Signal Processing 103 (2014) 45–59

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Fast blockwise SURE shrinkage for image denoising

Yue Wu [a,b,*], Brian H. Tracey [a], Premkumar Natarajan [b], Joseph P. Noonan [a]

[a] Department of Electrical and Computer Engineering, Tufts University, 161 College Ave., Medford, MA 02155, USA
[b] Speech, Language and Multimedia Business Unit, Raytheon BBN Technologies, 10 Moulton St., Cambridge, MA 02138, USA

ARTICLE INFO

Article history:
Received 9 July 2013
Received in revised form
28 October 2013
Accepted 11 January 2014
Available online 20 January 2014

Keywords:
Image denoising
Shrinkage estimation
Non-local means
Adaptive algorithm
Stein's unbiased risk estimator

ABSTRACT

In this paper, we investigate the shrinkage problem of image denoising for various methods under the additive white Gaussian noise (AWGN) model. Our main contribution is to derive the closed-form of the optimal shrinkage that minimizes the Stein's unbiased risk estimator (SURE) and thus allows direct blockwise shrinkage without additional optimizations. Simulation results show that the proposed method boosts the denoising performance for a variety of image denoising techniques including the moving average filter, the median filter, the wiener filter, the bilateral filter, the probabilistic non-local means, and the block matching 3D filter in terms of higher pixel signal noise ratio (PSNR) and structural similarity index (SSIM). We also case study the proposed shrinkage solution with respect to the classic NLM denoising, whose shrinkage solutions and equivalent forms have been widely researched, and further confirm its superiority. The proposed shrinkage solution can be used to improve arbitrary image denoising methods under the AWGN model, and it serves as a good remedy to save badly denoised images due to inappropriate parameters.

1. Introduction

Because of possible noise introduced in image acquisition, transmission, storage and other stages, a digital image is often corrupted by a certain level of noise [1–3]. Though a low level of noise does not cause a visual difference for human beings, a high level of noise often leads to a noticeable loss of image quality. Image denoising is a process aiming to restore an original clean image from its noisy counterpart. A well denoised image not only looks visually appealing, but also paves a way to many advanced analysis tasks in image processing and computer vision, *e.g.* object recognition and character recognition [4,5].

Assume a two-dimensional $H \times W$ clean image $\mathbf{x} = \{x_l\}$ with $l = (l_r, l_c)$ and $l_r \in [1, H]$, $l_c \in [1, W]$ is contaminated by noise $\mathbf{n} = \{n_l\}$ through a noise model $f(\cdot)$. Then its noisy counterpart which we observe can be denoted as $\mathbf{y} = f(\mathbf{x}, \mathbf{n})$. The image denoising process $g(\cdot)$ then seeks an estimation of the clean image only from the observed noisy image *i.e.* $\hat{\mathbf{x}} = g(\mathbf{y})$. Depending on the physical causes of image noise, a noise model $f(\cdot)$ can be of different forms. For example, the widely used additive white Gaussian noise (AWGN) model can be depicted as follows:

$$f : \mathbf{y} = \mathbf{x} + \mathbf{n} \quad \text{and} \quad \forall l,\ n_l \sim \mathcal{N}(0, \sigma^2) \tag{1}$$

where each observed image pixel y_l is the superposition of a clean image pixel x_l and a noisy pixel n_l following i.i.d. zero-mean Gaussian noise with an unknown variance σ^2. Other commonly used noise model includes the salt-and-pepper noise [6], shot noise (Poisson distribution) [7], quantization noise (uniform distribution) [8], etc.

Image denoising methods are often classified into two groups [9]: (1) spatial domain (SD) image denoising and (2) transform domain (TD) image denoising. SD image denoising is a category of methods that often denoise an image by exploring and using spatial correlations across an

* Corresponding author at: Department of Electrical and Computer Engineering, Tufts University, 161 College Ave., Medford, MA 02155, USA. Tel.: +1 617 678 3008.
E-mail address: ywu03@ece.tufts.edu (Y. Wu).

0165-1684/$ - see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.sigpro.2014.01.007

image. It includes the conventional moving average filter (MAF) [10], the classic Gaussian filter (GF) [10], the well-known bilateral filter (BF) [11], the popular non-local means (NLM) [12,13], and many others [14–18]. One common feature of many SD methods is that each of them can be considered as a special case within the kernel estimation framework

$$\widehat{x}_l = \frac{\sum_{k \in \mathbb{S}_l} K(l,k,y_l,y_k) y_k}{\sum_{k \in \mathbb{S}_l} K(l,k,y_l,y_k)} \tag{2}$$

where $K(\cdot)$ denotes a kernel function and $\mathbb{S}_l$ is a specified searching window centered at the lth pixel, *e.g.* $\mathbb{S}_l = \{k \mid |k_r - l_r| \le 1 \text{ and } |k_c - l_c| \le 1\}$ denotes a 3×3 window. It is the kernel $K(\cdot)$ that differentiates the various SD methods. To see this, the kernel functions with respect to the GS, BF, NLM and the probabilistic NLM (PNLM) [17] are

$$K_{\mathrm{GF}}(l,k,y_l,y_k) = \exp\left(-\frac{\| l-k \|^2 y_k}{h} \right) \tag{3}$$

$$K_{\mathrm{BF}}(l,k,y_l,y_k) = \exp\left(-\frac{\| l-k \|^2}{h_s} - \frac{(y_l - y_k)^2}{h_i} \right) \tag{4}$$

$$K_{\mathrm{NLM}}(l,k,y_l,y_k) = \exp\left(-\frac{\| y_{\mathbb{P}_l} - y_{\mathbb{P}_k} \|^2}{h_\sigma} \right) \tag{5}$$

$$K_{\mathrm{PNLM}}(l,k,y_l,y_k) = \chi^2_{\eta(l,k)}\left(\frac{\| y_{\mathbb{P}_l} - y_{\mathbb{P}_k} \|^2}{\gamma(l,k)} \right) \tag{6}$$

where $\| \cdot \|$ denotes the L_2 norm, $\mathbb{P}_l$ is a prescribed local patch centered at the lth pixel, and h, h_s, h_i, h_σ, $\eta(\cdot)$, $\gamma(\cdot)$ are all parameters or parametric functions specified in the methods. Many contributions have been made for the SD denoising in recent years. These mainly focus on four things: (1) finding better kernel functions [14,16,17,19–22], (2) robust parameter tuning [23,24], (3) iterative denoising [25–27] and (4) computational speed-up [15,28,29]. It is worthy to point out that the design idea of NLM is often adopted and extended in more advanced image denoising techniques [30], like K-clustering with singular value decomposition (K-SVD) [31], and more sophisticated single image super-resolution tasks like [32].

In contrast to SD methods, TD methods first transform an image from the spatial domain to an alternative domain, apply denoising in this domain and then transform the resulting image back to spatial domain to complete denoising. In this way, the original image denoising problem may be better represented in the transformed domain preserving image features. Depending on the basis functions of a selected domain, one may further group TD methods into (1) methods with orthogonal-basis and (2) methods with nonorthogonal-basis [9]. The former group uses orthogonal basis functions, like Fourier [33], wavelet [34,35], curvelet [36], ridgelet [37], etc. As a result, image noise corresponds to high frequency components with small coefficients in a transformed image, and can be differentiated from image edges that correspond to high frequency components with large coefficients and from image homogeneous regions that correspond to low frequency components. The latter group of non-orthogonal basis domain has attracted much attention recently, mainly because non-orthogonal basis functions are more adaptive to local image features than fixed orthogonal basis functions [9]. A recent trend of TD methods is to use redundant representations for each image patch (equivalent to decompose each image patch as a linear combination of several non-orthogonal basis functions), *e.g.* K-SVD [31] and the learned simultaneous sparse coding (LSSC) method [38].

It is worthy noting that both SD and TD image denoising have their own weaknesses [9]. For example, SD methods often tend to oversmooth edge pixels after denoising [9], while TD methods often turn i.i.d. noise in a spatial domain to dependent noise in a transform domain and make analysis more difficult. It is therefore not surprising to see that many recent methods use both SD and TD for image denoising. For example, the block matching 3D transform (BM3D) [33] method stacks similar image patches in spatial domain but applying wavelet shrinkage for denoising, and it achieves remarkable performance in both denoising quality and speed. The NLM with dimensionality reduction method [39] denoises an image by using a kernel function in a transformed PCA domain. For a more detailed literature review on image denoising, one may refer to [9,12].

Besides these contributions on specific image denoising methods, many efforts are also made to solve general image denoising problems. For example, the Monte-Carlo SURE blackbox [1] provides an effective solution to parameter selections for image denoising methods. NLM with shape-adaptive patches (NLM-SAPs) develops a method to linearly combine multiple denoising results of using different parameters into one single and better output [19]. In this paper, we are interested in the general shrinkage estimation problem for image denoising: how to construct a better denoised image by using a noisy image observation and its corresponding denoised image. The rest of the paper is organized as follows: Section 2 briefly reviews the SURE and the shrinkage problem; Section 3 derives the closed-form of optimal blockwise shrinkage, proposes the SURE-based pixel aggregations and implementation; Section 4 discusses our simulation results; and we conclude the paper in Section 5.

2. Background

2.1. The shrinkage estimation in spatial domain image denoising

The shrinkage problem seeks a *better* estimator by linearly combining an initial estimate and the raw data. In image denoising, this means a new estimate that is found as

$$\widehat{\mathbf{x}}' = (\mathbf{1} - \mathbf{q}) \circ \widehat{\mathbf{x}} + \mathbf{q} \circ \mathbf{y} \tag{7}$$

where $\circ$ denotes the elementwise product, $\mathbf{q} = \{q_l\}$ is the shrinkage coefficient matrix of the same size as the noisy image, and $\mathbf{1}$, $\widehat{\mathbf{x}}$, and $\mathbf{y}$ denote the all-one matrix, the initially denoised image, and the noisy image, respectively.

Though the shrinkage problem is often considered as a separated problem in the SD denoising, shrinkage is actually implicitly and deeply involved in the design of