Signal Processing 103 (2014) 60–68

Contents lists available at ScienceDirect

Signal Processing

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

A constrained optimization approach to combining multiple non-local means denoising estimates

B.H. Tracey [a,*], E.L. Miller [a], Y. Wu [b], P. Natarajan [b,1], J.P. Noonan [a]

[a] Department of Electrical and Computer Engineering, Tufts University, Medford, MA, 02155, United States
[b] Raytheon BBN, Cambridge, MA 02138, United States

ARTICLE INFO

Article history:
Received 9 July 2013
Received in revised form
11 November 2013
Accepted 16 December 2013
Available online 28 December 2013

Keywords:
Denoising
Non-local means
Optimization
ADMM
Total Variation

ABSTRACT

There is an ongoing need to develop image denoising approaches that suppress noise while maintaining edge information. The non-local means (NLM) algorithm, a widely used patch-based method, is a highly effective edge-preserving technique but is sensitive to parameter tuning. We use a variational approach to combine multiple NLM estimates, seeking a solution that balances positivity constraints and gradient penalties against Stein's Unbiased Risk Estimate (SURE). This method greatly reduces parameter sensitivity and improves denoising performance vs. other NLM variants.

1. Introduction

Non-local means denoising (NLM) [1,2] has attracted significant interest in recent years, largely due to its success in preserving edges in denoised images. NLM operates by averaging data from different regions in the image that have similar appearance, relying on the fact that natural images often contain repeated patterns. Multiple extensions of the algorithm have been proposed, focusing on computationally efficient implementation [3–5], robust parameter selection [6–8], alternative noise models [9–11], and many others. NLM and other methods exploiting patch concepts, in particular BM3D [12], give state-of-the-art denoising results.

Our work builds on previous work by De Ville and Kocher [7], in which a linear combination of multiple NLM estimates is found that minimizes Stein's Unbiased Risk Estimate (SURE) globally across the image. The authors demonstrated that this combination can improve performance and reduce the need for parameter tuning. The authors also noted that in some cases, particularly for low-texture images, their approach suffered from over-fitting. Here we address this problem by proposing a constrained optimization framework for finding a convex combination of multiple NLM estimates, in which we add Total Variation (TV) gradient penalties [13] as well as positivity constraints.

Insight can be gained by considering denoising as a diffusion process, in which the image values are evolved according to a partial differential equation. Applying isotropic diffusion (the heat equation) corresponds to convolving the image with a Gaussian, which suppresses noise but also blurs edges [14]. As a result, many approaches have been developed that preserve edges during denoising, such as anisotropic diffusion approaches [15,14], nonlinear diffusion approaches such as the ROF filter [13,16], and nonlocal methods such as the Yaroslavsky filter [17] and bilateral filters [18]. Patch-based approaches such as NLM or UNITA [19] are a further development of these

* Corresponding author.
E-mail address: btracey@eecs.tufts.edu (B.H. Tracey).

[1] P. Natarajan is currently at the Information Sciences Institute at USC.

0165-1684/$ - see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.sigpro.2013.12.021

methods. NLM has been shown to be equivalent to diffusion based on pixel values, rather than spatial location [20]. In general, these filters take the form (for a continuous image $I(y)$) of

$$u(x) = \frac{\int K(x,y)I(y)\,dy}{\int K(x,y)\,dy} \tag{1}$$

(see [21] for further discussion). In Section 2, we show that a linear expansion of NLM estimates gives a result which is similar in form to Eq. (1). We further demonstrate that the *constrained* linear expansion we propose here leads to a kernel with desirable properties (guaranteeing, for example, that a non-negative image remains non-negative after denoising).

The contributions of our work are as follows. First, we formulate in Section 2 an optimization problem that includes Total Variation terms and positivity constraints in order to find the optimal weight combination for multiple NLMs. The properties of the resulting filter are then discussed. Second, we develop an alternating direction method of multipliers (ADMM) approach to efficiently solve this problem. While a similar solution method was recently used to deblur cartoon-plus-texture images [22], we are not aware of previous work on applying ADMM to multi-patch denoising problems. Finally, we demonstrate in Section 3 that this approach gives improved denoising as compared to [7] and other NLM variants, and conclude in Section 4.

2. Methods

Denoising addresses the problem of recovering the true image values $\mathbf{u}$ given noisy observation $\mathbf{v}=\mathbf{u}+\mathbf{n}$, where $\mathbf{n}$ is additive noise. Consider a given pixel location j that is associated with a patch Δ. The NLM estimate $\hat{u}_{NLM}(j)$ is a weighted sum of values at other points k that are within some search neighborhood $N_{srch}(j)$:

$$\hat{u}_{NLM}(j) = \frac{1}{Z(j)} \sum_{k \in N_{srch}(j)} K_{NLM}(j,k|\lambda)v(k) \tag{2}$$

where $Z(j) = \sum_k K_{NLM}(j,k|\lambda)$, and the weights are given by [6]

$$K_{NLM}(j,k|\lambda) = \exp\left(- \frac{\sum_{\delta \in \Delta}(v(j+\delta)-v(k+\delta))^2}{2L_\Delta \lambda^2} \right) \tag{3}$$

(note [1] defined the denominator above as h^2). In Eq. (3), λ is a bandwidth parameter, while Δ represents a local patch of pixels surrounding j, containing L_Δ pixels; a patch of the same shape also surrounds k. Although a variety of patch shapes are possible [1,23], square patches centered on the points of interest are most common. The summation in (3) captures patch similarity from the summed, squared pixel-by-pixel difference between patches centered on j and k (here δ is the offset from the patch center). If similar patches can be found throughout the image, then ideally the neighborhood N_{srch} is taken to be the entire image, so the averaging process is fully *non-local*. More detail on NLM is found in [1,2].

Next, assume that we generate P different estimates of the denoised image $\hat{u}_p(j)$, calculated as discussed in Eqs. (2) and (3) but corresponding to different choices of denoising parameters. Like [7], we seek a linear combination of these estimates. Unlike [7], we seek to minimize a weighted sum of Stein's Unbiased Risk Estimate (SURE) and gradient penalties, subject to positivity constraints. The linear expansion of NLM estimates is given by

$$\hat{u}(j) = \sum_{p=1}^{P} w_p \hat{u}_p(j). \tag{4}$$

The anisotropic Total Variation is given by

$$TV(\hat{\mathbf{u}}) = \sum_l \left| \frac{\partial}{\partial x}\hat{u}(l) \right| + \left| \frac{\partial}{\partial y}\hat{u}(l) \right| \tag{5}$$

while the SURE term is given by [6]

$$\begin{aligned} J_{SURE} &= \frac{1}{2L}\sum_l (\hat{u}(l)-v(l))^2 - \sigma^2 + \frac{2\sigma^2}{L}\sum_l \frac{\partial \hat{u}(l)}{\partial v(l)} \\ &\equiv \frac{1}{2L}\sum_l (\hat{u}(l)-v(l))^2 - \sigma^2 + \frac{1}{L} div_{\mathbf{y}}\{\hat{\mathbf{u}}\} \end{aligned} \tag{6}$$

where L is the number of points in the image. The expansion in Eq. (4) is substituted into Eqs. (5) and (6). To simplify notation, we collect the P denoised image estimates into an $L \times P$ matrix $\mathbf{U}$. Similarly, we can collect the x and y gradients of the P images into $L \times P$ matrices $\mathbf{D}_x$ and $\mathbf{D}_y$, and can collect the P divergence fields into a matrix $\mathbf{DIV}$ (which includes the scaling $2\sigma^2$). We are using the fact that because the new $\hat{u}$ is a linear expansion of the component $\hat{u}_p$s, the new gradient and divergence are also linear expansions of the component gradients and divergences. This gives the following expression to be minimized:

$$J(\mathbf{w}) = \tfrac{1}{2}\|\mathbf{Uw}-\mathbf{v}\|_2^2 + \mathbf{DIVw} + \lambda_{TV}\|\mathbf{Dw}\|_1 \tag{7}$$

Here we have dropped the constant term σ^2, absorbed factors of $1/L$ into λ_{TV}, and defined $\mathbf{D} = [\mathbf{D}_x^T \mathbf{D}_y^T]^T$ (note that $\mathbf{D}$ is not simply the difference operator, but rather the difference operator applied to each of the denoising estimates). In addition, we impose the constraints

$$\sum_p w_p = 1, \quad \mathbf{w} > 0 \tag{8}$$

2.1. Comments

We can gain a better understanding of the proposed filter by inserting Eq. (2) into Eq. (4) and noting that our final denoised estimate can be written as

$$\hat{u}(j) = \sum_{k \in N_{srch}(j)} \left[\sum_{p=1}^{P} \frac{w_p}{Z_p(j)} K_{NLM}(j,k|\lambda_p) \right] v(k) \tag{9}$$

$$\hat{u}(j) \equiv \sum_{k \in N_{srch}(j)} K(i,j)v(k) \tag{10}$$

where $K(i,j)$ is a new denoising kernel, formed as the weighted sum of the individual NLM kernels. This result holds for both the unconstrained linear expansion method of [7] and for our proposed method. Thus linear expansion creates a new kernel, which has the flexibility to take on shapes not possible within the family of possible NLM kernels defined by Eq. (3). This added flexibility helps to