Nonlocal Similarity Image Filtering

Yifei Lou[1], Paolo Favaro[2], Stefano Soatto[1], and Andrea Bertozzi[1]

[1] University of California Los Angeles, USA
[2] Joint Research Institute on Image and Signal processing
Heriot-Watt University, Edinburgh, UK

Abstract. We exploit the recurrence of structures at different locations, orientations and scales in an image to perform denoising. While previous methods based on "nonlocal filtering" identify corresponding patches only up to translations, we consider more general similarity transformations. Due to the additional computational burden, we break the problem down into two steps: First, we extract similarity invariant descriptors at each pixel location; second, we search for similar patches by matching descriptors. The descriptors used are inspired by scale-invariant feature transform (SIFT), whereas the similarity search is solved via the minimization of a cost function adapted from local denoising methods. Our method compares favorably with existing denoising algorithms as tested on several datasets.

1 Introduction

Image "denoising" refers to a series of inference tasks whereby the effects of various nuisance factors in the image formation process are removed or mitigated. Like all inference tasks, denoising hinges on an underlying model – implicit or explicit – where nuisance factors are processes that affect the data, but whose inference is not directly of interest. The generic term "noise" then refers loosely to all unmodeled phenomena, so illumination could be treated as noise in one application, or signal in another.

In Computer Vision we are used to more explicit models of the underlying scene, and even simple ones such as "cartoon models" [1, 2], occlusion "layers" [3], multiresolution and scale-space processes [4] have had ramifications in image processing. However, one could argue that the image formation process is unduly complex, and modeling it explicitly just to remove noise or increase the resolution is overkill. This philosophy is at the core of so-called "exemplar-based methods," [5]: Instead of explicitly modeling the image-formation process, one can just "sample" its effects and manipulate the samples to yield the desired inference result. In the simpler forward problem, that of image synthesis, this philosophy has yielded so-called "procedural methods" in computer graphics, that have been rather successful especially in synthesizing complex textures (see [6] and references therein).

The basic model underlying nonlocal denoising [7] is that an image is generated by patches that are translated in different locations of the image, downsampled, and corrupted by additive noise. To perform denoising, then, one can search for all patches similar to the given one *up to translation*, then transform them, and then perform standard

image processing operations. This model can be generalized, whereby the transformation undergone by patches is not just a translation, but any homeomorphic transformation of the image domain. The more complex the transformation, the more powerful the model, the more costly the inference process is. Which begs the question of what is the right trade off between modeling power (fidelity) and computational costs (complexity)[3].

We have therefore conducted empirical studies of various procedural, or exemplar-based, models and their effects on image denoising, and have converged to a *similarity model* as the desirable tradeoff. Furthermore, projective transformations can be approximated locally by similarity transformations, for which efficient detectors and descriptors are available [8]. In this manuscript we propose a denoising algorithm that operates on patches with scale and rigid invariance, hence extending recent results on nonlocal image filtering.

In our method, we consider equivalent all patches that are similarity transformations of a given pattern. This generates equivalence classes of patches, and one can define a metric and probabilistic structure on the equivalence classes, so that patches can be compared. This can be done as part of the matching process (by "searching" the equivalence class for all possible transformations of a given patch) or by defining "canonical representatives" of the equivalence class. This way, one can generate a "descriptor" for every equivalence class, and then endow the space of descriptors with a distance, without solving an optimization or search problem at every step. We choose this second option, where we compute – at each pixel – a similarity-invariant descriptor, similar to the scale-invariant feature transform (SIFT) [8].

In the next section we start from the general formulation of nonlocal filtering, then extend it to the similarity model. We then briefly review SIFT, and how it relates to our goals, and finally propose our algorithm in Sect. 3. We then present empirical results in support of our approach.

1.1 Related Work

A variety of methods are available for image denoising, such as PDE-based methods [9–11], wavelet-based approaches [12, 13] and statistical filters [14, 15]. Among all these methods, the most related one to ours is the nonlocal means filter [7]. It recently emerged as a generalization of the Yaroslavsky filter [16], but also taps on "exemplar-based" methods in texture synthesis [17] and super-resolution [5], as well as on "procedural methods" in computer graphics [6, 18]. Buades *et al.* transposed the idea to image denoising. Its advantage is to exploit similar patches in the same image, without an explicit model of the image formation process. The approach is taken one step further in [19], where similarity is computed hierarchically and efficiently. Another accelerating method is proposed by Mahmoudi and Sapiro [20] via eliminating unrelated neighborhoods from the weighted average. There are several other methods based on the idea of nonlocal means filter [7]. For example, Kervrann, *et al.* [21] improve it by using an adaptive window size; [22, 23] formalize a variational nonlocal framework motivated

[3] As George E.P. Box said, "all models are wrong, some are useful," hence this question cannot be settled by means of analysis.

from graph theory [24]; Chatterjee, *et al.* [25] generalize nonlocal means to high-order kernel regression. Nonetheless, all the methods interpret the concept of "similarity" only up to translation, while we extend it to a more general similarity transformation, *i.e.*, scaling and rotation.

2 Nonlocal Similarity Image Filtering

In the next subsection we review the method proposed by Buades *et al.*, then SIFT, and then propose our approach to image denoising and super-resolution.

2.1 Nonlocal Means Filtering

The key idea of the nonlocal means filter is that a given noisy image $f : \Omega \subset \mathbb{R}^2 \mapsto \mathbb{R}$ is filtered by

$$u(\mathbf{x}) = \int w_f(\mathbf{x}, \mathbf{y}) f(\mathbf{y}) d\mathbf{y} \ , \tag{1}$$

where $u : \Omega \mapsto \mathbb{R}$ is the denoised image and $w_f : \Omega \times \Omega \mapsto \mathbb{R}^+$ is a normalized weight function written as

$$w_f(\mathbf{x}, \mathbf{y}) \doteq \frac{e^{-\frac{d_f^2(\mathbf{x},\mathbf{y})}{h^2}}}{\int e^{-\frac{d_f^2(\mathbf{x},\mathbf{y})}{h^2}} d\mathbf{y}} \ , \qquad \text{for} \qquad d_f^2(\mathbf{x}, \mathbf{y}) \doteq \|f_{\mathbf{x}} - f_{\mathbf{y}}\|_{G_\sigma}^2 \tag{2}$$

is the L_2-norm of the difference of $f_{\mathbf{x}}$ (*i.e.*, f centered in $\mathbf{x}$) and $f_{\mathbf{y}}$ (*i.e.*, f centered in $\mathbf{y}$), weighted against a Gaussian window G_σ with standard deviation σ. The map $d_f(\mathbf{x}, \mathbf{y})$ measures how similar two patches of f centered in $\mathbf{x}$ and $\mathbf{y}$ are. If two patches are similar, then the corresponding weight $w_f(\mathbf{x}, \mathbf{y})$ will be high. Vice versa, if the patches are dissimilar, the weight $w_f(\mathbf{x}, \mathbf{y})$ will be small (but positive). While the parameter σ defines the dimension of the patch where we measure the similarity of two patches, the parameter h regulates how strict or relaxed we are in considering patches similar. The final result of the nonlocal means filter is that several (similar) patches are used to reconstruct another one.

Notice that the similarity of patches in d_f is defined up to translation. In other words, we can only match patches that are simply in different locations, but otherwise unchanged – with the same orientation and scale. This motivates us to consider the larger class of similarity measures that discounts scale and rotation changes, *i.e.*, a similarity-invariant measure. In theory, defining this measure is just a matter of introducing two more integrals in d_f and an inverse similarity-transformation in eq. (1) to align the patches being averaged. In practice, however, because this similarity has to be computed multiple times for each patch, this introduces considerable computational burden that makes the ensuing algorithm all but impractical. One way to address this problem is to find a function that estimates a rotation and a scale at each patch with respect to a common reference system, so that each patch can be transformed into a "canonical" patch. Once this is done, one can apply the original nonlocal means filter. In the next section we will describe one such function.

2.2 Scale-Invariant Feature Descriptors

The idea of determining when two regions are similar up to a similarity transformation has been widely explored in the past to solve several tasks including object recognition, structure from motion, wide-baseline matching, and motion tracking [26, 8, 27–29]. In this paper, however, we will exploit the same idea of matching similarity-invariant regions for the purpose of image denoising.

One of the most successful methodologies to match regions up to a similarity transformation is the Scale Invariant Feature Transform (SIFT) [8]. The main steps in computing SIFT are

- *Scale-space extrema detection:* Scale is identified at each point by searching for extrema in the scale-space of the image via a difference-of-Gaussian convolution.
- *Keypoint localization:* Keypoints are selected based on the stability of fitting a 3-D quadratic function (obtained via Taylor expansion of the scale-space of the image).
- *Orientation assignment:* A rotation with respect to a canonical reference frame is computed based on local image gradients.
- *Keypoint descriptor:* A vector composed of local image gradients is built, so that it is not sensitive to similarity transformations and, to some extent, changes in illumination.

More details on how each step is implemented in practice can be found in [8].

Notice that there is a fundamental difference in how SIFT is commonly used and how it is employed in our algorithm. In our case the *Keypoint localization* step is not implemented as we are interested in computing a SIFT descriptor and in obtaining some consistent estimate of scale and orientation at each pixel. From now on, therefore, we will define our SIFT filter to estimate scale and orientation as $\rho(\mathbf{x}) : \Omega \mapsto [0, \infty)$ and $\theta(\mathbf{x}) : \Omega \mapsto [0, \pi]$ respectively.

3 Nonlocal Similarity-Invariant Filtering

In this section we define our nonlocal similarity mean filter, which is a combination of nonlocal mean filtering and SIFT. The nonlocal means can be regarded as one step of a fixed point iteration to solve the optimality conditions of the following functional [30]

$$J(u) \doteq \int \left(u(\mathbf{x}) - u(\mathbf{y})\right)^2 w_f(\mathbf{x}, \mathbf{y}) d\mathbf{x} d\mathbf{y} \ . \tag{3}$$

where w_f is defined in eq. (2). We reformulate the weight function to be similarity-invariant,

$$w_f(\mathbf{x}, \mathbf{y}) = \mathrm{e}^{-||P(\mathbf{x}) - P(\mathbf{y})||^2 / h^2} \ , \tag{4}$$

where $P(\mathbf{x})$ is the canonical form of the patch center at $\mathbf{x}$ and h is a parameter as in the Non-local means.

In Figure 1, we illustrate step-by-step how we align the patch to its canonical form: for each pixel $\mathbf{x}$,

1. Take a patch with size $\sim 10\rho(\mathbf{x})$ around the pixel;

2. Rotate this patch with the angle $\theta(\mathbf{x})$;
3. Extract the middle part with size $\sim 7\rho(\mathbf{x})$ for the boundary problem after rotating;
4. Down-sample to a uniform size (the smallest size among all patches) and save as $P(\mathbf{x})$.

In this way we can extract more meaningful patches than in previous nonlocal means methods, as shown in Figure 2. Since we assume additive Gaussian white noise, noise is invariant to rotation and scaling if the image is considered to be a continuous function. When aligning the patches, there are interpolation errors, but they are negligible two-pixels away from the center, if bilinear interpolation is used. We mitigate scale errors by using only patches that are larger, and therefore at higher resolution, than the reference patch.

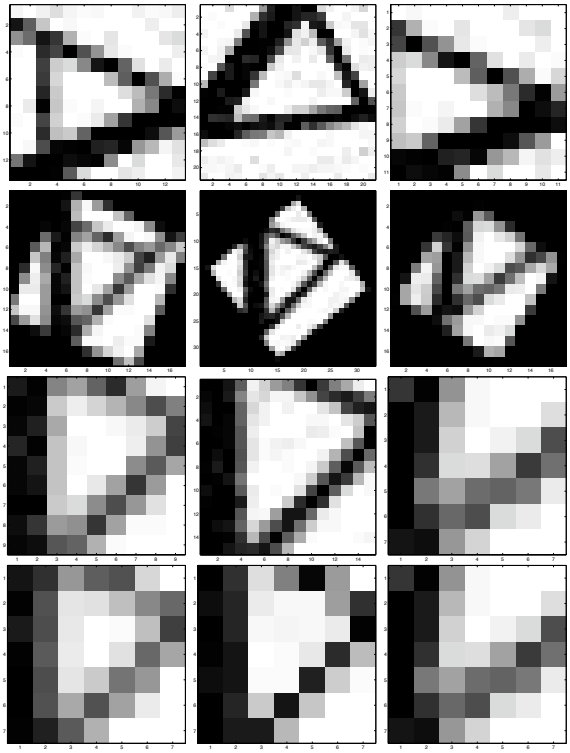

Fig. 1. Procedure to align patches. Three patches are selected to illustrate the alignment, as shown in the first row. From top to bottom: (1) noisy patches whose size corresponds to the scale of its center; (2) rotate the patch with the angle assigned by SIFT; (3) crop the black boundary due to the rotation; (4) down-sample to a uniform size patch 7×7.

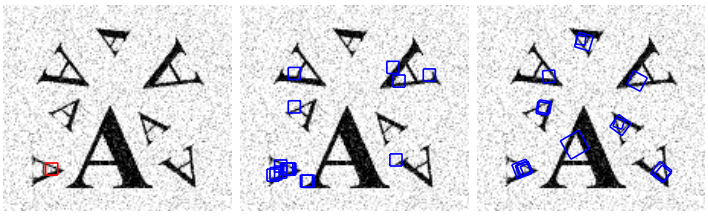

Fig. 2. Fifteen most similar patches to the target one (red square on the left) are selected (middle) and aligned via similarity (right). On the right, the pose of the patch corresponds to the scale and orientation of its center as obtained by SIFT.

3.1 Denoising

We add a convex fidelity term to the nonlocal functional J in (3), yielding a denoising model

$$\hat{u} = \arg\min_u J(u) + \frac{\lambda}{2}\int (f(\mathbf{x}) - u(\mathbf{x}))^2\, d\mathbf{x}\,. \tag{5}$$

To minimize the energy (5), we apply the gradient descent flow:

$$u_t(\mathbf{x}) = -\int (u(\mathbf{x}) - u(\mathbf{y}))w(\mathbf{x},\mathbf{y})d\mathbf{y} + \lambda(f(\mathbf{x}) - u(\mathbf{x}))\,. \tag{6}$$

Notice that the above equation is linear in $u(\mathbf{x})$, so an implicit time difference scheme is applied in order to make the iterations more stable.

$$\frac{u^{n+1}(\mathbf{x}) - u^n(\mathbf{x})}{dt} = -\int (u^{n+1}(\mathbf{x}) - u^n(\mathbf{y}))w(\mathbf{x},\mathbf{y})d\mathbf{y} + \lambda(f(\mathbf{x}) - u^{n+1}(\mathbf{x}))\,. \tag{7}$$

We can also extend this model to color image denoising in which the input image $\boldsymbol{f} := (f^R, f^G, f^B)$ is a three-channel signal. In a similar way, we can compute the weight $w_{\boldsymbol{f}}(\mathbf{x},\mathbf{y})$ using high-dimensional patches so that the weight is the same for all color channels. We express the total energy as follows,

$$\hat{u} = \arg\min_u \sum_{j=R,G,B} \int \left(u^j(\mathbf{x}) - u^j(\mathbf{y})\right)^2 w_{\boldsymbol{f}}(\mathbf{x},\mathbf{y})d\mathbf{x}d\mathbf{y} + \frac{\lambda}{2}\int \left(f^j(\mathbf{x}) - u^j(\mathbf{x})\right)^2 d\mathbf{x}\,. \tag{8}$$

Notice that we can perform color image denoising by treating the three color channels independently.

4 Experiments

We compare the performance of our method to that of the PDE-based method [11], the wavelet-based method [12] and the original nonlocal means [7]. Other denoising methods are examined and compared in [7].

We present the nonlocal similarity filtering on two synthetic images which are corrupted by additive Gaussian noise with standard deviation $\sigma = 20$ (Fig. 3) and $\sigma = 40$ (Fig. 4) respectively. For each method, the residual image $f - u$ is shown. Both the PDE-based method [11] and the wavelet based method [12] fail to preserve structures as they are left in the residual image. The traditional nonlocal method fails to denoise the central part in Fig 3 since these regions in the residual image are almost flat.

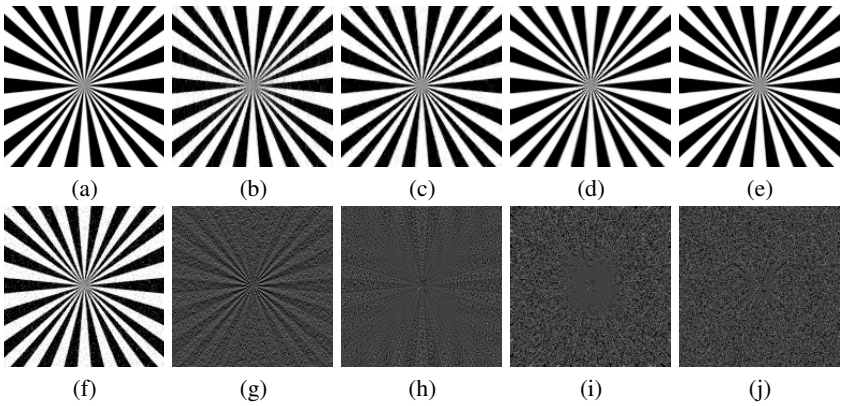

Fig. 3. Experiment with Gaussian noise: $\sigma = 20$**.** *Top: (a) original image, (b) PDE-based [11], (c) Wavelet-based [12], (d) nonlocal means and (e) NL similarity. Bottom: (f) noisy input f, (g)-(j) show the residual of each method (b)-(e) respectively. The flat regions in (h) show that the central part has not been denoised.*

An example of color image denoising is presented in Fig. 5. In computing the weight, the L_2 distance between 3-D patches (RGB) is used. As for denoising, we treat the three color bands independently. The results are presented in Fig. 5, which shows that our approach works better for stripes, while it is comparable to the original method for the stars.

We compare the quantitative evaluation of various denoising methods. Table 1 lists the root mean square (RMS) error of each method:

$$RMS(u) = \sqrt{\int_{\Omega} (I(x) - u(x))^2 dx / |\Omega|}\,, \tag{9}$$

where I is the original image and u is the reconstruction of the method, both of which are defined on the image domain Ω.

5 Conclusions

We extended the nonlocal means filtering by a more general similarity measurement. In particular, we applied SIFT to estimate a rotation and a scale at each patch so that it can

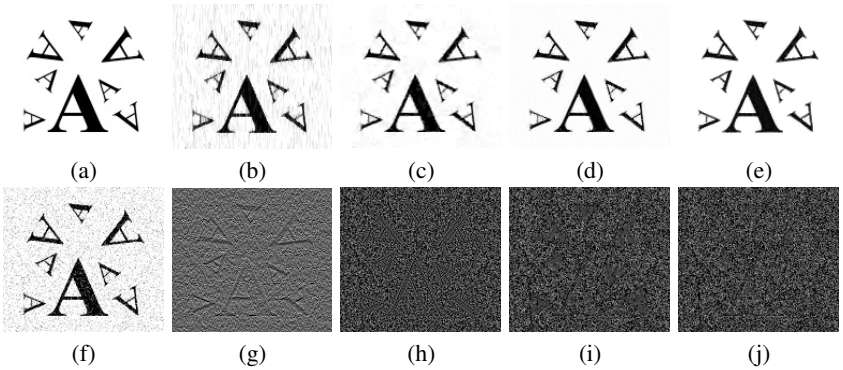

Fig. 4. Experiment with Gaussian noise: $\sigma = 40$**.** *Top: (a) original image, (b) PDE-based [11], (c) Wavelet-based [12], (d) nonlocal means and (e) NL similarity. Bottom: (f) noisy input f, (g)-(j) residual of each method (b)-(e) respectively. The flat regions in (h) show that the serifs of the A characters have not been captured.*

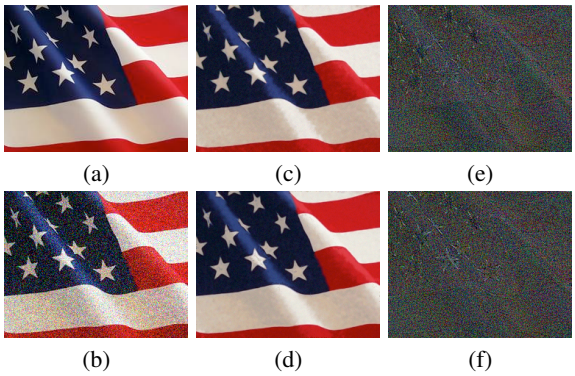

Fig. 5. Color image denoising with Gaussian noise with $\sigma = 30$. From left to right, top to bottom: (a) original image, (b) noisy input f, (c) nonlocal means u_1, (d) nonlocal Similarity u_2, (e) NL method noise $f - u_1$ and (f) NL similarity method noise $f - u_2$. The stripes tend to be restored better in (f) than in (e).

Table 1. Root mean square error for the input images and different denoising methods.

RMS	Input	PDE-based [11]	wavelet-based[12]	NL means[7]	NL similarity
testpat	20.00	14.65	12.56	9.80	**6.74**
letter	40.00	20.05	13.57	12.42	**11.66**
flag (color)	30.00	N/A	N/A	10.43	**9.11**

be transformed to a canonical form. Then we construct the weight based on the canonical form so that we could exploit more similar patches to help denoising. Experiments demonstrate that the proposed nonlocal similarity filtering outperforms the previous methods especially when applied to the restoration of patterns that are replicated with different scale and/or rotation.

Acknowledgement This work is supported by NSF grant ECS-0622245, ONR grants N000140810363 / N000140810414, and EPSRC grant EP/F023073/1(P). Dr. Favaro acknowledges the support of his position within the Joint Research Institute in Image and Signal Processing at Heriot-Watt University which is part of the Edinburgh Research Partnership in Engineering and Mathematics (ERPem).

References

1. Mumford, D., Shah, J.: Optimal approximation by piecewise smooth optimal approximation by piecewise smooth functions and associated variational problems. Commun. Pure Appl. Math **42**(577-685) (1989)
2. W. Yin, D.G., Osher, S.: Image cartoon-texture decomposition and feature selection using the total variation regularized l1 functional. In: Variational, Geometric, and Level Set Methods in Computer Vision. Volume 3752. (2005) 73–84
3. Wang, J.Y.A., Adelson, E.H.: Representing moving images with layers. IEEE Transactions on Image Processing Special Issue: Image Sequence Compression **3**(5) (1994) 625–638
4. Lindeberg, T.: Scale-space theory: A basic tool for analysing sturctures at different scale. Journal of Applied Statistics **21**(2) (1994) 224–270
5. Freeman, W.T., Jones, T.R., Pasztor, E.C.: Example-based super-resolution. IEEE Computer Graphics and Applications (2002)
6. Wei, L.Y., Levoy, M.: Fast texture synthesis using tree-structured vector quantization. In: Proc. of conf. on Computer Graphics and interactive techniques. (2000) 479–488
7. Buades, A., Coll, B., Morel, J.M.: On image denoising methods. SIAM Multiscale Modeling and Simulation **4**(2) (2005) 490–530
8. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision **60**(2) (2004) 91–110
9. Perona, P., Malik, J.: Scale space and edge detection using anisotropic diffusion. In: IEEE Trans. Pattern Anal. Math. Intell. Volume 12. (1990) 629–639
10. Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D **60** (1992) 259–268
11. Chambolle, A.: An algorithm for total variation minimization and applications. In: Journal of Mathematical Imaging Vision. Volume 20. (2004) 89–97
12. Portilla, J., Strela, V., Wainwright, M.J., Simoncelli, E.P.: Image denoising using scale mixtures of gaussians in the wavelet domain. In: IEEE Trans. Image Process. Volume 12. (2003) 1338–1351
13. Mignotte, M.: Image denoising by averaging of piecewise constant simulations of image partitions. In: IEEE Trans. Image Processing. Volume 16. (2007) 523–533
14. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: Proc. International Conference Computer Vision. (1998) 839–846
15. Awate, S.P., Whitaker, R.T.: Unsupervised, information-theoretic, adaptive image filtering for image restoration. IEEE Trans. Pattern Anal. Mach. Intell. **28**(3) (2006) 364–376

16. Yaroslavsky, L.P.: Digital Picture Processing, an Introduction. Springer-Verlag, Berlin (1985)
17. Efros, A.A., Leung, T.K.: Texture synthesis by non-parameteric sampling. In: ICCV. Volume 2. (1999) 1033–1038
18. Criminisi, A., Perez, P., Toyama, K.: Region filling and object removal by exemplar-based inpainting. IEEE Transactions on Image Processing (2004) 1200–1212
19. Brox, T., Cremers, D.: Iterated nonlocal means for texture restoration. In Sgallari, F., Murli, A., Paragios, N., eds.: Proc. International Conference on Scale Space and Variational Methods in Computer Vision. Volume 4485 of LNCS., Ischia, Italy, Springer (May 2007) 13–24
20. Mahmoudi, M., Sapiro, G.: Fast image and video denoising via nonlocal means of similiar neighborhoods. In: IEEE Signal Processing Letter. Volume 12. (2005) 839–842
21. Kervrann, C., Boulanger, J.: Optimal spatial adaptatio for patch-based image denoising. In: IEEE Trans. Image Processing. Volume 15. (2006) 2866–2878
22. Gilboa, G., Osher, S.: Nonlocal operators with applications to image processing. Technical report, UCLA CAM report 07-23 (2007)
23. Kindermann, S., Osher, S., Jones, P.: Deblurring and denoising of images by nonlocal functionals. SIAM Multiscale Modeling and Simulation **4**(4) (2005) 1091–1115
24. Zhou, D., Scholkopf, B.: A regularization framework for learning from graph data. In: ICML Workshop on Stat. Relational Learning and Its Connections to Other Fields. (2004)
25. Chatterjee, P., Milanfar, P.: A generalization of non-local means via kernel regression. In: Proc. of SPIE Conf. on Computational Imaging. (2008)
26. Schmid, C., Mohr, R.: Local greyvalue invariants for image retrieval. Pattern Analysis and Machine Intelligence (1997)
27. Tuytelaars, T., Gool, L.V.: Wide baseline stereo based on local, affinely invariant regions. In: british Machine vision conference. (2000) 412–422
28. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robusts wide baseline stereo from maximally stable extremal regions. In: british Machine vision conference. (2002) 384–393
29. Vedaldi, A., Soatto, S.: Local features, all grown up. In: Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). (2006) 1753–1760
30. Gilboa, G., Osher, S.: Nonlocal linear image regularization and supervised segmentation. SIAM Multiscale Modeling and Simulation **6**(2) (2007) 595–630