be large. If a blur kernel is operated on this image, the high-frequency bands will be attenuated, leading to the magnitudes being un-sparse. To recover the original sparsity, a natural choice is the  $\ell_0$  measure, an important property of which is the scale-invariance, i.e.,  $\min \ell_0(\nabla x) = \min \ell_0(a \cdot \nabla x)$  for any positive values of  $a$ . Minimizing  $\ell_0$  will only lead to a sparse effect, without destroying the magnitudes of large values, thus preserving the energy of original gradients. However,  $\ell_0$  is difficult to optimize because of the lack of derivative information everywhere, and then  $\ell_1$  is utilized as an alternative to approximate  $\ell_0$ . Unfortunately, the blurring process in itself reduces the  $\ell_1$  norm of the gradients. Minimizing  $\ell_1$  fails to preserve or recover the energy of the original gradients. Additionally, the scale variant property makes  $\ell_1$  sensitive to the setting of the regularization parameter  $\lambda$ . Therefore, various methods of approximating the  $\ell_0$  norm while maintaining the scale-invariance property are proposed.

Krishnan et al (2011) recently extended the  $\ell_1$  norm to a normalized version:

$$\Psi(\nabla x) = \frac{\|\nabla x\|_1}{\|\nabla x\|_2}. \quad (4.8)$$

To understand this regularizer, let us focus on the denominator,  $\ell_2$  norm. The blurring process reduces the  $\ell_2$  norm of the gradients as well. Fortunately,  $\ell_2$  is reduced more than the numerator  $\ell_1$  norm, leading to an increased ratio of the two terms. Fig. 8c illustrates that the minimum of this ratio is along the axes. The blurry effect will drive the ratio away from the axes. Therefore, minimizing this regularizer will deduce the blurry effect in the image without destroying the magnitude of the true gradient because  $\ell_1/\ell_2$  is evidently scale invariant, just as we expected.

Another example of the approximation is the unnatural  $\ell_0$  regularizer which is proposed by Xu et al (2013). The unnaturalness stems from the observation that in most iterative deblurring methods, the intermediate

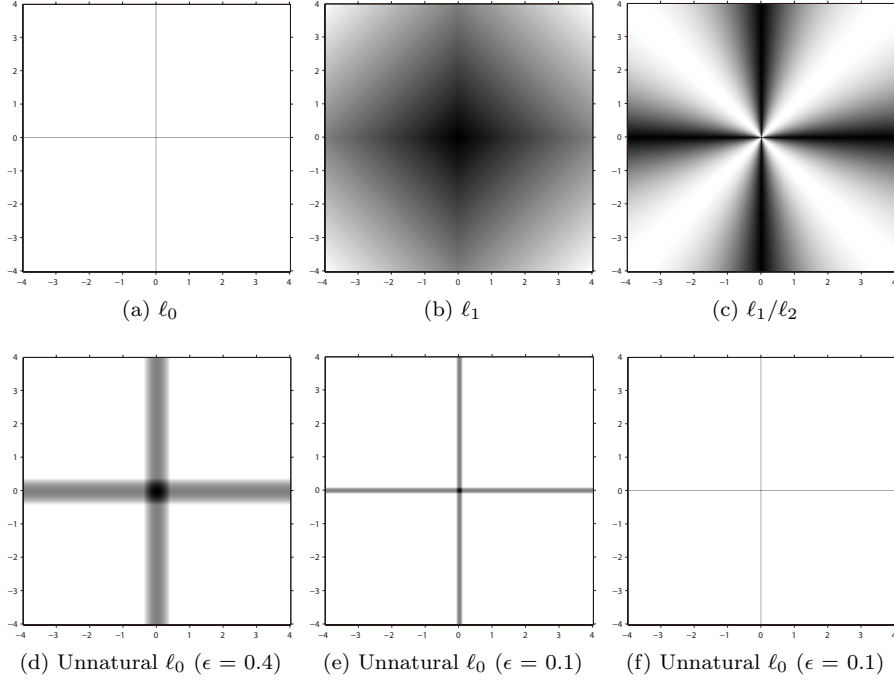


Figure 8: Visualization of different measures

image results only contain high-contrast and step-like structures while suppressing others. These images are different from natural scenes, and hence the term ‘unnatural’ is exploited. To incorporate the step-edge properties in an unnatural representation, the authors utilized the unnatural  $\ell_0$  scheme to preserve the salient changes (i.e., the gradients) in the image. The resultant regularizer is formulated as

$$\Psi(\nabla_h x) = \sum_i \psi(\nabla_h x_i), \quad (4.9)$$

where

$$\psi(z) = \begin{cases} \frac{1}{\epsilon^2} |z|^2, & \text{if } |z| \leq \epsilon, \\ 1, & \text{otherwise.} \end{cases} \quad (4.10)$$

The definition on the vertical derivative  $\nabla_v x_i$  is similar. Depending on the formulation, the gradient magnitudes smaller than  $\epsilon$  are penalized by  $\psi(\cdot)$  while the larger values result in a constant 1 in the objective function. Minimizing this regularizer will remove fine structures and keep useful salient details in the result. Fig. 8d-8f illustrates three plots under different values of  $\epsilon$ . When  $\epsilon$  approaches to zero, this regularizer can be fitted perfectly to the  $\ell_0$  norm. Another property ensuring the unnatural  $\ell_0$  superior to  $\ell_1$  is its scale invariance property, as previously stated. By using this regularization technique in the estimation of blur kernels, the deblurring performance has been notably improved.

While the above regularizers are all based on first-order derivatives, second-order regularization techniques have also proven to be useful in image denoising tasks, and have recently been introduced to deblurring images. Lefkimmiatis et al (2012) extended the first-order TV functional to two second-order cases by defining the mixed norms including  $\ell_1 - \ell_\infty$  and  $\ell_1 - \ell_2$ . These regularizers maintain favorable properties of TV (such as convexity, homogeneity, rotation and translation invariance) well, and can effectively suppress the staircase effect. To solve the resultant variational problem, an efficient algorithm is proposed based on the majorization-minimization approach. Rather than only enforcing the second-order regularization in deblurring tasks, Papafitsoros and Schönlieb (2014) handled the combined problem involving both first- and second-order functionals. The benefit is that the first-order term recovers the step-edges as well as possible, while the second-order term eliminates the artifacts of the staircase produced by the first-order regularizer, without introducing any serious blur in the reconstructed image. Further, the existence and uniqueness of



the solution to the combined problem is proved, and numerical solutions are provided based on the split Bregman iteration (Goldstein and Osher 2009).

A limitation of most regularizers is noted to be based on the local principle, i.e. regularizing the local structures, which can be overcome by the nonlocal ideas. Inspired by the development of nonlocal TV (Gilboa and Osher 2007, 2008) in the image deblurring task (Lou et al 2010), Jung et al (2011) derived a nonlocal *Mumford-Shah* (MS) regularizer by applying the nonlocal operators to the multichannel approximations of the MS regularizer. Due to the repetitiveness of the textures in nature, this regularizer performs better than the local counterpart in various image applications.

#### 4.1.2 Regularization in Multi-image Deblurring

In the case of multi-image deblurring, the regularizers should not only express human knowledge but also exhibit mutual constraints among different images. In this setting, we need to involve a multiple blurring model:

$$y_k = x * h_k + n_k, \text{ for } k = 1, 2, \dots \quad (4.11)$$

Different observation  $y_k$  is obtained by convolving the same latent image  $x$  with a different kernel  $h_k$ , interrupted by specific noise  $n_k$ . Given equation (4.11), the blind deblurring process can be written as a joint variational formulation:

$$\begin{aligned} (x^*, \{h_k^*\}) &= \underset{x, \{h_k\}}{\operatorname{argmin}} \sum_k \|y_k - x * h_k\|_2^2 + \lambda_x \Psi_x(x) \\ &\quad + \lambda_h \sum_k \Psi_h(h_k). \end{aligned} \quad (4.12)$$

Detailed discovery of the relationship among the kernels  $\{h_k\}$  can help us handle the above inverse problem. Specifically in terms of two-image deblurring, Li et al (2011a) presented a theory of *Coprime Blurred Paris* (CBP). CBP means that in the blurry image pair, the  $z$ -transforms of the two kernels are coprime. Mathematically, the equations in (4.11) are transformed into

$$\begin{cases} \tilde{y}_1(z_1, z_2) &= \tilde{x}(z_1, z_2) \cdot \tilde{h}_1(z_1, z_2) + \tilde{n}_1(z_1, z_2) \\ \tilde{y}_2(z_1, z_2) &= \tilde{x}(z_1, z_2) \cdot \tilde{h}_2(z_1, z_2) + \tilde{n}_2(z_1, z_2) \end{cases}, \quad (4.13)$$

where  $\tilde{y}$ ,  $\tilde{x}$ ,  $\tilde{h}$  and  $\tilde{n}$  are the  $z$ -transform of  $y$ ,  $x$ ,  $h$  and  $n$ , respectively. The coprimeness between  $\tilde{h}_1(z_1, z_2)$  and  $\tilde{h}_2(z_1, z_2)$  will theoretically lead to an approximation of the optimal solution  $x^*$ , that is

$$\tilde{x}^* = \operatorname{GCD}\{\tilde{y}_1(z_1, z_2), \tilde{y}_2(z_1, z_2)\}, \quad (4.14)$$

where GCD is the abbreviation of the greatest common divisor (Pillai and Liang 1999), a classic method in *Number Theory*. However, computing (4.14) requires great computation power and system memory, making GCD impractical in deblurring tasks. To address this issue, the authors suggested recovering the sharp image by factoring the Bézout Matrix of  $\tilde{y}_1$  and  $\tilde{y}_2$  to determine the coprimality and can be efficiently solved.

With the assistance of additional hardware, Li et al (2011b) captured two images of the same scene disturbed by camera shake blur. These two images are correlated according to a mapping  $g$ , and can be formulated as

$$\begin{cases} y_1 &= x * h + n_1 \\ y_2 &= g(x) * h + n_2 \end{cases}. \quad (4.15)$$

The mapping  $g$ , as expected, needs to be bijective isometry, and satisfies that  $g^{-1}(a * b) = g^{-1}(a) * g^{-1}(b)$  for all  $a, b \in \mathbb{R}^n$ . Imposing the inverse mapping of  $g$  to the second equation of (4.15), we obtain

$$g^{-1}(y_2) = x * g^{-1}(h) + n_2 = x * h' + n_2, \quad (4.16)$$

where  $h' = g^{-1}(h)$ . Substituting these two observation models into equation (4.12), the two-step optimization is

$$\begin{cases} h^* = \underset{h}{\operatorname{argmin}} \|y_1 - x * h\|_2^2 + \|y_2 - g(x) * h\|_2^2 \\ \quad + \lambda_h \Psi_h(h), \\ x^* = \underset{x}{\operatorname{argmin}} \|y_1 - x * h\|_2^2 + \|g^{-1}(y_2) - x * g^{-1}(h)\|_2^2 \\ \quad + \lambda_x \Psi_x(x). \end{cases} \quad (4.17)$$

In practice, the mapping  $g$  is instantiated as *rot90*, i.e. an image is the 90°-rotated version of the other. This is achieved by employing a beam splitter in the consumer camera. An advantage of this method is the avoidance of the blurry image alignment which, if not properly processed, brings serious artifacts into the restored images.

Even though the alignment problem is crucial in multi-image deblurring, it is possible to handle without an accurate alignment. [Hacohen et al \(2013\)](#) addressed the deblurring task by discovering the dense correspondence ([HaCohen et al 2011](#)) between a blurry image and a sharp reference image. The reference image acts as a regularizer such that the blurry information can be restored from the corresponding sharp information. An assumption in this method is that the two images are required to share the same content but undergo non-rigid geometric transformations; and there is no need to construct an alignment between them.

## 4.2 Optimization Methods

In a variational framework, as well as in other related schemes, a good optimization algorithm can achieve a fast convergence rate and produce an accurate solution. For the problem of deblurring, the general formulation is

$$\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 + \lambda \Psi(\mathbf{x}), \quad (4.18)$$

where we utilize the matrix-vector expression in (2.4). Even though this is for the non-blind deblurring problem, we can see from the discussion in previous sections that the blind case is generally decomposed into a two-step procedure, in which the blur kernel is first estimated by a similar function to (4.18) and the sharp image is then calculated according to (4.18).

A standard algorithm for solving the problem of (4.18) is the so-called *iterative shrinkage/thresholding* (IST) algorithm, which depends on the shrinkage/thresholding function. For example, if we set the  $\Psi$  as the  $\ell_0$  norm on  $x$ , the corresponding shrinkage/thresholding function is the hard-threshold function ([Donoho and Johnstone 1994](#)):

$$\mathcal{T}_{\lambda \ell_0}(\mathbf{w}) = \mathbf{w} \odot \mathbf{1}_{\mathbf{w} \geq \sqrt{2\lambda}}, \quad (4.19)$$

where  $\mathbf{w}$  is the observation to be approximated,  $\mathbf{1}_{\mathbf{w} \geq \sqrt{2\lambda}}$  is the indicator function determined by the condition of the subscript. If  $\Psi$  is  $\ell_1$  norm, the soft-threshold function ([Donoho and Johnstone 1994](#)) is then utilized:

$$\mathcal{T}_{\lambda \ell_1}(\mathbf{w}) = \operatorname{sign}(\mathbf{w}) \odot \max(|\mathbf{w}| - \lambda, 0). \quad (4.20)$$

For solving (4.18), the IST iteration is given by

$$\mathbf{x}^{t+1} = \mathcal{T}_{\lambda \Psi} \left( \mathbf{x}^t - \frac{1}{\gamma} \mathbf{H}^* (\mathbf{H} \mathbf{x}^t - \mathbf{y}) \right), \quad (4.21)$$

where  $\mathbf{H}^*$  is the adjoint of the matrix  $\mathbf{H}$ , and  $\frac{1}{\gamma}$  is the step size. The convergence rate of IST is determined by the parameter  $\lambda$  and the matrix  $\mathbf{H}$ . Small values of  $\lambda$  and/or the ill-condition of  $\mathbf{H}$  results in slow convergence. To accelerate IST, variants have been proposed including the two-step IST algorithm ([Bioucas-Dias and Figueiredo 2007](#)), fast IST algorithm ([Beck and Teboulle 2009](#)), and sparse reconstruction by separable approximation ([Wright et al 2009b](#)). [Michailovich \(2011\)](#) worked on a more general version of TV which is based on multidirectional gradients rather than only the horizontal and vertical derivatives in  $\operatorname{TV}_i$  and  $\operatorname{TV}_a$ , and then derived an IST-type algorithm to solve the model.

Another popular scheme for solving the problem (4.18) is the *alternating direction method of multipliers* (ADMM), a variant of the *augmented Lagrangian Method* (ALM). Formally, (4.18) can be transformed into the following constrained problem by introducing an auxiliary variable  $\mathbf{z}$ :

$$\begin{aligned} \min_{\mathbf{x}, \mathbf{z}} \quad & f_1(\mathbf{x}) + f_2(\mathbf{z}), \\ \text{s.t.} \quad & \mathbf{x} = \mathbf{z}, \end{aligned} \quad (4.22)$$

where  $f_1(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2$  and  $f_2(\mathbf{z}) = \lambda\Psi(\mathbf{z})$ . This procedure is called *variable splitting*. By incorporating the ALM techniques, (4.22) can be further written as

$$\min_{\mathbf{x}, \mathbf{z}, \boldsymbol{\beta}} \mathcal{L}_\mu(\mathbf{x}, \mathbf{z}, \boldsymbol{\beta}) = f_1(\mathbf{x}) + f_2(\mathbf{z}) + \frac{\mu}{2}\|\mathbf{x} - \mathbf{z}\|_2^2 - \boldsymbol{\beta}^T(\mathbf{x} - \mathbf{z}), \quad (4.23)$$

where  $\mu \geq 0$  is the penalty parameter and  $\boldsymbol{\beta}$  is a vector of Lagrange multipliers. Then ADMM alternatively optimizes  $\mathbf{x}$  and  $\mathbf{z}$  (Boyd et al 2011), associated with an additional step to estimate the Lagrange parameters

$$\mathbf{x}^{t+1} = \underset{\mathbf{x}}{\operatorname{argmin}} \mathcal{L}_\mu(\mathbf{x}, \mathbf{z}^t, \boldsymbol{\beta}^t), \quad (4.24)$$

$$\mathbf{z}^{t+1} = \underset{\mathbf{z}}{\operatorname{argmin}} \mathcal{L}_\mu(\mathbf{x}^{t+1}, \mathbf{z}, \boldsymbol{\beta}^t), \quad (4.25)$$

$$\boldsymbol{\beta}^{t+1} = \boldsymbol{\beta}^t + \mu(\mathbf{x}^{t+1} - \mathbf{z}^{t+1}). \quad (4.26)$$

If a transformed signal (say the gradients) of  $\mathbf{x}$  is regularized, the constraint in (4.22) is replaced by the corresponding equations ( $\mathbf{G}\mathbf{x} = \mathbf{z}$  where  $\mathbf{G}$  is the matrix expression of the derivatives). Under this framework, Afonso et al (2010) recently proposed a so-called SALSA (*split augmented Lagrangian shrinkage algorithm*) to tackle the problem (4.18) with a non-smooth regularizer (e.g. TV or  $\ell_1$  norm on wavelet coefficients). Caravan and Blanc-Féraud (2012) used the same scheme to deblur Poisson noisy images. Yang et al (2013) presented the multidimensional anisotropic TV problem which is then solved by ADMM. To accelerate the algorithm, it is recommended to decompose the original multidimensional problem into a set of independent 1D TV problems, which can be processed in parallel, thus reducing the time cost. Ng et al (2013) unified the image deblurring, decomposition and inpainting problem into one formulation, and recommended an algorithm based on ADMM to solve it.

Newton's method and its variants are also used to solve (4.18). Barbero and Sra (2011) developed a Newton-type method under the dual formulation of the TV-proximity problem to solve the  $\ell_2$ -TV problem. Similarly, in Chan et al (2010)'s work, Fenchel-duality and semi-smooth Newton techniques are utilized to handle a  $\ell_1$ -TV problem.

A crucial issue in solving the variational problem is the determination of the regularization parameter. A good selection of the parameter will result in a promising deblurring result, whereas a bad choice may lead to slow convergence as well as the existence of severe artifacts in the results. Generally, when the degradation in the blurry image is significant, the value of  $\lambda$  needs to be set large, to reduce the blur as much as possible. However, in the continuing iterations, the blurry effect is decreased gradually. In this case, small value of  $\lambda$  is required since a large value will damage the fine detail in the image. By considering these effects, a direct implementation is to set  $\lambda$  from large to small according to an empirical reduction rule (Tai et al 2011; Almeida and Almeida 2010; Faramarzi et al 2013):

$$\lambda^{t+1} = \max(\lambda^t \cdot r, \lambda_{min}), \quad (4.27)$$

which depends on the initial value  $\lambda_0$ , the minimal value  $\lambda_{min}$  and the reduction factor  $r \in (0, 1)$ . Usually,  $r = 0.5$ . This setting ensures the improvement of the convergence speed of the algorithm if, at each step in the outer iteration, the optimal solution of its immediate predecessor is used as a starting point for the inner iterative steps (*warm start*). The adaptive adjustment of the parameter by considering the intermediate images in each iteration is more favorable. For example, Montefusco and Lazzaro (2012) proposed an update rule based on the values of the loss function in two previous iterations. If we define the loss function in (4.18) as  $\mathcal{L}(\mathbf{x}, \lambda)$ , we then have

$$\lambda^{t+1} = \lambda^t \cdot \frac{\mathcal{L}(\mathbf{x}^t, \lambda^t)}{\mathcal{L}(\mathbf{x}^{t-1}, \lambda^{t-1})}. \quad (4.28)$$

If the penalty in (4.18) is  $\ell_1$ -based functional, the utilization of this adaptive rule can guarantee the convergence of the forward-backward splitting algorithm. Almeida and Figueiredo (2013b) suggested using the residual whiteness measure as a guideline for the parameter selection and the stopping criterion. The rationale is that if the restored image is well-estimated, the residual image is spectrally white. Their strategy is experimentally demonstrated in both blind and non-blind deblurring tasks.

### 4.3 Analysis & Synthesis

There are two different kinds of formulation for modelling the variational image restoration problem, which are called *analysis* and *synthesis* respectively (Elad 2010; Elad et al 2007). In the analysis formulation, the image  $\mathbf{x}$  is associated with an analysis operator  $\mathbf{U}$  by which  $\mathbf{x}$  is effectively decomposed in a transformed domain, forming the analysis representation  $\mathbf{U}\mathbf{x}$ . Constraining this representation and merging it with the Gaussian noise assumption, the problem is formalized as

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda \|\mathbf{U}\mathbf{x}\|_p, \quad (4.29)$$

where  $\|\cdot\|_p$  is the  $\ell_p$  norm. The TV problem is an instance of the above formulation by setting  $\mathbf{U}$  as the gradient operator and  $p = 1$ . For different forms of  $\mathbf{U}$  and  $p$ , we can find various examples of the analysis representation in Section 4.1, and iterative optimization methods in Section 4.2.

The synthesis formulation indicates that the vectorized image  $\mathbf{x} \in \mathbb{R}^N$  is to be represented as a linear combination of atoms taken from the columns of a full-rank matrix  $\mathbf{D} \in \mathbb{R}^{N \times M}$ , where  $N \leq M$ . The synthesis representation is then taken as  $\mathbf{x} = \mathbf{D}\boldsymbol{\alpha}$ , where  $\boldsymbol{\alpha} \in \mathbb{R}^M$  is expected to be sparse. Under the same noise assumption as in the analysis case, the synthesis formulation is given by

$$\boldsymbol{\alpha}^* = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \|\mathbf{y} - \mathbf{H}\mathbf{D}\boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_p, \quad (4.30)$$

and the optimal solution  $\mathbf{x}^* = \mathbf{D}\boldsymbol{\alpha}^*$ . Synthesis-based methods stem from the basis pursuit method by Chen et al (1998) and have been well developed in the past years. A particularly popular synthesis representation is sparse representation, which will be detailed in the next section. Curvelet (Candes and Donoho 2000), contourlet (Do and Vetterli 2005), ridgelet (Candès and Donoho 1999) and their variants are all excellent works under this framework and most of them are focused on the denoising problem. In terms of deblurring, Tzeng et al (2010) recently exploited the collaborative property of multiband deblurring in the fluid lens camera system. Since the green plane of the image remains sharp in the image formation process, this sharp information can be used to recover the contourlet coefficients of other blurry planes. Zhang and Hirakawa (2013) proposed the double discrete wavelet transform (DDWT) whose coefficients possess near-blur-invariant properties, thus enabling the identity of the blur kernel in the DDWT domain. Under the tight wavelet frame system, Ji and Wang (2012a) derived a deblurring scheme by considering the kernel error in a unified formulation, whereas Cai et al (2012) used framelet decomposition to derive an analysis-based sparse prior rather than the synthesis-based prior.

Analysis and synthesis modeling share a similar structure in their formulation, but the resultant  $\mathbf{x}$  from them are not necessarily equal, and may even be significantly different from each other in most cases. According to (Elad et al 2007), the equivalence between the two solutions occurs in anyone of the following three situations: 1) if  $\mathbf{U}$  takes as a square and non-singular analyzing operator and  $\mathbf{D} = \mathbf{U}^{-1}$ ; 2) if  $\mathbf{H} = \mathbf{I}$ ,  $\mathbf{U} \in \mathbb{R}^{L \times N}$  ( $L \leq N$ ) is a full-rank operator, and  $\mathbf{D} = \mathbf{U}^+$ ; or 3) if  $p = 2$ ,  $\mathbf{U} \in \mathbb{R}^{L \times N}$  ( $L > N$ ) is a full-rank operator, and  $\mathbf{D} = \mathbf{U}^+$ . Even if one of these equivalences exist, the difference between the analysis scheme and the synthesis scheme suggests each has merits and demerits in specific tasks. For example, the redundancy of the operators  $\mathbf{U}$  and  $\mathbf{D}$  has a different effect on the analysis approach and the synthesis approach, respectively. While a redundant  $\mathbf{D}$  enables the synthesis scheme to describe more complex signals, redundant  $\mathbf{U}$  in the analysis case makes the signal inconsistent. On the other hand, in the synthesis scheme, considering the dependence among the atoms in  $\mathbf{D}$ , the incorrect selection of one atom can propagate to the others; in contrast, for the analysis scheme, considering each pixel is related to its neighbor pixels, this dependence is significantly reduced, which ensures the imprecise estimation of one pixel will not seriously affect the estimation of the other pixels (Elad et al 2007).

The above analysis is based on a single analysis formulation or a single synthesis formulation. Nevertheless, a recent state-of-the-art work (Danielyan et al 2012, 2010) suggested a combined scheme to address

the deblurring problem, i.e. *block-matching 3-D (BM3D) image modeling* (originally proposed for the image denoising task (Dabov et al 2007)). Danielyan et al (2012, 2010) deduced the BM3D modeling from both the analysis formulation and the synthesis formulation, proposing the so-called BM3D frames. Within the formulations of (4.29) and (4.30), the authors extended BM3D modeling to modern variational image restoration techniques. Furthermore, a combination of analysis/synthesis is formulated as a *generalized Nash equilibrium* (GNE) problem including two objective functions, which are

$$\begin{cases} \mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2, & s.t. \|\mathbf{x} - \mathbf{D}\boldsymbol{\alpha}^*\|_2^2 \leq \varepsilon_1, \\ \boldsymbol{\alpha}^* = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \lambda \|\boldsymbol{\alpha}\|_p, & s.t. \|\boldsymbol{\alpha} - \mathbf{U}\mathbf{x}^*\|_2^2 \leq \varepsilon_2, \end{cases} \quad (4.31)$$

where  $\varepsilon_1, \varepsilon_2 > 0$  and  $p = 0, 1$ . The optimization algorithm iteratively solves the two functions. As noted by the authors, simultaneously minimizing the two functions cannot be achieved since minimization of either one of them will lead to an increase of the other. This effect is called the *Nash equilibrium* and provides a balance between the fit of the restoration  $\mathbf{x}$  to observation  $\mathbf{y}$  and the complexity of the model  $\|\boldsymbol{\alpha}\|_p$ . This balance plays a vital role in the image restoration.

Xie and Rahardja (2012) and Shen et al (2011) worked on the balanced formulation:

$$\min_{\boldsymbol{\alpha}} \|\mathbf{y} - \mathbf{H}\mathbf{D}\boldsymbol{\alpha}\|_2^2 + \gamma \|(\mathbf{I} - \mathbf{D}^T\mathbf{D})\boldsymbol{\alpha}\|_2^2 + \boldsymbol{\lambda}^T |\boldsymbol{\alpha}|, \quad (4.32)$$

where  $\mathbf{D}$  and  $\boldsymbol{\alpha}$  are same as those in (4.30),  $\mathbf{I}$  is the identity matrix,  $\gamma > 0$ ,  $\boldsymbol{\lambda}$  is a given nonnegative weight vector, and  $|\cdot|$  denotes the element-wise absolute operator. This formulation can be viewed as a combination of the analysis formulation and the synthesis formulation. Specifically, when  $\gamma = 0$ , the problem (4.32) is turned into the synthesis scheme, and while  $\gamma = \infty$ , (4.32) is then reduced as the analysis problem. Therefore, by taking  $0 < \gamma < \infty$ , this formulation balances the sparsity of the coefficients  $\boldsymbol{\alpha}$  and the smoothness of the reconstructed image. To solve (4.32), Xie and Rahardja (2012) proposed a fast ADMM-type algorithm by exploiting the special structures of  $\mathbf{H}$  and  $\mathbf{D}$ . In Shen et al (2011)'s work, it is noted that the problem in (4.32) is not strictly convex, and thus the optimal solution is not unique. The authors addressed this issue by adding an additional regularizer  $\|\boldsymbol{\alpha}\|_2^2$  to constrain the solution space, and an accelerated proximal gradient algorithm was developed to solve the problem.

## 5 Sparse Representation-based Methods

Sparse representation accounts for a decomposition that represents a signal  $\mathbf{x} \in \mathbb{R}^N$  as a sparse linear combination of basis atoms  $\mathbf{D}_m \in \mathbb{R}^N (m = 1, \dots, M)$ . Given an over-complete dictionary  $\mathbf{D} = [\mathbf{D}_1, \dots, \mathbf{D}_M]$  where  $N \ll M$ , the underlying sharp image  $\mathbf{x}$  in (2.4) is sparsely represented as follows

$$\begin{aligned} \mathbf{x} &= \mathbf{D}\boldsymbol{\alpha}, \\ s.t. \quad & \|\boldsymbol{\alpha}\|_0 \ll N, \\ & \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2 < \epsilon, \end{aligned} \quad (5.1)$$

where  $\boldsymbol{\alpha} \in \mathbb{R}^M$ ,  $\epsilon > 0$  and the second constraint stems from the Gaussian noise assumption. Taking into consideration the noise and the representation error, the above problem can be formulated as

$$\min_{\boldsymbol{\alpha}} \|\mathbf{y} - \mathbf{H}\mathbf{D}\boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_0. \quad (5.2)$$

Since solving this  $\ell_0$  regularized problem is NP-hard and computationally prohibitive (Bruckstein et al 2009), approximation algorithms are usually considered. A general scheme is to replace the  $\ell_0$  with an  $\ell_1$  norm, leading to a reformulation of (5.2) as

$$\min_{\boldsymbol{\alpha}} \|\mathbf{y} - \mathbf{H}\mathbf{D}\boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1, \quad (5.3)$$

which can be solved by *basis-pursuit* (BP) (Chen et al 1998) or LASSO (Tibshirani 1996). More general optimization methods for (5.3) are reviewed in (Zibulevsky and Elad 2010).

## 5.1 Smooth Enhancement

The sparse representation framework has been successfully applied to various image processing tasks, such as denoising (Elad and Aharon 2006), deblurring (Cai et al 2009), inpainting (Elad et al 2005), and super-resolution (Yang et al 2010). In these approaches, the sign  $\mathbf{x}$  in the above equations typically represents a patch in the image. Mathematically, the minimization over the whole image can be written as

$$\min_{\alpha} \sum_i \|\mathbf{R}_i \mathbf{y} - \mathbf{H} \mathbf{D} \alpha_i\|_2^2 + \lambda \sum_i \|\alpha_i\|_1, \quad (5.4)$$

where  $\alpha = [\alpha_1, \dots, \alpha_K]$ ,  $K$  is the total number of patches,  $\mathbf{R}_i$  denotes the extraction of the patch at location  $i$ , and the patch  $\mathbf{x}_i = \mathbf{D} \alpha_i$ . To reconstruct the sharp image, the estimated patches need to be fused. A general operation is to chop the image into overlapping patches, which are then averaged to calculate each pixel's value. However, as noted by Cai et al (2012), a crude fusing scheme in the synthesis approach often leads to visible artifacts along the image edges, exhibiting less smoothness in the reconstructed image.

To suppress the artifacts and enhance the smoothness, additional constraints have been proposed that involve kinds of formations. For example, Ma et al (2013) added a TV regularizer about the estimated image into the sparse formulation, since the TV regularization in the analysis approaches has the effect of enhancing edges and restraining the changes in non-edge regions. Based on the Poisson noise assumption, the objective (5.4) changes to

$$\begin{aligned} \min_{\alpha, \mathbf{x}, \mathbf{D}} \quad & \gamma < \mathbf{H} \mathbf{x} - \mathbf{y} \log \mathbf{H} \mathbf{x}, \mathbf{1} > + \sum_i \|\mathbf{R}_i \mathbf{x} - \mathbf{D} \alpha_i\|_2^2 \\ & + \sum_i \mu_i \|\alpha_i\|_0 + \eta \|\nabla \mathbf{x}\|_1, \end{aligned} \quad (5.5)$$

where  $\gamma$  and  $\eta$  balance the corresponding terms, and  $< \cdot, \cdot >$  denotes the inner product. An algorithm based on the variable splitting method is also proposed to iteratively solve different unknown variables.

Inspired by the superiority of nonlocal strategy over the local estimation strategy, Dong et al (2011b) introduced a combination of two adaptive regularizers, one of which involves a set of *auto-regression* (AR) models  $\{\mathbf{a}_1, \dots, \mathbf{a}_L\}$  characterizing the local structures, while the other encodes the nonlocal similarity. The resultant formulation is

$$\begin{aligned} \min_{\alpha} \quad & \sum_i \|\mathbf{R}_i \mathbf{y} - \mathbf{H} \mathbf{D} \alpha_i\|_2^2 + \sum_i \lambda_i \|\alpha_i\|_1 \\ & + \underbrace{\gamma \sum_i \|x_i - \mathbf{a}_l^T \mathbf{x}_{\setminus i}\|_2^2}_{\text{AR regularization}} + \underbrace{\eta \sum_i \|x_i - \mathbf{b}_i^T \beta_i\|_2^2}_{\text{Nonlocal regularization}}, \end{aligned} \quad (5.6)$$

where  $\{\lambda_i\}$ ,  $\gamma$ ,  $\eta$  are regularization parameters. For the  $i$ -th patch,  $x_i$  and  $\mathbf{x}_{\setminus i}$  are the central pixel and the non-central pixels respectively,  $\mathbf{a}_l$  denotes the parameters of the selected  $l$ -th AR model,  $\beta_i$  collects the central pixels of similar patches searched across the whole image, and  $\mathbf{b}_i$  lists the corresponding nonnegative weights. The elements of  $\mathbf{b}_i$  are required to sum to 1, i.e.  $\sum_j \mathbf{b}_{i,j} = 1$ . This scheme considers both local smoothness and nonlocal structure similarity, which in cooperation with sparse representation, will provide an accurate estimation of the pixel by excavating as much of the information in the image as possible. In Dong et al (2011a, 2013)'s subsequent work, the same idea of nonlocal regularization is employed. Rather than directly estimating the central pixel according to the nonlocal similarity, the authors impose this regularization onto the sparse codes  $\alpha$  by replacing the last term in (5.6) with

$$\sum_i \|\alpha_i - \mathbf{A}_i \omega_i\|_p, \quad (5.7)$$

where  $p = 1$  or  $2$ , the columns of  $\mathbf{A}_i$  correspond to the sparse codes of the similar patches which are collected via block matching over all extracted patches, and  $\omega_i$  is the same weight vector as  $\mathbf{b}_i$  in (5.6). The resultant model is named *centralized sparse representation*. The rationale behind this nonlocal regularization comes from an empirical investigation, showing that the sparse codes  $\alpha^e$  calculated according to (5.4) usually have



sparse coding noise with respect to the true codes  $\alpha^t$ , i.e.  $\alpha^e = \alpha^t + \mathbf{n}_\alpha$  where  $\mathbf{n}_\alpha$  is the noise that can be approximately characterized by Laplace distribution. The authors proposed using  $\mathbf{A}_i \omega_i$  to approximate  $\alpha^t$ , and thus minimizing the regularization in (5.7) would restrain the noise emerging in the optimization procedure.

## 5.2 Dictionary Learning

The selection of the dictionary  $\mathbf{D}$  is crucial in the sparse representation-based methods, since a suitable  $\mathbf{D}$  will result in low reconstruction error, otherwise high error will be caused. A classic selection is the predefined regular dictionary, such as the redundant discrete cosine transform (DCT) (Guleryuz 2006a,b) and the overcomplete Haar dictionary. These dictionaries can be pre-calculated before the optimization of problem (5.2) or (5.3), saving a large amount of computational source, but these choices provide limited performance due to the ignoring of the real data.

Alternatively, a task-specific dictionary learning strategy has been proven to be advantageous. A typical example is the K-SVD (Elad and Aharon 2006; Aharon et al 2006), in which the redundant dictionary is learned from a training set of high quality images or the currently processed degraded images. In this method, the optimization alternates between solving the sparse codes  $\alpha$  and the dictionary  $\mathbf{D}$ , thus increasing the computational cost compared with using a predefined regular dictionary. Fortunately, the performance of K-SVD is enhanced, and it is also shown that the scheme in which the dictionary is learned from the currently processed image is superior to other schemes that utilize a training set to learn the dictionary. This highlights the importance of task-specificity.

The K-SVD algorithm is utilized by Ma et al (2013) to update the dictionary in each iteration for deblurring. Besides, Perrone et al (2012) used a dictionary  $\mathbf{D}$  composed of two parts, one of which consists of blurry patches extracted from the current processed image  $y$ , denoted as  $\mathbf{D}_y$ , while another collects the patches from a sharp dataset and corrupts them using the known blur kernel, denoted as  $\mathbf{D}_d$ . Thus we can write  $\mathbf{D} = [\mathbf{D}_y, \mathbf{D}_d]$ , corresponding to the self-similarity-based and dataset-based dictionaries, respectively. Instead of solving the  $\ell_1$  optimization problem, the authors computed the representation of the blurry image on  $\mathbf{D}$  by employing the nonlocal mean method and a subsequent refinement. This representation, denoted as  $\alpha^{nlm}$ , is then regarded in a variational formulation as a constraint on the sharp image. The problem is written as

$$\begin{aligned} \min_{\mathbf{x}, \mathbf{n}, \mathbf{e}} \quad & \frac{1}{2} \|\mathbf{x} - \mathbf{D} \alpha^{nlm}\|_2^2 + \lambda \|\nabla \mathbf{x}\|_2 + \frac{\eta}{2} \|\mathbf{n}\|_2^2 + \gamma \|\mathbf{e}\|_1, \\ \text{s.t.} \quad & \mathbf{y} = \mathbf{H} \mathbf{x} + \mathbf{n} + \mathbf{e}, \end{aligned} \quad (5.8)$$

where  $\lambda, \eta, \gamma$  balance the different terms, and  $\mathbf{e}$  denotes the error introduced by the inaccurate blur kernel. The effectiveness of this formulation is experimentally demonstrated when a noisy blur kernel is assumed.

Generally, the local structures of the natural images exhibit a certain level of correlations and can be synthesized by similar atoms in a redundant dictionary. By intentionally arranging the dictionary atoms, the estimated sparse patterns will be constrained to a specific subset of atoms. This is the so-called structured sparse representation. By exploring structured sparse representation, Dong et al (2011a,b, 2013) introduced a PCA-based dictionary learning method. In the training phase, a large number of patches are extracted from an additional dataset of natural images, followed by a clustering procedure to create a set of clusters (say  $K$  clusters), each of which encodes a certain type of local structure. Then PCA is applied to each cluster (denoted by its centroid  $\mu_k$ ,  $1 \leq k \leq K$ ), in which all the affiliated patches are vectorized. The calculated principal components, which encode the structural information, are utilized as the subdictionary (denoted as  $\mathbf{D}_k$ ) for the corresponding cluster. Thus the final dictionary is a concatenation of all subdictionaries, i.e.  $\mathbf{D} = [\mathbf{D}_1, \dots, \mathbf{D}_K]$ . In the sparse coding phase, a patch  $\mathbf{x}$  is adaptively assigned a subdictionary whose index is

$$k^* = \operatorname{argmin}_k \|\tilde{\mathbf{D}}_k \mathbf{x} - \tilde{\mathbf{D}}_k \mu_k\|_2, \quad (5.9)$$

where  $\tilde{\mathbf{D}}_k$  collects the first several most significant components in  $\mathbf{D}_k$ . Following the selection of the subdictionary, the sparse codes of  $\mathbf{x}$  are solved by optimizing (5.3). Since  $\mathbf{x}$  is degraded by blur and noise, the selection in (5.9) may be not optimal. The authors proposed to iteratively implement (5.9) and (5.3) until the estimation of  $\mathbf{x}$  converged. This PCA-based structured sparsity is shown to be equivalent to *piecewise*

*linear estimation* (PLE) under the GMM assumption (Yu et al 2012). By contrast, a MAP-EM algorithm is proposed in (Yu et al 2012) to solve the PLE problem instead of using PCA, showing that PLE can stabilize and improve traditional nonlinear sparse inverse problems.

### 5.3 Combined Tasks under Sparse Representation

Image deblurring is a low-level task aiming to produce high quality images for the subsequent high-level vision tasks. Several recent sparse representation-based methods combine the deblurring problem and other elements, such as compressed sensing (Amizic et al 2013; Rostami et al 2012) and object recognition (Zhang et al 2011), into a unified scheme. In Amizic et al (2013)’s work, the blind image deconvolution technique is integrated into the compressed sensing (CS)-based imaging system. In this system, a blurry image is represented as

$$\mathbf{y} = \mathbf{M}\tilde{\mathbf{x}} + \mathbf{n} = \mathbf{M}\mathbf{H}\mathbf{x} + \mathbf{n}, \quad (5.10)$$

where  $\mathbf{M}$  is the CS measurement matrix, and  $\tilde{\mathbf{x}} = \mathbf{H}\mathbf{x}$ . To sample the signal  $\tilde{\mathbf{x}}$ , CS involves the following optimization problem:

$$\min_{\boldsymbol{\alpha}} \|\mathbf{y} - \mathbf{M}\mathbf{W}\boldsymbol{\alpha}\|_2^2 + \tau\|\boldsymbol{\alpha}\|_1, \quad (5.11)$$

which is similar to (5.3), and  $\mathbf{W}$  denotes the redundant transformation by which  $\tilde{\mathbf{x}} = \mathbf{W}\boldsymbol{\alpha}$  can be sparsely represented. Combined with the variational framework for blind image deblurring in (4.3), a unified objective function can be obtained

$$\begin{aligned} \min_{\mathbf{x}, \mathbf{H}, \boldsymbol{\alpha}} \quad & \|\mathbf{y} - \mathbf{M}\mathbf{W}\boldsymbol{\alpha}\|_2^2 + \tau\|\boldsymbol{\alpha}\|_1 + \lambda_x\Psi_x(\mathbf{x}) + \lambda_h\Psi_h(\mathbf{H}), \\ \text{s.t.} \quad & \mathbf{H}\mathbf{x} = \mathbf{W}\boldsymbol{\alpha}. \end{aligned} \quad (5.12)$$

As noted, the advantage of (5.12) compared to the sequential approach (a two-step approach in which CS is followed by blind image deblurring) is the capacity of imposing an additional structural constraint on the sparse codes  $\boldsymbol{\alpha}$  because in the sequential case,  $\boldsymbol{\alpha}$  is determined when completing CS and cannot be changed in the subsequent deblurring procedure. Rostami et al (2012) introduced a derivative compressed sensing (DCS) technique to derive the generalized pupil function (GPF) of the short-exposure optical system, where the GPF can uniquely determine the PSF introduced by air turbulence. Once the estimation of PSF via DCS is completed, a non-blind deconvolution procedure under the variational framework is conducted to produce the restored optical images.

The performance of face recognition heavily depends on the quality of the captured face images. When images are degraded by blur or noise, the extracted features may be significantly disturbed. To address this issue, Zhang et al (2011) proposed a framework to simultaneously blind-restore the face images and recognize the face labels. The whole framework stems from the sparse representation-based classification (SRC) (Wright et al 2009a) which assumes that the data samples belonging to the same class lie in the same low-dimensional subspace. Different from SRC, the target image to be labeled in this method is a blurry face image, and the results include an estimated kernel, a restored image and its associated label. Formally, the proposed objective function is

$$\begin{aligned} \min_{\mathbf{x}, \mathbf{H}, \boldsymbol{\alpha}, c} \quad & \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \eta\|\mathbf{x} - \mathbf{D}\boldsymbol{\alpha}\|_2^2 + \tau\|\boldsymbol{\alpha}\|_1 \\ & + \lambda_x\Psi_x(\mathbf{x}) + \lambda_h\Psi_h(\mathbf{H}), \end{aligned} \quad (5.13)$$

where the dictionary  $\mathbf{D}$  includes a set of  $C$  subdictionaries, each of which contains the training face images of one specific class, i.e.  $\mathbf{D} = [\mathbf{D}_1, \dots, \mathbf{D}_C]$ , and  $c$  is the class label embedded in  $\mathbf{D}$  and  $\boldsymbol{\alpha}$ . In the above formulation,  $\mathbf{x}$ ,  $\mathbf{H}$ ,  $\boldsymbol{\alpha}$ ,  $c$  are optimized alternatively. It should be noted that  $\mathbf{D}$  is realized as the collection of sharp training images when optimizing  $\mathbf{x}$  and  $\mathbf{H}$ , whereas in the estimation of  $\boldsymbol{\alpha}$  and  $c$ ,  $\mathbf{D}$  is corrupted by the intermediate estimation of  $\mathbf{H}$  to generate the blurry dictionary  $\mathbf{D}^b$ , i.e.  $\mathbf{D}^b = \mathbf{H}\mathbf{D}$ . Compared with the state-of-the-art blind deblurring algorithms and SRC, this framework obtains a significant performance improvement.

## 6 Homography-based Methods

In this section and the next section, our discussion will mainly focus on the modeling and associated processing of the spatial variant blur effect.

*Homography-based modeling* is generally proposed to simulate the blur effect induced by the camera's egomotion or camera shake. Recall that in the image formation process, a 3D scene point  $(u, v, m)$  is mapped to a 2D image plane point  $(i, j)$ , which can be formulated in the homogeneous coordinates as

$$(i_t, j_t, 1)^T = \mathcal{P}_t(u, v, m, 1)^T, \quad (6.1)$$

where  $t$  denotes the time index. In the case of camera motion,  $\mathcal{P}_t$  may vary with time as a function of camera translation and rotation, causing a fixed point in the scene to be projected onto different locations in the image plane at each time. As is well-known, when using a pinhole camera, all views seen by the camera are projectively equivalent except for the boundaries (Whyte et al 2010; Joshi et al 2010a). This means that for a static scene with constant depth, the 2D images projected at different instances of time are related via a *homography*. Denoting the image point at time  $t = 0$  as  $(i_0, j_0)$ , the homography and projected point at time  $t$  are modeled as

$$\mathcal{H}_t(d) = \mathbf{K} \left( \mathbf{R}_t + \frac{1}{d} \mathbf{T}_t \mathbf{N}^T \right) \mathbf{K}^{-1}, \quad (6.2)$$

$$(i_t, j_t, 1)^T = \mathcal{H}_t(d)(i_0, j_0, 1)^T, \quad (6.3)$$

for a particular depth  $d$ , where  $\mathbf{K}$  is the camera's internal calibration matrix,  $\mathbf{R}_t$  and  $\mathbf{T}_t$  are the rotation matrix and translation vector at time  $t$ , and  $\mathbf{N}$  is the unit vector orthogonal to the image plane. Based on this formulation, the image captured at any time is expressed by the initial image  $x_0$ , i.e.

$$x_t(i, j) = x_0(\mathcal{H}_t(d)(i, j, 1)^T). \quad (6.4)$$

For simplicity, we express the coordinates as a column vector  $\mathbf{i}$  and the homography as  $\mathcal{H}_t$ , and then (6.4) can be rewritten as

$$x_t(\mathbf{i}) = x_0(\mathcal{H}_t \mathbf{i}). \quad (6.5)$$

Thus we can define the blurry image  $y$  as the accumulated result over the exposure duration  $\tau$ , which is given by

$$y(\mathbf{i}) = \int_{t=0}^{\tau} x_0(\mathcal{H}_t \mathbf{i}) dt, \quad (6.6)$$

where we omit the noise term. We rewrite (6.5) in the matrix-vector form as

$$\mathbf{x}_t = \mathbf{H}_t \mathbf{x}_0, \quad (6.7)$$

where  $\mathbf{H}_t$  is a sparse resampling matrix that implements the image warping and resampling due to the homography, and then (6.6) becomes

$$\mathbf{y} = \int_{t=0}^{\tau} \mathbf{H}_t \mathbf{x}_0 dt. \quad (6.8)$$

However, due to the successive duration  $[0, \tau]$  in (6.6) and (6.8), infinite instantiations of  $\mathcal{H}_t$  will be created, causing the ill-posedness of the problem. To handle this issue, two methods can be employed: one which discretizes the time duration and one which decomposes  $\mathbf{H}_t$  into the basis transformations.

### 6.1 Time Discretization

If we suppose that the duration  $\tau$  is segmented to  $N$  equivalent periods, in each of which the homography  $\mathcal{H}$  is approximately consistent, the equation (6.6) becomes

$$y(\mathbf{i}) \approx \frac{1}{N} \sum_{t=1}^N x_0(\mathcal{H}_t \mathbf{i}). \quad (6.9)$$

The left hand side of (6.9) is equal to the right hand side when  $N \rightarrow \infty$ . If  $N$  is set to a limited value, the unknown variables in  $\{\mathcal{H}_t\}$  will be well-constrained, reducing the ill-posedness of the problem. Equation (6.9) is nothing but the *projective motion blur* model proposed by Tai et al (2011). To estimate each homography, the motion is assumed to be uniform in the exposure time. Each  $\mathcal{H}_t$  can therefore be computed directly according to the whole homography generated from  $t = 0$  to  $\tau$ , i.e.  $\mathcal{H}_t = \sqrt[N]{\mathcal{H}_{[0,\tau]}}$ . Based on this formulation, Tai et al (2011) also developed the projective motion Richardson-Lucy algorithm which has been mentioned in Section 3.1.1. Following this model, Tai et al (2010b) proposed a coded exposure technique to introduce discontinuities during the exposure period for the purpose of reducing the discretization error existing in equation (6.9), while at the same time preserving the high-frequency spatial information. The captured image provides more details for the recovery of the blur kernel. In conjunction with the matting technique, the moving object in the image is isolated and deblurred appropriately.

A strong assumption in equation (6.9) is the consistent homography in each equivalent period, which may not fit to the reality. A more general formulation is

$$y(\mathbf{i}) \approx \sum_{t=1}^N w_t x_0(\mathcal{H}_t \mathbf{i}), \quad (6.10)$$

where  $w_t$  denotes the proportion of the period occupied by  $\mathcal{H}_t$ , and  $\sum_t w_t = 1$ . Under this model, Cho et al (2012b) proposed a registration-based method to estimate the homographies  $\{\mathcal{H}_t\}$  and the weights  $\{w_t\}$  by using the *Lucas-Kanade* algorithm (Szeliski and Shum 1997). This estimation method is extended into a multiple image deblurring scheme.

## 6.2 Homography Decomposition

This strategy decomposes the homography into a set of basic operations, i.e. representing  $\mathbf{H}_t$  in (6.7) and (6.8) as a weighted sum of *predefined* transformations or homographies. This leads to the formulation as

$$\mathbf{y} = \sum_{l=1}^L w_l \mathbf{H}_l \mathbf{x}_0, \quad (6.11)$$

where  $\{\mathbf{H}_l\}$  denotes the basis set with cardinality of  $L$ ,  $w_l$  is the weight assigned to the  $l$ -th basis, and  $\sum_l w_l = 1$ .

A typical example is the *transformation spread function* (TSF)-based modeling proposed by Chandramouli and Rajagopalan (2010). Their method focuses on the modeling of the camera's translations and in-plane rotations.  $\{\mathbf{H}_l\}$  is assumed to be the set of possible geometric transformations the image points can undergo during the exposure. TSF  $w(\mathbf{H}_l)$  is defined as

$$w(\mathbf{H}_l) : \{\mathbf{H}_l\} \rightarrow \mathbb{R}_+, \quad (6.12)$$

and (6.11) is rewritten as

$$\mathbf{y} = \sum_{l=1}^L w(\mathbf{H}_l) \mathbf{x}_{\mathbf{H}_l}, \quad (6.13)$$

where  $\mathbf{x}_{\mathbf{H}_l} = \mathbf{H}_l \mathbf{x}_0$  denotes the reference image  $\mathbf{x}_0$  warped by the transformation  $\mathbf{H}_l$ . A direct understanding of this formulation is similar to that in (6.10), i.e. the transformed images are weighted according to their exposure time. Depending on (6.13), Chandramouli et al. derived the relationship between TSF and the spatial variant PSF. Let  $h(\mathbf{i}; \mathbf{u})$  denote the PSF where  $\mathbf{i}$  is the location in the image plane and  $\mathbf{u}$  is the position in the 2D PSF. Then

$$h(\mathbf{i}; \mathbf{u}) = \sum_{l=1}^L w(\mathbf{H}_l) \delta(\mathbf{u} - (\mathbf{i} - \mathbf{i}_l)), \quad (6.14)$$

in which  $\delta(\cdot)$  is the Kronecker delta function, and  $\mathbf{i}_l$  denotes the coordinate of the point projected from  $\mathbf{i}$  according to  $\mathbf{H}_l$ . Applying this model to the deblurring problem, TSFs  $w(\cdot)$  are estimated by solving a least-squares problem. Vijay et al (2013) subsequently employed TSF-based modeling in the high dynamic range

(HDR) image reconstruction problem which involves blurry/noisy image pairs or multiple blurry images. In their formulation, the TV regularization and the image pair collaborative constraint are utilized to ensure the smoothness and recoverability of fine details. Note that the equation (6.13) is limited to the situation of constant depth (Chandramouli and Rajagopalan 2010). If there are depth variations in the scene, the blurry effect cannot be modeled using (6.13) where only one TSF is involved. Paramanand and Rajagopalan (2013) proposed a method involving two TSFs to handle the blurry image by capturing a bilayer scene which consists of two dominant layers of different depths. The two TSFs, each of which corresponds to a certain depth, are estimated as follows. Local blur kernels are first calculated at different image locations, followed by a grouping operation on the kernels according to the depth of each location. Each TSF is then estimated from the blur kernels in the corresponding depth layer under a regularization framework.

Another scheme for setting the basis set  $\{\mathbf{H}_l\}$  was proposed by Whyte et al (2010, 2012). Considering all-directional rotation of the camera, the homographies are correlated with the camera’s orientation. The resultant formulation is

$$y(\mathbf{i}) = \int_{\Theta} w(\theta) x_0(\mathbf{H}_{\theta} \mathbf{i}) d\theta, \quad (6.15)$$

and its discretized version is

$$y(\mathbf{i}) = \sum_{\theta} w(\theta) x_0(\mathbf{H}_{\theta} \mathbf{i}), \quad (6.16)$$

where  $\mathbf{H}_{\theta}$  is the homography specifying the orientation  $\theta$  of the camera rotation, and  $w(\theta)$  denotes the weighting function of  $\theta$ . By setting  $\{\mathbf{H}_{\theta}\}$  according to the rotations along the three Cartesian axes, the spatially variant blur kernel can be well-approximated. This model is applied to a single image deblurring problem using the variational Bayesian method, as well as a blurry/noisy image pair deblurring problem under the regularized least-squares formulation. Note that an integral over  $\Theta$  implies a search over the full orientation space, incurring a high computational cost when the camera rotates significantly. Instead, Hu and Yang (2012a) proposed to constrain the camera poses by imposing an initial guess of the pose subspace and searching the optimal solution within this subspace. This initialization of the pose subspace is implemented using back-projection. Compared with the method in (Whyte et al 2010, 2012), Hu and Yang (2012a)’s method produces more favorable results.

*Motion density function* (MDF) proposed by Gupta et al (2010) is similar to the above two schemes, and has the formulation of (6.11). Here the basis set  $\{\mathbf{H}_l\}$  is called the *motion response basis* (MRB) while  $\{w_l\}$  is called MDF. In the method, MRB is again pre-computed according to both camera translation and rotation, while the sharp image  $\mathbf{x}_0$  and the MDF  $\{w_l\}$  are iteratively optimized using the proposed RANSAC (short for "Random Sample consensus")-based scheme. Zheng et al (2013) applied this type of method to a more practical scenario that is forward/backward motion blur removal.

## 7 Region-based Methods

The spatial variant blur kernel implies that the local kernels vary from region to region, or more strictly, from point to point. Region-based methods express the blur model in that each region, or ideally, each point is assigned to a local consistent kernel (Levin 2006). The local kernel at the location  $i$  of the image is denoted as  $h_i$ , and the windowing operation extracting the fixed size patch whose center is located at  $i$  is denoted as  $\omega_i$ . The blurry image is then modeled as

$$y_i = c(h_i * (\omega_i \odot x)), \quad (7.1)$$

where  $c(\cdot)$  is a function to extract the center point of the patch. Thus the set  $\{h_i\}$  forms the spatial variant blur model. Note that in equation (7.1), only the correspondence between  $y_i$  and the center of the patch  $h_i * (\omega_i \odot x)$  is accessible due to the operation of  $c(\cdot)$ , and thus it is impractical to estimate  $x$  and all  $\{h_i\}$ . By considering the redundancy information (in particular, the overlaps between sampled patches) in the modeling, the operator  $c(\cdot)$  is generally replaced by the averaging of overlaps, which results in the *efficient filter flow* (EFF) model proposed by Hirsch et al (2010) and Harmeling et al (2010),

$$y = \sum_{\{i\}} h_i * (\omega_i \odot x). \quad (7.2)$$

In (7.2), for computational efficiency, the overlapping patches are only extracted at the locations indicated by the set  $\{i\}$  (not all locations in the image).  $\omega_i$  encodes not only the extracting operation, but also the averaging operation. Thus its values are not necessarily 0 or 1 but in the range of  $[0, 1]$ , and furthermore the normalization constraint  $\sum_{\{i\}} \omega_i = \mathbf{1}$  should be satisfied. Thanks to the matrix-vector-multiplication expression of convolution, equation (7.2) can be rewritten as  $\mathbf{y} = \mathbf{H}\mathbf{x}$ , where

$$\mathbf{H} = \mathbf{Z}_y^T \sum_{\{i\}} \mathbf{C}_i^T \mathbf{F}^H \text{Diag}(\mathbf{F} \mathbf{Z}_h h_i) \mathbf{F} \mathbf{C}_i \text{Diag}(\omega_i). \quad (7.3)$$

Here  $\text{Diag}(v)$  is a diagonal matrix that has vector  $v$  along its diagonal,  $\mathbf{C}_i$  is the chopping matrix for the  $i$ -th patch in the image,  $\mathbf{F}$  and  $\mathbf{F}^H$  are respectively the Discrete Fourier Transform matrix and its Hermitian,  $\mathbf{Z}_h$  is the zero-padding matrix used to expand the  $h_i$  by zero to the patch size, and  $\mathbf{Z}_y^T$  chops out the valid part of the space-variant convolution. Regarding equation (7.2) as  $\mathbf{y} = \mathbf{X}\mathbf{h}$ , in which the vector  $\mathbf{h}$  is obtained by stacking  $h_1, \dots, h_R$ ,  $\mathbf{X}$  is given by

$$\mathbf{X} = \mathbf{Z}_y^T \sum_{\{i\}} \mathbf{C}_i^T \mathbf{F}^H \text{Diag}(\mathbf{F} \mathbf{C}_i \text{Diag}(\omega_i) x) \mathbf{F} \mathbf{Z}_h \mathbf{B}_i, \quad (7.4)$$

where  $\mathbf{B}_i$  denotes a matrix that chops the blur kernel  $h_i$  from  $\mathbf{h}$ . Equations (7.3) and (7.4) provide an efficient way to obtain  $\mathbf{H}$ ,  $\mathbf{H}^T$ ,  $\mathbf{X}$  and  $\mathbf{X}^T$ , which are essential to iteratively estimate  $\{h_i\}$  and  $x$ . The experiments on processing of both atmospheric turbulence blur and camera shake blur show promising recovery accuracy and computational efficiency.

In their subsequent work, Schuler et al (2011) apply this framework to the problem of correcting optical aberrations, which exploits three channels (i.e. RGB) of the captured image. The problem is formulated as a joint objective function that combines respective constraints on three channels. Additionally, by incorporating the idea discussed in Section 6.2, Hirsch et al (2011) and Schuler et al (2012) reformulated (7.2) to model the camera rotation blur by representing the local blur kernel as the weighted sum of a set of bases, i.e.

$$y = \sum_{\{i\}} \left( \sum_{k=1}^K \mu_k b_{k,i} \right) * (\omega_i \odot x), \quad (7.5)$$

where  $\{\mu_k\}$  are the weights assigned to the bases  $\{b_{k,i}\}$ . It should be noted that  $\mu_k$  is independent of the patch with index  $i$ , whereas  $b_{k,i}$  depends that patch. This is because when modeling the camera rotation blur, different locations in the sensor may undergo very different routes, i.e. the local kernel in a specific point is expressible only by the basis kernels at the same point, each of which is induced by different rotation. Specifically, for modeling the camera rotation blur, we need to first generate the kernel maps, each of which consists of the local kernel for all points in an image, for different rotations, and then extract the local kernels at the same point, say  $i$ , of all maps to constitute the bases  $\{b_{k,i}\}$ . In such a situation,  $\{\mu_k\}$  and  $x$  are unknown variables in equation (7.5), whose number is significantly decreased. By employing the edge emphasizing operation in the pre-steps of optimization, the deblurring results on shaken blurred images and optical aberrant images are impressive. Apart from modeling camera rotation blur, this scheme is applicable to modeling general camera shaken blur, air turbulence blur, as well as defocus blur.

In contrast to the above EFF methods, Ji and Wang (2012b) proposed an intuitive approach to estimate point-wise blur kernels. In camera egomotion, the sensor is forced to undergo rigid translation or rotation. This rigidness naturally implies that without the consideration of depth change, the variation of the local kernels between two adjacent points is limited, meaning that if we have two kernels located in nearby positions, the other kernels located within the line between these two points can be recovered by a kernel interpolation scheme. Based on this strategy, Ji and Wang (2012b)'s method initially estimates the local kernels for overlapping regions using existing spatially invariant blind deblurring methods. However, because the scene depth varies in reality, some of the estimated kernels are erroneous and should be corrected according to the affine relationship between the adjacent regions. The authors then presented a kernel interpolation algorithm for transforming the region-wise kernels into point-wise kernels. Note that either kernel initialization or interpolation inevitably introduces errors to the kernel estimation. Based on (Ji and Wang 2012a), the induced error term is considered in a non-blind deblurring model which is an analysis formulation (cf. equation (4.29) in Section 4.3).



Camera translational blur is spatially variant when capturing a scene with depth variation. [Xu and Jia \(2012\)](#) developed a model by discovering the depth information of the scene. The depth layers, each of which corresponds to a local region, are produced by a stereo-matching-based disparity estimation algorithm that involves two blurry images of different views. In the estimation of the blur kernel for each layer, a tree structure is constructed over all depth layers, where the nodes for two layers having neighboring depths are merged to form their parent node. The blur kernels are then estimated from the top-level of the tree to the bottom-level, in which the kernel refinement is operated to make the estimation robust. Once the kernels are estimated, the two images are simultaneously deblurred in a unified formulation. To produce a pleasant result, the whole process from disparity estimation to blur removal is required to repeat at least once.

In terms of object motion blur, the blurry image can typically be segmented into two regions, where one is for the moving object inducing the blur and the other is generally the static background that has probably been corrupted by defocus blur. In this case, [Chakrabarti et al \(2010\)](#) determined the spatially variant blur model by mining the statistics of the localized frequency representation, in which each local region of the blurry image is independently transformed into the frequency domain. To reduce the number of unknown parameters, pixels in one region correspond to a same local kernel. To handle the boundary effects in convolution, the statistics, in particular the first- and second-order moments, are involved in a  $\text{MAP}_k$  formulation. Solving the  $\text{MAP}_k$  provides the estimation of the blur kernel. What follows is a binary segmentation of the motion-blurred foreground from the background via a Markov Random Field modeling strategy. The non-blind deblurring procedure is conducted independently in each region. Similarly, [Kim et al \(2013\)](#) addressed to deblur an image which contains multiple moving objects. Each object occupies a region and corresponds to a single kernel. In this method, the blur segmentation, the kernel estimation and the image restoration are alternatively processed in a unified variational formulation (cf. equation (4.3) in Section 4).

## 8 Others

In this section, we will discuss several methods which cannot be grouped into the above categories.

**Projection-based method:** The projection-based methods involve two terms, one of which concerns knowledge about the true solution that can be incorporated into the prior constraint set, while the other encodes the noisy blurry image that specifies the observation constraint set. Similar to the variational framework, this type of method also has the objective of (4.1), containing the above two terms. Specifically, [Li \(2011\)](#) proposed the following weighted deblurring function:

$$\min_{\mathbf{x}} \frac{1}{2} \sum_{i=0}^r \|\mathcal{B}_i(\mathbf{y} - \mathbf{H}\mathbf{x})\|_2^2 + \lambda \Psi(\mathbf{x}), \quad (8.1)$$

where  $r$  is a positive integral parameter, and the operator  $\mathcal{B}_i$  is defined as

$$\mathcal{B}_i = (\mathbf{I} - \beta \mathbf{H}\mathbf{H}^*)^i, \quad i = 0, \dots, r, \quad (8.2)$$

and  $0 < \beta < 1$ . The first term in (8.1) forms a variant of the problem in the traditional Landweber-iteration algorithm ([Landweber 1951](#)) which is used to deal with the ill-posed linear inverse problem, and can be solved by the  $r$ -times Landweber iteration. As noted by Li, the introduction of the  $r$  operators  $\{\mathcal{B}_i\}$  has three advantages. The first is that in the variational framework, the estimated solution is sensitive to the setting of the regularization parameter  $\lambda$ . However in (8.1), the degree of regularization is scaled by a factor of  $1/r$ , because of the addition of  $\{\mathcal{B}_i\}$ . Thus we can adjust  $r$  instead of  $\lambda$  to control the impact of regularization. Second, operators  $\{\mathcal{B}_i\}$  generalize the definition of deblurring error which is determined by  $r$ . When  $r = 0$ , the functional (8.1) reduces to the standard variational formulation. When  $r \rightarrow \infty$ , (8.1) becomes the standard Landweber iteration without regularization. Therefore by setting  $r \in (0, \infty)$ , the global minimum can fall everywhere except the two extremes. The third advantage is that based on the eigenvalue analysis,  $\{\mathcal{B}_i\}$  can be deemed as a weighting scheme for improving the numerical stability of the Landweber iteration. To solve (8.1), Li proved that the minimization can be realized by a concatenation of  $r$ -times Landweber iteration (projection on the observation constraint set) and regularized filtering (projection on the prior constraint set). Formally, denoting the projection operator of  $r$ -times Landweber iteration as

$\mathcal{P}_l$  and the projection operator of regularized filtering as  $\mathcal{P}_r$ , the optimization is then iterated between  $\mathbf{x}^{k+1/2} = \mathcal{P}_l \mathbf{x}^k$  and  $\mathbf{x}^{k+1} = \mathcal{P}_r \mathbf{x}^{k+1/2}$  until convergence.

**Kernel regression:** Kernel regression techniques have been widely used in image denoising. However, for deblurring, this type of method has received relatively less attention. To understand how a kernel regression framework handles the ill-posedness, local similarity is conventionally investigated to ensure the smoothness of local structures, whereas more recently, the focus has moved to non-local kernels that can efficiently exploit the repetitive patterns in the blurry image. Mathematically, kernel regression computes the central pixel of a window as follows:

$$\mathbf{x}_i^* = \underset{\mathbf{x}_i}{\operatorname{argmin}} \sum_{j \in \mathcal{N}(i)} (\mathbf{y}_j - \mathbf{x}_i)^2 \mathbf{K}_i(j - i), \quad (8.3)$$

where  $i$  and  $j$  are the locations,  $\mathbf{x}_i$  and  $\mathbf{y}_j$  are the corresponding pixels,  $\mathcal{N}(i)$  denotes the neighbors of location  $i$ , and  $\mathbf{K}_i(\cdot)$  is a generic spatial kernel at  $i$  which typically assigns large weights to the nearby similar pixels while assigning small weights to the farther dissimilar pixels. To apply (8.3) to deblurring, Takeda et al (2008) derived a formulation for kernel-based deblurring through exploiting the Taylor expansion of local structures. In their work, two types of locally adaptive kernel, i.e. bilateral kernel function and steering kernel function, are utilized in the proposed framework. The bilateral kernel function considers both the similarity of pixel values and the distance of spatial locations, while the steering kernel function is adaptive to the local gradients. Zhang et al (2013b) extended the local kernel regression to the nonlocal case, in which the kernel  $\mathbf{K}(\cdot)$  encodes the nonlocal similarity of high orders by searching similar patches across the whole image and calculating the zero-, first- and second-order similarities. In their formulation, both local kernel and nonlocal kernel are incorporated since the local structures regularize the noisy candidates found by the nonlocal similarity search, and nonlocal similarity provides the redundancy that prevents possible overfitting of the local kernel regression. In the application of deblurring, the authors pre-deconvolved the image using Wiener filtering, which amplifies the effect of the noise. The proposed nonlocal kernel regression is then employed to suppress the noise, resulting in a deblurred image.

**Stochastic deconvolution:** It is not easy to develop specific optimization algorithms to solve the deblurring problems defined under the Bayesian framework (or the variational framework) due to the complex priors (or the regularizers). To derive a general-purpose deconvolution algorithm, Gregson et al (2013) proposed stochastic deconvolution, which is not only capable of effectively handling arbitrary priors, but also of tackling the boundary conditions, saturated pixels and spatial variant kernels. Stochastic deconvolution is based on the random walk optimization strategy from Stochastic Tomography (Gregson et al 2012), in which a stochastic coordinate-descent method employs a Metropolis-Hasting style heuristic for picking the next coordinate axis to descend along. Specifically, this strategy picks a single pixel in each iteration and checks whether the objective can be improved by adding (or removing) an energy quantum to (or from) this pixel. Any change will be recorded if the objective is improved, otherwise the process will jump to the next pixel.

**Spectral analysis:** The power spectra of sharp natural images exhibit canonical behaviors expressing strong regularities (Field 1987; Burton and Moorhead 1987):

$$|\hat{x}(\boldsymbol{\omega})|^2 \propto \|\boldsymbol{\omega}\|^{-\beta}, \quad (8.4)$$

for  $\boldsymbol{\omega} \neq (0,0)$ , where  $\hat{x}$  is the Fourier transform of the sharp image  $x$ ,  $\boldsymbol{\omega}$  is the frequency coordinates. However, such behaviors will be disturbed by the degradation of the blur effect, which might display statistical irregularities. Goldstein and Fattal (2012) derived the power spectrum of the blurry image  $y$  as

$$|\widehat{y * d}(\boldsymbol{\omega})|^2 \approx c |\hat{h}(\boldsymbol{\omega})|^2, \quad (8.5)$$

where  $d$  denotes the spectral whitening operation and  $c$  is a constant. Based on this relationship, the power spectrum of the blur kernel  $h$  can be recovered from the statistics of the whitened spectrum of  $y$  (Goldstein and Fattal 2012; Hu et al 2012). Further, recovering the kernel  $h$  given its power spectrum  $|\hat{h}|^2$ , requires the estimation of the phase component of  $\hat{h}(\boldsymbol{\omega})$ . This mission is generally accomplished by a phase retrieval technique. For example, Goldstein and Fattal (2012) proposed a robust version of the relaxed averaged alternating reflections (RAAR), which was originally developed by Luke (2005) for phase retrieval. Hu et al (2012) recovered the phase using the error-reduction phase retrieval algorithm (Fienup 1982).

## 9 Practical Issues

### 9.1 Boundary

In image deblurring tasks, pixels located around the boundary of the blurry image are dependent upon the unknown pixels outside the observed region. Inappropriate processing on these pixels can bring severe artifacts. Typically, several kinds of *boundary condition* (BC) are utilized to formulize the boundary issue. For example, the periodic BC, which assumes a periodic convolution, is frequently used. This condition facilitates the implementation of fast Fourier transform (FFT) and thus speeds up the optimization. Alternatively, two other ways to define the condition are zero BC and reflexive BC. The zero BC pads the external region with zero values, while the reflexive BC indicates that the pixels outside the image are a reflection of those near the boundary but in the image. Even though these BCs make the deblurring tasks addressable and computationally convenient, they are intrinsically an approximate procedure and do not correspond to the real imaging systems. Deconvolution using these BCs could produce staircase artifacts in the deblurred image. To properly handle the boundary issue, Almeida and Figueiredo (2013a) and Matakos et al (2013) proposed the integration of a masking scheme into the image blurring model, which can be formulated as

$$\mathbf{y}_M = \mathbf{M}\mathbf{H}\mathbf{x} + \mathbf{n}, \quad (9.1)$$

where  $\mathbf{M}$  is a masking matrix designed to select only the subset of pixels which do not depend on the boundary pixels, and  $\mathbf{y}_M = \mathbf{M}\mathbf{y}$ . In this case, the assumed BC for the convolution is then irrelevant to the deconvolution.

Note that the above model implies that the pixels around the boundary but inside the image cannot be estimated. These pixels, however, are calculated in Ji and Wang (2012a)’s method and Sorel (2012)’s method by extending the size of the mask to involve all the pixels of the observed image. This model can be summarized as

$$\mathbf{y} = \mathbf{M}\mathbf{H}\mathbf{x} + \mathbf{n}, \quad (9.2)$$

which biasedly estimates the boundary pixels. In practice, the above models (defined by (9.1) and (9.2)) can be accordingly employed by considering their specific properties.

### 9.2 Noise and Outliers

Noises in images are usually caused by insufficient exposure. The longer the exposure time is, the lower the noise level will be exhibited. In general conditions, we can therefore expect that the noises in blurry images have not reached a sensitive level, and these noises can be effectively removed by appropriately choosing the parameters of the noise model in the Bayesian inference framework (Section 3) or the regularization parameters in variational methods (Sections 4 and 5). Nevertheless, noises should be carefully handled in extreme cases, such as in low-light conditions, or when capturing a very fast object. This is because deblurring an image with noticeable noise will produce ringing artifacts in the results. To handle this issue, Tai and Lin (2012) applied an existing denoising algorithm as a preprocessing step, and successively conducted blind deconvolution on the denoised image to estimate the blur kernel and the sharp image. However, a drawback of this method has been noticed by Zhong et al (2013), i.e. the denoising operation can bias the accurate estimation of the kernel. Thus, Zhong et al. designed a set of denoising filters based on the directional filters so that the denoising operation has no effect on the estimated kernel.

Another source of deblurring that should be adequately addressed is the outliers. According to Cho et al (2011a)’s definition, outliers include all factors which cannot be explained by the linear model (2.3), e.g. saturated/clipped pixels, non-Gaussian noise and nonlinear CRF. If these outliers are processed in the same way as the inliers, the resultant image will show severe ringing artifacts. In Cho et al (2011a)’s method, the blurry image is classified into two parts, i.e. inliers and outliers. Different statistical assumptions are respectively imposed on these two parts, and the estimation of the sharp image and the classification of the inlier/outlier are alternated until a reasonable result is obtained.

## 10 Promising Future Directions

Most of conventional methods for image deblurring process a single blurry image, in either non-blind or blind ways. The non-blind methods try to suppress the artifacts in the resultant image, while the blind methods are developed to recover an accuracy blur kernel. However, a single image has very limited information for the deblurring task. Although the blur kernel is known in the non-blind case, the high frequency information which is lost in the blurring process, is difficult to be restored from the degraded image. On the other hand, the deblurring problem is underdetermined in the blind case since the parameters to be estimated (i.e.,  $x$  and  $h$ ) outnumber the observations. In both cases, the performance of deblurring using single image is limited.

Effective options to resolve the above problem include 1) **involving multiple images, each of which can provide different information in recovering the sharp target image, and 2) improving the camera systems to capture more detailed information that can be exploited in the recovering process.** Regarding these options, two promising directions, *learning-based deblurring* and *hardware modifications*, have emerged in recent years, which are discussed in this section.

Another possible but difficult direction is to develop new models for single image deblurring, especially for blind single image deblurring. The recent progresses in this area include the  $\text{MAP}_h$  scheme, homography-based modeling, efficient filter flow modeling (please refer to Sections 3, 6, 7 respectively). However, the performance of these methods are far from optimal, and potential techniques should be exploited.

### 10.1 Learning-based Deblurring

Learning-based deblurring methods explore machine learning techniques to learn how to restore a target image from additional images which can be easily obtained from other resources, e.g. the Internet. Since different images share similar local patterns, the images used for learning can provide both sharp information and high frequency details that can be used to deduce the target image, even though they have different contents. This strategy has also been incorporated in various image processing tasks, such as image denoising and image super-resolution.

**The first type of learning-based method is to learn a subspace where the sharp image may be located.** By extracting the local patterns from multiple sharp images, a subspace can be constructed, in which the target image and the sharp images share similar details and thus details in the target image can be accurately represented. For example, [Joshi et al \(2010b\)](#) proposed to construct one person’s eigenfaces (which form the subspace) by using this person’s multiple sharp facial images. To restore the same person’s blurry facial image, these eigenfaces are employed to constrain the corresponding sharp image. Similarly, [Dong et al \(2011b\)](#) and [Ni et al \(2011\)](#) introduced to build the patch space or the patch manifold by first clustering the patches extracted from a sharp image database, and then applying PCA on each cluster to form its corresponding subspace. The whole patch space is composed of the resultant subspaces, each of which provides specific local structural information, and can therefore constrain the local patterns of the target image. An alternative scheme to specify the subspace is to estimate the patch distribution from sharp image databases, by assuming that the local patterns of different natural images should follow the common patch distribution. A typical example was developed by [Zoran and Weiss \(2011\)](#), in which the GMM was learned to model the prior distribution of sharp patches. Straightforwardly using this prior under the MAP framework, however, could produce a deblurred image in which the restored patches might not follow the distribution of the patches sampled from the original sharp image. Thus, to reduce this problem, the authors proposed a regularization technique that minimizes the difference between the patch distribution of the target image and that of the learned GMM. [Sun et al \(2013\)](#) developed a nonparametric method to model the distribution of the edge patches extracted from the BSDS500 dataset ([Arbelaez et al 2011](#)). In this method, the nonparametric distribution is learned by clustering the extracted patches and then counting the percentage of the patches falling into each cluster.

**The second type of learning-based method is to learn a restoring function which recovers the sharp target image from its blurry version.** In this case, multiple sharp images together with their blurry correspondences are generally utilized to train the parameters of the restoring function. For example, [Schmidt et al \(2013\)](#) employed the *regression tree field* (RTF) to model a non-linear regressor that specifies the local deblurring parameters. Before deblurring an image, RTFs are learned by maximizing a peak signal-to-noise ratio-based loss function. Rather than directly learning a deblurring function, [Schuler et al \(2013\)](#) estimated

the deblurred image by using *direct deconvolution* (Hirsch et al 2011) in the Fourier domain, and then incorporated the *multilayer perceptron* (MLP) to remove the artifacts of that deblurred image. Note that in both methods, RTF and MLP are trained based on a collection of sharp images and their corresponding synthesized blurry counterparts.

Finally, the learning scheme has been employed to facilitate the blur-related tasks by exploiting additional sharp information. For example, Hu and Yang (2012b) developed a novel strategy to determine which region of the blurry image is good to deblur, or in other words, is good to estimate the blur kernel. To detect the good regions, *conditional random field* is employed as the learning scheme from which the local structural information can be well exploited. The regions are considered to be good if they could provide as many informative structures as possible. In addition, Couzinié-Devy et al (2013) presented a novel scheme to handle the non-uniform blur such as defocus blur and linear motion blur. In this scheme, the kernel size is regarded as the class label indicating how large the blur kernel is for each blurry pixel. A multi-label segmentation method is introduced to estimate each pixel’s kernel size. Once the kernel sizes for all pixels are obtained, the blurry image is restored within a variational framework.

The sharp information of additional images are beneficial not only for recovering the details of a target blurry image, such as in the cases of learning a subspace and learning a restoring function, but also for providing assistant information for deblurring tasks, such as in the cases of specifying the good regions and determining the kernel sizes. As demonstrated by the attractive experimental results in these methods, learning-based deblurring is a promising direction in this domain.

## 10.2 Hardware Modifications

Image deblurring has recently benefited from the development of imaging systems. A typical instance is that most camera systems have been integrated with the image stabilization modular, reducing the blur caused by slight shakes of the systems during exposure. However, this image stabilization technique has limited performance in some extreme conditions, such as when the captured object is moving fast, or in the low light environments. Fortunately, more advanced techniques have been proposed to facilitate the image deblurring task.

Computational photography, which is an emerging field, inspires many researchers to develop new techniques for deblurring. Coded exposure photography (Ding et al 2010; McCloskey et al 2012) is one kind of technique in computational photography. In conventional exposure technique, the camera records the photons in a successive duration. In this technique, however, the camera’s shutter is fluttered open and closed during the chosen exposure time, while the open-close period is determined by a binary pseudo-random shutter sequence. Its superiority is that the coded exposure photography can produce a blurry image preserving more high-frequency spatial details than the conventional technique, especially when the image contains large motions, textured backgrounds, or partially occluded regions. Considering this, the preserved details can make the deconvolution to be a well-posed problem. Additionally, the shutter sequence can be designed as velocity-dependent when capturing very fast moving object, such that the details of the moving object can be retained (McCloskey 2010). Another kind of technique is the coded aperture photography. The aperture allows the light to go through it, and at the same time, can block the light with different wavelengths by reshaping itself. The rationale behind this technique is that the light through different shapes of apertures provides different information, because the light with different wavelengths has different optical characteristics. Using this property, Zhou et al (2011) derived a criterion for selecting a pair of coded apertures, by which the light through the two apertures can preserve the complementary information covering a broad band of scene frequencies. The third kind of computational photography is the coded flash technique, which is specifically designed for low light conditions (McCloskey 2011). In traditional lighting technique, the flash is generally activated in a successive duration. In contrast, by coding the flash timing sequence, this technique can control the illuminance and the exposure time in photographing, which has a similar effect to the coded exposure. When capturing a moving object in low light environment, coded flash system can produce a well-exposed image, and meanwhile can make the PSF estimation easier.

Besides the computational photography, constructing a multi-camera system can provide multiple images of the same scene, each of which carries different information that benefits deblurring tasks. For example, Cho et al (2010b) developed an orthogonal parabolic camera to capture two images of the scene by moving the sensor with parabolic velocities and in two orthogonal directions. Tai et al (2010a) designed a hybrid



camera system combined of a high-resolution, low-frame-rate camera and a low-resolution, high-frame-rate camera. This system captures a high-resolution but blurry image, as well as multiple low-resolution but sharp images. A beam splitter is used to ensure that the alignment of different images is well solved. By exploiting the information in the sharp images, the blur in the high-resolution image can be removed. [Li et al \(2011b\)](#) proposed a hybrid system to capture two well-aligned images with a specific relationship (for example, one image is a 90°-rotated version of the other). This relationship facilitates the mathematical derivation of the sharp image from two blurry images, making the deblurring task more effective.

For deblurring tasks, researchers have also added additional components into the camera systems. For instance, instead of using the conventional color image sensors which only have red (R), green (G), blue (B) patterns, [Wang et al \(2012b\)](#) employed a new color image sensor which adds the panchromatic (pan) pixels to the original RGB pixels. Due to the high light sensitivity, these pan pixels can help to restore the images captured in low light conditions. [Bando et al \(2013\)](#) employed the focus sweep technique during exposure. Since the scene has varying depths or in-plane motion, the captured image may exhibit a spatially variant blur effect. By using Bando et al.’s method, the blur effect of the image can be turned to a near-spatially invariant scenario. [Joshi et al \(2010a\)](#) integrated a gyroscope and an accelerometer to estimate the camera’s acceleration and angular velocity, from which the translation and rotation of the camera can be calculated and used for deblurring.

Hardware modification is always an effective manner to remove the blurry effects in captured images. Benefitting from the advanced hardware techniques, more information of the scenes can be easily recorded, and thus facilitate the deblurring task. Although some achievements have been obtained, various techniques are waiting to be invented in the future.

## 11 Performance Evaluation

To complement the discussions of the previous sections, we provide experimental evaluations for representative techniques. The evaluation for image deblurring can be subjective quality assessment or objective quality assessment. Subjective quality assessment predicts the observers’ perception without a well-defined numerical quantification. Although the subjective image quality assessment is the most direct and most accurate metric to reflect a person’s perception, it is too subject to cater for different persons. In contrast, objective quality assessment metrics can operate in an automatic and numerical manner. The metrics used in this section include *peak signal-to-noise ratio* (PSNR) and *structural similarity* (SSIM), which are widely utilized to evaluate the performance of image deblurring algorithms. Assuming the restored image  $x$  and the un-degraded image  $\tilde{x}$  are in the range of  $[0, 255]$ , PSNR and SSIM are defined as

$$\text{PSNR} = 10 \cdot \log_{10} \left( \frac{N \times 255 \times 255}{\sum_i (x_i - \tilde{x}_i)^2} \right), \quad (11.1)$$

$$\text{SSIM} = \frac{(2\mu_x\mu_{\tilde{x}} + c_1)(2\sigma_{x\tilde{x}} + c_2)}{(\mu_x^2 + \mu_{\tilde{x}}^2 + c_1)(\sigma_x^2 + \sigma_{\tilde{x}}^2 + c_2)}, \quad (11.2)$$

where  $N$  is the total number of pixels,  $\mu_x$  and  $\mu_{\tilde{x}}$  are the means of  $x$  and  $\tilde{x}$ ,  $\sigma_x^2$  and  $\sigma_{\tilde{x}}^2$  are the variances of  $x$  and  $\tilde{x}$ ,  $c_1$  and  $c_2$  are two variables to stabilize the division with weak denominator.

According to different task settings, existing methods can be categorized as non-blind uniform deblurring, blind uniform deblurring and non-uniform deblurring. Non-uniform deblurring is always blind since the kernel is hard to obtain in real applications. For each kind of setting, we evaluate the representative methods where the codes are publicly accessible.

### 11.1 Non-blind Uniform Deblurring

The test images used for non-blind uniform deblurring come from ([Levin et al 2009, 2011b](#)), including 4 images and 8 blur kernels (see Fig. 9). To generate blurry images, the 4 sharp images are corrupted by each blur kernels and additive white Gaussian noise with standard deviation  $\sigma = 0.01$ , resulting in 32 degraded images. The methods are evaluated on these synthetic blurry images, including [Schuler et al \(2013\)](#), [Zoran and Weiss \(2011\)](#), [Schmidt et al \(2011\)](#), [Danielyan et al \(2012\)](#), [Dong et al \(2013\)](#), [Cho et al \(2011a\)](#), [Afonso et al \(2010\)](#), [Almeida and Figueiredo \(2013a\)](#), and [Gregson et al \(2013\)](#).





Figure 9: First row: 4 test images. Second row: 8 blur kernels.

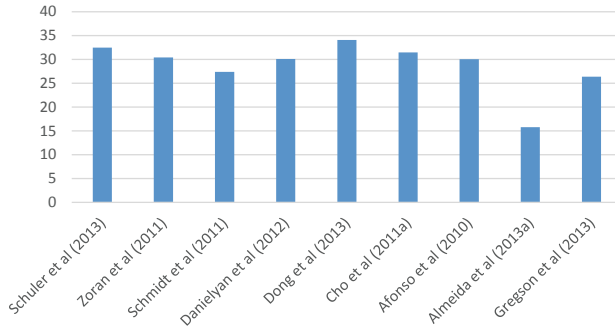


Figure 10: Mean PSNR (dB) results of non-blind uniform deblurring.

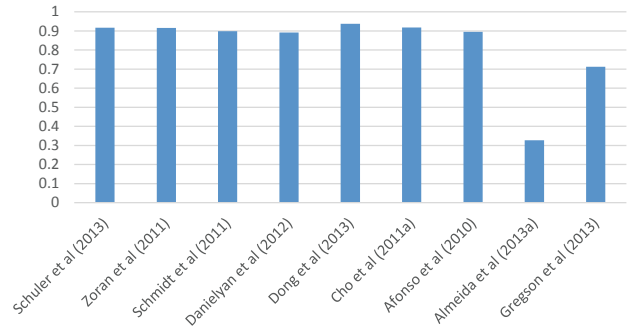


Figure 11: Mean SSIM results of non-blind uniform deblurring.

The objective assessment on the deblurred images obtained by these methods are shown in Fig. 10 and Fig. 11, while the visual comparisons of the deblurring results on an exemplary image are shown in Fig. 12. In summary, [Schuler et al \(2013\)](#), [Zoran and Weiss \(2011\)](#) and [Schmidt et al \(2011\)](#) employ learning schemes to constrain the restored images by exploring additional sharp information, and the resultant images exhibit promising recoveries. Compared with these three methods, the best performance acquired by [Dong et al \(2013\)](#) suggests that the nonlocal similar patterns discovered in the blurry image are more helpful for restoration than the information from other irrelevant sharp images. This is mainly because the nonlocal patterns in a blurry image are more stable and consistent than those in other images, and therefore have good expressivity. The results of both [Cho et al \(2011a\)](#) and [Schmidt et al \(2011\)](#) recommend that an appropriate processing of outliers and noises is important, otherwise like [Almeida and Figueiredo \(2013a\)](#), the noise in the restored image (Fig. 12(i)) might be severe.

## 11.2 Blind Uniform Deblurring

We consider 10 evaluated methods for blind uniform deblurring, including [Zhang et al \(2013a\)](#), [Sun et al \(2013\)](#), [Xu and Jia \(2010\)](#), [Levin et al \(2011a\)](#), [Babacan et al \(2012\)](#), [Krishnan et al \(2011\)](#), [Cai et al \(2012\)](#), [Cho et al \(2011b\)](#), [Goldstein and Fattal \(2012\)](#), and [Zhong et al \(2013\)](#). The algorithms are applied on the same set of test images as shown in Fig. 9.

Fig. 13 and Fig. 14 illustrate the performance of the evaluated methods, while the difference in visual quality between the methods can be inspected in the examples shown in Fig. 15. Through the joint restoration of multiple blurry images, [Zhang et al \(2013a\)](#)'s method acquires the best averaged results among the methods. [Sun et al \(2013\)](#) obtained comparable performance by using a learning strategy to estimate



Figure 12: An illustrative example of non-blind uniform deblurring results by different methods. The blurry image is synthesized with the first image and the first kernel in Fig. 9.

the patch prior. These observations are consistent with the discussions in Section 10. The superiority of (Xu and Jia 2010) reveals the importance of detecting useful edges in kernel estimation phase. Both Levin et al (2011a) and Babacan et al (2012) incorporated the  $\text{MAP}_h$  strategy to estimate the blur kernel and then conducted a non-blind deconvolution, whose effectiveness can be justified as in the figures, even though the averaged performance of (Levin et al 2011a) is limited. The difference between these two methods indicates that a general image prior (i.e. Gaussian scale model) is crucial in the MAP framework. Krishnan et al (2011)’s method, which uses a normalized sparsity measure, is expected to behave promisingly; but in the experiments the performance is sensitive to parameter settings. As in the non-blind deblurring tasks, the performance obtained by Zhong et al (2013)’s method also suggests that the noise handling is important in blind deblurring task. In kernel estimation, neither the spectral analysis (Goldstein and Fattal 2012) nor relatively simple methods such as the framelet (Cai et al 2012) and Radon transform (Cho et al 2011b) appear to be competitive.

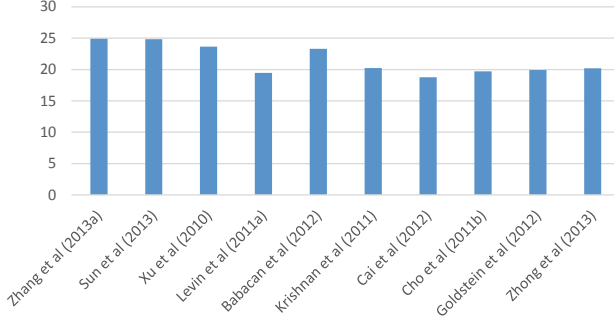


Figure 13: Mean PSNR (dB) results of blind uniform deblurring.

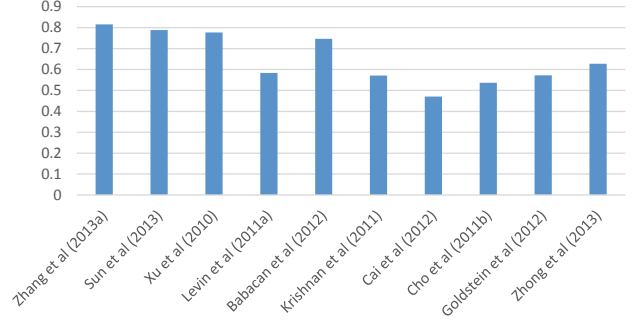


Figure 14: Mean SSIM results of blind uniform deblurring.

### 11.3 Non-uniform Deblurring

Non-uniform blur simulates the effect in real captured images, but is hard to synthesize. As a result, [Köhler et al \(2012\)](#) presented a benchmark dataset for simulating the camera shakes, which cause non-uniform blur. However, existing methods cannot perform well on this dataset, and are generally sensitive to parameter settings. To provide illustrative examples, we employ 3 real blurry images for evaluation, where the blur is mainly caused by camera shake, as shown in Fig. 16. Non-uniform deblurring methods considered for evaluation here include [Xu et al \(2013\)](#), [Whyte et al \(2010, 2012\)](#), and [Hu and Yang \(2012a\)](#). Since the corresponding sharp images are inaccessible, we perform the visual comparison in Fig. 17. From the resultant images, we can see that non-uniform deblurring is a difficult task because of the limited available information. The models used in these methods can only simulate specific motion blur, and therefore have strong restrictions. This property could help them perform well in some cases (e.g. the left-column images and the middle-column images in Fig. 17 which become sharper than the original), whereas reduce their effectiveness in other cases (e.g. the right-column images in Fig. 17 which are still blurry).

## 12 Conclusion

In this paper, we reviewed recent developments for image deblurring, including non-blind/blind, and spatially invariant/variant techniques. We mainly classify these methods into five categories according to the manners of handling ill-posedness: Bayesian inference framework, variational methods, sparse representation-based methods, homography-based modeling, and region-based methods. Bayesian inference framework-based approaches can be grouped into three sub-categories: MAP methods, MMSE methods, and variational Bayesian methods. Researches on variational framework are focused on the development of various regularization techniques and optimization methods, as well as the interpretation of analysis operators, synthesis operators and their combinations. Sparse representation-based methods assume that the sharp images share similar geometric structures which form a redundant dictionary possessing strong expressivity. How to induce the dictionary as well as how to enhance the smoothness in the resultant images under the dictionary are two critical issues of the sparse representation framework. Homography-based modeling and region-based modeling are two techniques designed for spatially variant deblurring, where the former accounts for the temporary accumulation of photons, while the later concerns the spatial variation of the blur kernel. Other types of technique include projection-based methods, kernel regression, stochastic deconvolution, spectral analysis. Besides these mathematical considerations, hardware modification techniques are also attractive to improve the performance of image deblurring.

In these conclusive comments, we would like to discuss some theoretical and practical aspects of these developments which, we expect, will help readers regarding their future research:

- 1) In blind uniform deblurring, the  $\text{MAP}_{x,h}$  schemes under a sparse gradient prior will result in a delta kernel and a blurry image. However, the  $\text{MAP}_h$  settings are able to approximate the true blur kernel, which can be solved by variational Bayesian methods.
- 2) Even though the estimator (such as  $\text{MAP}_h$  estimator) is important in deblurring task, the selection of

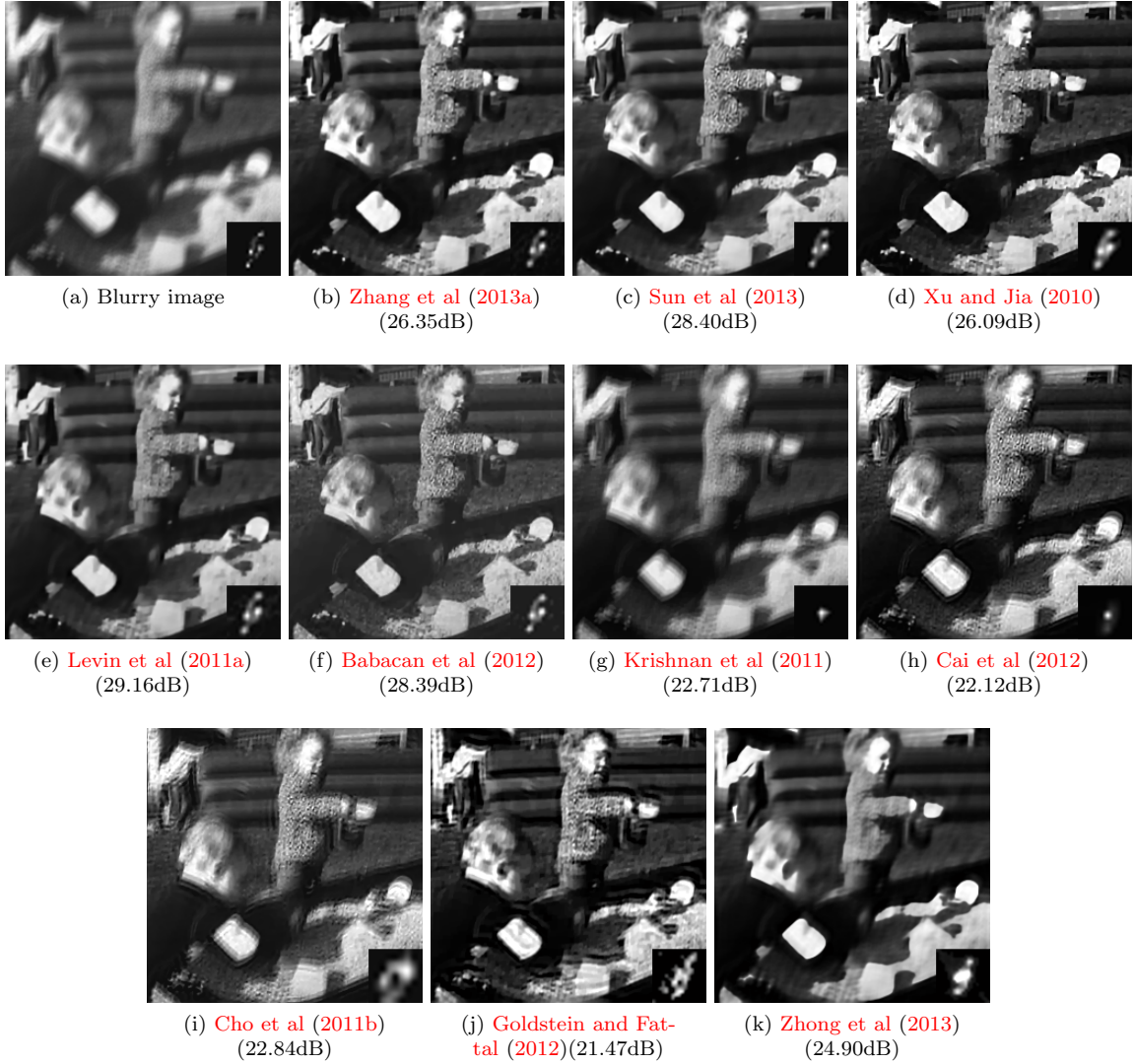


Figure 15: An illustrative example of blind uniform deblurring results by different methods. The blurry image is synthesized with the first image and the first kernel in Fig. 9.

prior is always critical. In blind uniform deblurring,  $\text{MAP}_h$  is guaranteed to provide an accurate kernel when the size of the blurry image is infinitely large. In real applications, however, the size of the obtained image is generally limited, under which condition an accurate kernel would be estimated by incorporating a proper prior (such as GSM). The same conclusion can also apply to the variational framework where the prior is replaced by the regularizer.

3) Either the analysis operator or the synthesis operator has specific properties benefitting the deblurring task. Combining these operators (such as BM3D) is appreciated for improving the deblurring performance.

4) Compared with human heuristics, data-driven techniques are promising for deriving priors in Bayesian inference framework, regularizers in variational framework, and dictionaries in sparse representation-based framework. The "data-driven" includes two types of aspect: the first is to explore the information from the target blurry image, while the second is to obtain the additional information from other sharp images (cf. the learning scheme in Section 5.2 and 10.1). The difference between these two types of aspect concerns that the first type can provide a better performance but in low efficiency, whereas the second type results in a slight degradation in the results but in high efficiency, since the learning procedure can be deployed in an offline manner.





Figure 16: Real captured blurry images.



Figure 17: Restored images by Xu et al (2013) (the first row), Whyte et al (2010, 2012) [103, 104] (the second row), and Hu and Yang (2012a) (the third row).

- 5) For deblurring, since an images can be viewed as a composition of large amount of repetitive patterns, exploiting the repetitive nature of images, which is often termed as a non-local strategy, is an effective way to boost the deblurring performance. This has been already proven in various image processing tasks including image deblurring, image denoising, image super-resolution, etc. Besides, the conventional local strategies can overcome a common issue of the non-local strategies that is the loss of local smoothness. Therefore, the combination of local and non-local strategies is recommended.
- 6) Spatially variant deblurring is a difficult task and the progress in this aspect is very limited. Incorporating a certain physical model (such as the modeling of object motion and the modeling of camera motion) is an option to improve the performance. However, this cannot satisfy the complex conditions in real applications.
- 7) Benefitting from the massive high-quality images available on the Internet and the advanced technologies of hardware, extra information can be recorded and employed for the subsequent deblurring as well as other restoration tasks. Learning-based deblurring and hardware modifications are therefore two promising directions to improve the performance of image deblurring.
- 8) Image deblurring is a crucial step to produce high-quality images for high-level vision tasks. However, when limited information is accessible, the deblurring performance could be restricted. In this case, a recommended option is to combine image deblurring and high-level vision tasks into a unified problem, such as classification of blurry images and detection of blurry objects. Then the deblurring task can benefit from the high-level vision tasks because they are generally human supervised. Therefore, the cross information among the images and among the tasks are worth to be exploited.

## References

- Afonso MV, Bioucas-Dias JM, Figueiredo MA (2010) Fast image recovery using variable splitting and constrained optimization. *IEEE Transactions on Image Processing* 19(9):2345–2356
- Aharon M, Elad M, Bruckstein A (2006) K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing* 54(11):4311–4322
- Almeida MS, Almeida LB (2010) Blind and semi-blind deblurring of natural images. *IEEE Transactions on Image Processing* 19(1):36–52
- Almeida MS, Figueiredo MA (2013a) Deconvolving images with unknown boundaries using the alternating direction method of multipliers. *IEEE Transactions on Image Processing* 22(8):3074–3086
- Almeida MS, Figueiredo MA (2013b) Parameter estimation for blind and non-blind deblurring using residual whiteness measures. *IEEE Transactions on Image Processing* 22(7):2751–2763
- Amizic B, Spinoulas L, Molina R, Katsaggelos AK (2013) Compressive blind image deconvolution. *IEEE Transactions on Image Processing* 22(10):3994–4006
- Arbelaez P, Maire M, Fowlkes C, Malik J (2011) Contour detection and hierarchical image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33(5):898–916
- Babacan SD, Molina R, Do MN, Katsaggelos AK (2012) Bayesian blind deconvolution with general sparse image priors. In: *Proceedings of European Conference on Computer Vision*, pp 341–355
- Bando Y, Holtzman H, Raskar R (2013) Near-invariant blur for depth and 2d motion via time-varying light field analysis. *ACM Transactions on Graphics* 32(2):13:1–13:15
- Barbero A, Sra S (2011) Fast newton-type methods for total variation regularization. In: *Proceedings of International Conference on Machine Learning*, pp 313–320
- Beck A, Teboulle M (2009) A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences* 2(1):183–202
- Bertero M, Boccacci P (1998) *Introduction to Inverse Problems in Imaging*. CRC press



- Bioucas-Dias JM, Figueiredo MA (2007) A new twist: two-step iterative shrinkage/thresholding algorithms for image restoration. *IEEE Transactions on Image Processing* 16(12):2992–3004
- Bishop CM (2006) *Pattern recognition and machine learning*. New York: Springer
- Boyd S, Parikh N, Chu E, Peleato B, Eckstein J (2011) Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning* 3(1):1–122
- Brainard DH, Freeman WT (1997) Bayesian color constancy. *Journal of the Optical Society of America A* 14(7):1393–1411
- Bruckstein AM, Donoho DL, Elad M (2009) From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM review* 51(1):34–81
- Burton G, Moorhead IR (1987) Color and spatial structure in natural scenes. *Applied Optics* 26(1):157–170
- Cai JF, Ji H, Liu C, Shen Z (2009) Blind motion deblurring from a single image using sparse approximation. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp 104–111
- Cai JF, Chan RH, Nikolova M (2010) Fast two-phase image deblurring under impulse noise. *Journal of Mathematical Imaging and Vision* 36(1):46–53
- Cai JF, Ji H, Liu C, Shen Z (2012) Framelet-based blind motion deblurring from a single image. *IEEE Transactions on Image Processing* 21(2):562–572
- Candès EJ, Donoho DL (1999) Ridgelets: A key to higher-dimensional intermittency? *Philosophical Transactions of the Royal Society of London Series A: Mathematical, Physical and Engineering Sciences* 357(1760):2495–2509
- Candès EJ, Donoho DL (2000) Curvelets: A surprisingly effective nonadaptive representation for objects with edges. Tech. rep., DTIC Document
- Cannon M (1976) Blind deconvolution of spatially invariant image blurs with phase. *IEEE Transactions on Acoustics, Speech and Signal Processing* 24(1):58–63
- Carlavan M, Blanc-Féraud L (2012) Sparse poisson noisy image deblurring. *IEEE Transactions on Image Processing* 21(4):1834–1846
- Chakrabarti A, Zickler T, Freeman WT (2010) Analyzing spatially-varying blur. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp 2512–2519
- Chan RH, Dong Y, Hintermüller M (2010) An efficient two-phase  $\ell_1$ -tv method for restoring blurred images with impulse noise. *IEEE Transactions on Image Processing* 19(7):1731–1739
- Chandramouli P, Rajagopalan A (2010) Inferring image transformation and structure from motion-blurred images. In: *Proceedings of British Machine Vision Conference*, pp 73.1–73.12
- Chen SS, Donoho DL, Saunders MA (1998) Atomic decomposition by basis pursuit. *SIAM journal on scientific computing* 20(1):33–61
- Chen X, He X, Yang J, Wu Q (2011) An effective document image deblurring algorithm. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp 369–376
- Chen X, Li F, Yang J, Yu J (2012) A theoretical analysis of camera response functions in image deblurring. In: *Proceedings of European Conference on Computer Vision*, pp 333–346
- Cho H, Wang J, Lee S (2012a) Text image deblurring using text-specific properties. In: *Proceedings of European Conference on Computer Vision*, Springer, pp 524–537
- Cho S, Lee S (2009) Fast motion deblurring. In: *Proceedings of ACM SIGGRAPH*, pp 145:1–145:8

- Cho S, Wang J, Lee S (2011a) Handling outliers in non-blind image deconvolution. In: Proceedings of IEEE International Conference on Computer Vision, pp 495–502
- Cho S, Cho H, Tai YW, Lee S (2012b) Registration based non-uniform motion deblurring. *Computer Graphics Forum* 31(7):2183–2192
- Cho TS, Joshi N, Zitnick CL, Kang SB, Szeliski R, Freeman WT (2010a) A content-aware image prior. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 169–176
- Cho TS, Levin A, Durand F, Freeman WT (2010b) Motion blur removal with orthogonal parabolic exposures. In: IEEE International Conference on Computational Photography, pp 1–8
- Cho TS, Paris S, Horn BK, Freeman WT (2011b) Blur kernel estimation using the radon transform. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 241–248
- Cho TS, Zitnick CL, Joshi N, Kang SB, Szeliski R, Freeman WT (2012c) Image restoration by matching gradient distributions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34(4):683–694
- Couzin  -Devy F, Sun J, Alahari K, Ponce J (2013) Learning to estimate and remove non-uniform image blur. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 1075–1082
- Dabov K, Foi A, Katkovnik V, Egiazarian K (2007) Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on Image Processing* 16(8):2080–2095
- Danielyan A, Katkovnik V, Egiazarian K (2010) Image deblurring by augmented lagrangian with bm3d frame prior. In: Workshop on Information Theoretic Methods in Science and Engineering, Tampere, Finland, pp 16–18
- Danielyan A, Katkovnik V, Egiazarian K (2012) Bm3d frames and variational image deblurring. *IEEE Transactions on Image Processing* 21(4):1715–1728
- Delbracio M, Mus   P, Almansa A, Morel JM (2012) The non-parametric sub-pixel local point spread function estimation is a well posed problem. *International journal of computer vision* 96(2):175–194
- Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L (2009) Imagenet: A large-scale hierarchical image database. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 248–255
- Ding Y, McCloskey S, Yu J (2010) Analysis of motion blur with a flutter shutter camera for non-linear motion. In: Proceedings of European Conference on Computer Vision, pp 15–30
- Do MN, Vetterli M (2005) The contourlet transform: an efficient directional multiresolution image representation. *IEEE Transactions on Image Processing* 14(12):2091–2106
- Dong W, Zhang L, Shi G (2011a) Centralized sparse representation for image restoration. In: Proceedings of IEEE International Conference on Computer Vision, pp 1259–1266
- Dong W, Zhang L, Shi G, Wu X (2011b) Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization. *IEEE Transactions on Image Processing* 20(7):1838–1857
- Dong W, Zhang L, Shi G, Li X (2013) Nonlocally centralized sparse representation for image restoration. *IEEE Transactions on Image Processing* 22(4):1620–1630
- Donoho DL, Johnstone JM (1994) Ideal spatial adaptation by wavelet shrinkage. *Biometrika* 81(3):425–455
- Elad M (2010) *Sparse and redundant representations: from theory to applications in signal and image processing*. Springer
- Elad M, Aharon M (2006) Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing* 15(12):3736–3745
- Elad M, Starck JL, Querre P, Donoho DL (2005) Simultaneous cartoon and texture image inpainting using morphological component analysis (mca). *Applied and Computational Harmonic Analysis* 19(3):340–358

- Elad M, Milanfar P, Rubinstein R (2007) Analysis versus synthesis in signal priors. *Inverse problems* 23(3):947–968
- Faramarzi E, Rajan D, Christensen MP (2013) Unified blind method for multi-image super-resolution and single/multi-image blur deconvolution. *IEEE Transactions on Image Processing* 22(6):2101–2114
- Fergus R, Singh B, Hertzmann A, Roweis ST, Freeman WT (2006) Removing camera shake from a single photograph. In: *Proceedings of ACM SIGGRAPH*, pp 787–794
- Field D (1994) What is the goal of sensory coding? *Neural computation* 6(4):559–601
- Field DJ (1987) Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A* 4(12):2379–2394
- Fienup JR (1982) Phase retrieval algorithms: a comparison. *Applied optics* 21(15):2758–2769
- Gilboa G, Osher S (2007) Nonlocal linear image regularization and supervised segmentation. *Multiscale Modeling & Simulation* 6(2):595–630
- Gilboa G, Osher S (2008) Nonlocal operators with applications to image processing. *Multiscale Modeling & Simulation* 7(3):1005–1028
- Gilboa G, Sochen N, Zeevi YY (2002) Forward-and-backward diffusion processes for adaptive image enhancement and denoising. *IEEE Transactions on Image Processing* 11(7):689–703
- Goldstein A, Fattal R (2012) Blur-kernel estimation from spectral irregularities. In: *Proceedings of European Conference on Computer Vision*, pp 622–635
- Goldstein T, Osher S (2009) The split bregman method for l1-regularized problems. *SIAM Journal on Imaging Sciences* 2(2):323–343
- Gregson J, Krimerman M, Hullin MB, Heidrich W (2012) Stochastic tomography and its applications in 3d imaging of mixing fluids. In: *Proceedings of ACM SIGGRAPH*, pp 52:1–52:10
- Gregson J, Heide F, Hullin M, Rouf M, Heidrich W (2013) Stochastic deconvolution. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp 1043–1050
- Grossberg MD, Nayar SK (2004) Modeling the space of camera response functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26(10):1272–1282
- Guleryuz OG (2006a) Nonlinear approximation based image recovery using adaptive sparse reconstructions and iterated denoising-part i: theory. *IEEE Transactions on Image Processing* 15(3):539–554
- Guleryuz OG (2006b) Nonlinear approximation based image recovery using adaptive sparse reconstructions and iterated denoising-part ii: Adaptive algorithms. *IEEE Transactions on Image Processing* 15(3):555–571
- Gupta A, Joshi N, Zitnick CL, Cohen M, Curless B (2010) Single image deblurring using motion density functions. In: *Proceedings of European Conference on Computer Vision*, pp 171–184
- HaCohen Y, Shechtman E, Goldman DB, Lischinski D (2011) Non-rigid dense correspondence with applications for image enhancement. In: *Proceedings of ACM SIGGRAPH*, pp 70:1–70:10
- Hacohen Y, Shechtman E, Lischinski D (2013) Deblurring by example using dense correspondence. In: *Proceedings of IEEE International Conference on Computer Vision*, pp 2384–2391
- Harmeling S, Michael H, Schölkopf B (2010) Space-variant single-image blind deconvolution for removing camera shake. In: *Proceedings of Advances in Neural Information Processing Systems*, pp 829–837
- Hirsch M, Sra S, Schölkopf B, Harmeling S (2010) Efficient filter flow for space-variant multiframe blind deconvolution. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp 607–614

- Hirsch M, Schuler CJ, Harmeling S, Schölkopf B (2011) Fast removal of non-uniform camera shake. In: Proceedings of IEEE International Conference on Computer Vision, pp 463–470
- Hu W, Xue J, Zheng N (2012) Psf estimation via gradient domain correlation. *IEEE Transactions on Image Processing* 21(1):386–392
- Hu Z, Yang MH (2012a) Fast non-uniform deblurring using constrained camera pose subspace. In: Proceedings of British Machine Vision Conference, pp 136.1–136.11
- Hu Z, Yang MH (2012b) Good regions to deblur. In: Proceedings of European Conference on Computer Vision, pp 59–72
- Ji H, Wang K (2012a) Robust image deblurring with an inaccurate blur kernel. *IEEE Transactions on Image Processing* 21(4):1624–1634
- Ji H, Wang K (2012b) A two-stage approach to blind spatially-varying motion deblurring. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 73–80
- Jordan MI, Ghahramani Z, Jaakkola TS, Saul LK (1999) An introduction to variational methods for graphical models. *Machine learning* 37(2):183–233
- Joshi N, Szeliski R, Kriegman D (2008) Psf estimation using sharp edge prediction. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 1–8
- Joshi N, Kang SB, Zitnick CL, Szeliski R (2010a) Image deblurring using inertial measurement sensors. In: Proceedings of ACM SIGGRAPH, pp 30:1–30:9
- Joshi N, Matusik W, Adelson EH, Kriegman DJ (2010b) Personal photo enhancement using example images. *ACM Transactions on Graphics* 29(2):12:1–12:15
- Jung M, Bresson X, Chan TF, Vese LA (2011) Nonlocal mumford-shah regularizers for color image restoration. *IEEE Transactions on Image Processing* 20(6):1583–1598
- Kee E, Paris S, Chen S, Wang J (2011) Modeling and removing spatially-varying optical blur. In: Proceedings of IEEE International Conference on Computational Photography, pp 1–8
- Kenig T, Kam Z, Feuer A (2010) Blind image deconvolution using machine learning for three-dimensional microscopy. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32(12):2191–2204
- Keuper M, Schmidt T, Temerinac-Ott M, Padeken J, Heun P, Ronneberger O, Brox T (2013) Blind deconvolution of widefield fluorescence microscopic data by regularization of the optical transfer function (otf). In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 2179–2186
- Kim S, Tai YW, Kim SJ, Brown MS, Matsushita Y (2012a) Nonlinear camera response functions and image deblurring. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 25–32
- Kim S, Tai YW, Kim SJ, Brown MS, Matsushita Y (2012b) Nonlinear camera response functions and image deblurring, supplementary. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition
- Kim TH, Ahn B, Lee KM (2013) Dynamic scene deblurring. In: Proceedings of IEEE International Conference on Computer Vision, pp 3160–3167
- Köhler R, Hirsch M, Mohler B, Schölkopf B, Harmeling S (2012) Recording and playback of camera shake: Benchmarking blind deconvolution with a real-world database. In: Proceedings of European Conference on Computer Vision, pp 27–40
- Krishnan D, Tay T, Fergus R (2011) Blind deconvolution using a normalized sparsity measure. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 233–240

- Landweber L (1951) An iteration formula for fredholm integral equations of the first kind. *American Journal of Mathematics* 73(3):615–624
- Lefkimmiatis S, Unser M (2013) Poisson image reconstruction with hessian schatten-norm regularization. *IEEE Transactions on Image Processing* 22(11):4314–4327
- Lefkimmiatis S, Bourquard A, Unser M (2012) Hessian-based norm regularization for image restoration with biomedical applications. *IEEE Transactions on Image Processing* 21(3):983–995
- Levin A (2006) Blind motion deblurring using image statistics. In: *Proceedings of Advances in Neural Information Processing Systems*, pp 841–848
- Levin A, Fergus R, Durand F, Freeman WT (2007) Image and depth from a conventional camera with a coded aperture. In: *Proceedings of ACM SIGGRAPH*
- Levin A, Weiss Y, Durand F, Freeman WT (2009) Understanding and evaluating blind deconvolution algorithms. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp 1964–1971
- Levin A, Weiss Y, Durand F, Freeman WT (2011a) Efficient marginal likelihood optimization in blind deconvolution. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp 2657–2664
- Levin A, Weiss Y, Durand F, Freeman WT (2011b) Understanding blind deconvolution algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33(12):2354–2367
- Li F, Li Z, Saunders D, Yu J (2011a) A theory of coprime blurred pairs. In: *Proceedings of IEEE International Conference on Computer Vision*, pp 217–224
- Li W, Zhang J, Dai Q (2011b) Exploring aligned complementary image pair for blind motion deblurring. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp 273–280
- Li X (2011) Fine-granularity and spatially-adaptive regularization for projection-based image deblurring. *IEEE Transactions on Image Processing* 20(4):971–983
- Lin S, Zhang L (2005) Determining the radiometric response function from a single grayscale image. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp 66–73
- Lou Y, Zhang X, Osher S, Bertozzi A (2010) Image recovery via nonlocal operators. *Journal of Scientific Computing* 42(2):185–197
- Lucy LB (1974) An iterative technique for the rectification of observed distributions. *The astronomical journal* 79:745–754
- Luke DR (2005) Relaxed averaged alternating reflections for diffraction imaging. *Inverse Problems* 21(1):37–50
- Ma J, Le Dimet FX (2009) Deblurring from highly incomplete measurements for remote sensing. *IEEE Transactions on Geoscience and Remote Sensing* 47(3):792–802
- Ma L, Moisan L, Yu J, Zeng T (2013) A dictionary learning approach for poisson image deblurring. *IEEE Transactions on Medical Imaging* 32(7):1277–1289
- Matakos A, Ramani S, Fessler JA (2013) Accelerated edge-preserving image restoration without boundary artifacts. *IEEE Transactions on Image Processing* 22(5):2019–2029
- McCloskey S (2010) Velocity-dependent shutter sequences for motion deblurring. In: *Proceedings of European Conference on Computer Vision*, pp 309–322
- McCloskey S (2011) Temporally coded flash illumination for motion deblurring. In: *Proceedings of IEEE International Conference on Computer Vision*, pp 683–690

- McCloskey S, Ding Y, Yu J (2012) Design and estimation of coded exposure point spread functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34(10):2071–2077
- Michaeli T, Sigalov D, Eldar YC (2012) Partially linear estimation with application to sparse signal recovery from measurement pairs. *IEEE Transactions on Signal Processing* 60(5):2125–2137
- Michailovich OV (2011) An iterative shrinkage approach to total-variation image restoration. *IEEE Transactions on Image Processing* 20(5):1281–1299
- Miskin J, MacKay DJ (2000) Ensemble learning for blind image separation and deconvolution. In: *Advances in independent component analysis*, Springer, pp 123–141
- Montefusco LB, Lazzaro D (2012) An iterative  $\ell_1$ -based image restoration algorithm with an adaptive parameter estimation. *IEEE Transactions on Image Processing* 21(4):1676–1686
- Ng MK, Yuan X, Zhang W (2013) Coupled variational image decomposition and restoration model for blurred cartoon-plus-texture images with missing pixels. *IEEE Transactions on Image Processing* 22(6):2233–2246
- Ni J, Turaga P, Patel VM, Chellappa R (2011) Example-driven manifold priors for image deconvolution. *IEEE Transactions on Image Processing* 20(11):3086–3096
- Nishiyama M, Hadid A, Takeshima H, Shotton J, Kozakaya T, Yamaguchi O (2011) Facial deblur inference using subspace analysis for recognition of blurred faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33(4):838–845
- Osher S, Rudin LI (1990) Feature-oriented image enhancement using shock filters. *SIAM Journal on Numerical Analysis* 27(4):919–940
- Palmer J, Kreutz-Delgado K, Rao BD, Wipf DP (2005) Variational em algorithms for non-gaussian latent variable models. In: *Proceedings of Advances in neural information processing systems*, pp 1059–1066
- Papafitsoros K, Schönlieb CB (2014) A combined first and second order variational approach for image reconstruction. *Journal of mathematical imaging and vision* 48(2):308–338
- Paramanand C, Rajagopalan AN (2013) Non-uniform motion deblurring for bilayer scenes. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp 1115–1122
- Perrone D, Ravichandran A, Vidal R, Favaro P (2012) Image priors for image deblurring with uncertain blur. In: *Proceedings of British Machine Vision Conference*, pp 114.1–114.11
- Peyré G (2009) Manifold models for signals and images. *Computer Vision and Image Understanding* 113(2):249–260
- Pillai SU, Liang B (1999) Blind image deconvolution using a robust gcd approach. *IEEE Transactions on Image Processing* 8(2):295–301
- Richardson WH (1972) Bayesian-based iterative method of image restoration. *Journal of the Optical Society of America* 62(1):55–59
- Rostami M, Michailovich O, Wang Z (2012) Image deblurring using derivative compressed sensing for optical imaging application. *IEEE Transactions on Image Processing* 21(7):3139–3149
- Rudin LI, Osher S, Fatemi E (1992) Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena* 60(1):259–268
- Russo F, Ramponi G (1992) Fuzzy operator for sharpening of noisy images. *Electronics Letters* 28(18):1715–1717
- Schavemaker JG, Reinders MJ, Gerbrands JJ, Backer E (2000) Image sharpening by morphological filtering. *Pattern Recognition* 33(6):997–1012



- Schmidt U, Gao Q, Roth S (2010) A generative perspective on mrfs in low-level vision. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 1751–1758
- Schmidt U, Schelten K, Roth S (2011) Bayesian deblurring with integrated noise estimation. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 2625–2632
- Schmidt U, Rother C, Nowozin S, Jancsary J, Roth S (2013) Discriminative non-blind deblurring. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 604–611
- Schuler CJ, Hirsch M, Harmeling S, Schölkopf B (2011) Non-stationary correction of optical aberrations. In: Proceedings of IEEE International Conference on Computer Vision, pp 659–666
- Schuler CJ, Hirsch M, Harmeling S, Schölkopf B (2012) Blind correction of optical aberrations. In: Proceedings of European Conference on Computer Vision, pp 187–200
- Schuler CJ, Burger HC, Harmeling S, Schölkopf B (2013) A machine learning approach for non-blind image deconvolution. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 1067–1074
- Shaked E, Michailovich O (2011) Iterative shrinkage approach to restoration of optical imagery. *IEEE Transactions on Image Processing* 20(2):405–416
- Shen Z, Toh KC, Yun S (2011) An accelerated proximal gradient algorithm for frame-based image restoration via the balanced approach. *SIAM Journal on Imaging Sciences* 4(2):573–596
- Shepp LA, Vardi Y (1982) Maximum likelihood reconstruction for emission tomography. *IEEE Transactions on Medical Imaging* 1(2):113–122
- Sorel M (2012) Removing boundary artifacts for real-time iterated shrinkage deconvolution. *IEEE Transactions on Image Processing* 21(4):2329–2334
- Stockham Jr TG, Cannon TM, Ingebretsen RB (1975) Blind deconvolution through digital signal processing. *Proceedings of the IEEE* 63(4):678–692
- Sun L, Cho S, Wang J, Hays J (2013) Edge-based blur kernel estimation using patch priors. In: IEEE International Conference on Computational Photography, pp 1–8
- Szeliski R, Shum HY (1997) Creating full view panoramic image mosaics and environment maps. In: Proceedings of Annual Conference on Computer Graphics and Interactive Techniques, pp 251–258
- Tai YW, Lin S (2012) Motion-aware noise filtering for deblurring of noisy and blurry images. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 17–24
- Tai YW, Du H, Brown MS, Lin S (2010a) Correction of spatially varying image and video motion blur using a hybrid camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32(6):1012–1028
- Tai YW, Kong N, Lin S, Shin SY (2010b) Coded exposure imaging for projective motion deblurring. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 2408–2415
- Tai YW, Tan P, Brown MS (2011) Richardson-lucy deblurring for scenes under a projective motion path. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33(8):1603–1618
- Tai YW, Chen X, Kim S, Kim SJ, Li F, Yang J, Yu J, Matsushita Y, Brown MS (2013) Nonlinear camera response functions and image deblurring: Theoretical analysis and practice. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35(10):2498–2512
- Takeda H, Farsiu S, Milanfar P (2008) Deblurring using regularized locally adaptive kernel regression. *IEEE Transactions on Image Processing* 17(4):550–563

- Temerinac-Ott M, Ronneberger O, Ochs P, Driever W, Brox T, Burkhardt H (2012) Multiview deblurring for 3-d images from light-sheet-based fluorescence microscopy. *IEEE Transactions on Image Processing* 21(4):1863–1873
- Tibshirani R (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B (Methodological)* 58(1):267–288
- Tikhonov A (1963) Solution of incorrectly formulated problems and the regularization method. In: *Soviet Math. Dokl.*, vol 5, pp 1035–1038
- Tzeng J, Liu CC, Nguyen TQ (2010) Contourlet domain multiband deblurring based on color correlation for fluid lens cameras. *IEEE Transactions on Image Processing* 19(10):2659–2668
- Vijay CS, Paramanand C, Rajagopalan AN, Chellappa R (2013) Non-uniform deblurring in hdr image reconstruction. *IEEE Transactions on Image Processing* 22(10):3739–3750
- Wang C, Sun L, Cui P, Zhang J, Yang S (2012a) Analyzing image deblurring through three paradigms. *IEEE Transactions on Image Processing* 21(1):115–129
- Wang C, Yue Y, Dong F, Tao Y, Ma X, Clapworthy G, Lin H, Ye X (2013) Nonedge-specific adaptive scheme for highly robust blind motion deblurring of natural images. *IEEE Transactions on Image Processing* 22(3):884–897
- Wang S, Hou T, Border J, Qin H, Miller RL (2012b) High-quality image deblurring with panchromatic pixels. *ACM Transactions on Graphics* 31(5):120:1–120:11
- Wehrwein S, Scharstein AD (2010) Computational photography: Coded exposure and coded aperture imaging
- Whyte O, Sivic J, Zisserman A, Ponce J (2010) Non-uniform deblurring for shaken images. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp 491–498
- Whyte O, Sivic J, Zisserman A, Ponce J (2012) Non-uniform deblurring for shaken images. *International journal of computer vision* 98(2):168–186
- Wiener N (1949) *Extrapolation, interpolation, and smoothing of stationary time series*, vol 2. Cambridge, MA: MIT press
- Wright J, Yang AY, Ganesh A, Sastry SS, Ma Y (2009a) Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31(2):210–227
- Wright SJ, Nowak RD, Figueiredo MA (2009b) Sparse reconstruction by separable approximation. *IEEE Transactions on Signal Processing* 57(7):2479–2493
- Xie S, Rahardja S (2012) Alternating direction method for balanced image restoration. *IEEE Transactions on Image Processing* 21(11):4557–4567
- Xu L, Jia J (2010) Two-phase kernel estimation for robust motion deblurring. In: *Proceedings of European Conference on Computer Vision*, pp 157–170
- Xu L, Jia J (2012) Depth-aware motion deblurring. In: *Proceedings of IEEE International Conference on Computational Photography*, pp 1–8
- Xu L, Lu C, Xu Y, Jia J (2011) Image smoothing via  $l_0$  gradient minimization. In: *Proceedings of ACM SIGGRAPH*, pp 174:1–174:12
- Xu L, Zheng S, Jia J (2013) Unnatural  $l_0$  sparse representation for natural image deblurring. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp 1107–1114
- Yang J, Wright J, Huang TS, Ma Y (2010) Image super-resolution via sparse representation. *IEEE Transactions on Image Processing* 19(11):2861–2873

- Yang S, Wang J, Fan W, Zhang X, Wonka P, Ye J (2013) An efficient admm algorithm for multidimensional anisotropic total variation regularization problems. In: Proceedings of ACM SIGKDD, pp 641–649
- Yu G, Sapiro G, Mallat S (2012) Solving inverse problems with piecewise linear estimators: From gaussian mixture models to structured sparsity. *IEEE Transactions on Image Processing* 21(5):2481–2499
- Zhang H, Wipf D (2013) Non-uniform camera shake removal using a spatially-adaptive sparse penalty. In: Proceedings of Advances in Neural Information Processing Systems, pp 1556–1564
- Zhang H, Yang J, Zhang Y, Nasrabadi NM, Huang TS (2011) Close the loop: Joint blind image restoration and recognition with sparse representation prior. In: Proceedings of IEEE International Conference on Computer Vision, pp 770–777
- Zhang H, Wipf D, Zhang Y (2013a) Multi-image blind deblurring using a coupled adaptive sparse prior. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 1051–1058
- Zhang H, Yang J, Zhang Y, Huang TS (2013b) Image and video restorations via nonlocal kernel regression. *IEEE Transactions on Cybernetics* 43(3):1035–1046
- Zhang L, Deshpande A, Chen X (2010) Denoising vs. deblurring: Hdr imaging techniques using moving cameras. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 522–529
- Zhang Y, Hirakawa K (2013) Blur processing using double discrete wavelet transform. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 1091–1098
- Zheng S, Xu L, Jia J (2013) Forward motion deblurring. In: Proceedings of IEEE International Conference on Computer Vision, pp 1465–1472
- Zhong L, Cho S, Metaxas D, Paris S, Wang J (2013) Handling noise in single image deblurring using directional filters. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 612–619
- Zhou C, Lin S, Nayar SK (2011) Coded aperture pairs for depth from defocus and defocus deblurring. *International journal of computer vision* 93(1):53–72
- Zhu X, Milanfar P (2010) Image reconstruction from videos distorted by atmospheric turbulence. In: Proceedings of SPIE, pp 75,430S–75,430S–8, DOI 10.1117/12.840127
- Zhu X, Milanfar P (2011) Stabilizing and deblurring atmospheric turbulence. In: Proceedings of IEEE International Conference on Computational Photography, pp 1–8
- Zhu X, Milanfar P (2013) Removing atmospheric turbulence via space-invariant deconvolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35(1):157–170
- Zhuo S, Guo D, Sim T (2010) Robust flash deblurring. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 2440–2447
- Zibulevsky M, Elad M (2010) L1-l2 optimization in signal and image processing. *IEEE Signal Processing Magazine* 27(3):76–88
- Zoran D, Weiss Y (2011) From learning models of natural image patches to whole image restoration. In: Proceedings of IEEE International Conference on Computer Vision, pp 479–486