CONFIDENCE-BASED DENOISING RELYING ON A TRANSFORMATION-INVARIANT, ROBUST PATCH SIMILARITY
Exploring Ways to Improve Patch Synchronous Summation

Cesario V. Angelino, Eric Debreuve and Michel Barlaud

Laboratoire I3S, Université de Nice-Sophia Antipolis/CNRS, 2000 route des lucioles, 06903 Sophia Antipolis, France

Keywords: Denoising, Image patch, Patch denoising confidence, Robust patch similarity.

Abstract: Patch-based denoising techniques have proved to be very efficient. Indeed, they account for the correlations that exist among the patches of natural images, even when degraded by noise. In this context, we propose a denoising method which tries to minimize over-smoothing of textured areas (an effect observed with NL-means), to avoid staircase effects in monotonically varying areas (an effect observed with BM3D), and to limit spurious patterns in areas with virtually no variations. The first step of the proposed method is to perform patch denoising by averaging similar patches of the noisy image (the equivalent in the space of patches to synchronous summation for temporal signals). From there, our contribution is twofold. *(a)* We proposed to combine the resulting overlapping denoised patches accounting for an assessed patch denoising confidence. *(b)* Since a crucial aspect is the definition of a similarity between two patches, we defined a patch similarity invariant to some transformations and robust to noise thanks to a polynomial patch approximation, instead of a classical weighted $\mathcal{L}^2$-similarity. The experimental results show an arguably better visual quality of images denoised using the proposed method compared to NL-means and BM3D. In terms of PSNR, the results are significantly above NL-means and comparable to BM3D.

1 INTRODUCTION

Some studies on the distribution of patches forming natural images (Carlsson et al., 2008; Huang and Mumford, 1999; Lee et al., 2003) have motivated recent patch-based processing methods, for example, for image and video denoising (Buades et al., 2005; Awate and Whitaker, 2006; Dabov et al., 2007; Boulanger et al., 2007) or inpainting (Criminisi et al., 2004). Indeed, these studies showed that there exist correlations among patches composing images. As a consequence, the probability is high that patches similar to a given image patch be encountered in the image itself. It is reasonable to think that these correlations remain in the presence of (some reasonable amount of) noise or after similar image degradations, offering the opportunity to reduce noise or to recover missing information in a patch. As a matter of fact, the nonlocal means algorithm (NL-means) (Buades et al., 2005) and BM3D (3D transform-domain collaborative filtering) (Dabov et al., 2007) proved to be successful in image denoising. Given an image patch in a noisy image, the main step of NL-means is to perform patch denoising by averaging some

patches of the noisy image similar to the given one. This can be viewed as the equivalent in the space of patches to synchronous summation for temporal signals. Then, the denoised patch is projected onto the pixel space by simply extracting the central denoised pixel. Some works have established a variational justification relying on an information-theoretic energy to such an approach (Awate and Whitaker, 2006; Angelino et al., 2008). Nevertheless, some developments rely on other projections which make better use of the denoised patches (Dabov et al., 2007; Salmon and Strozecki, 2010). In this context, our contribution is twofold. *(a)* The definition of a similarity between two patches is crucial for the denoising efficiency. Instead of a classical weighted $\mathcal{L}^2$-norm (Buades et al., 2005) (possibly computed after a dimension reduction such as a PCA (Principal Component Analysis)) (Tasdizen, 2008) or an SVD (Singular Value Decomposition) (Orchard et al., 2008)), we defined a patch similarity invariant to some transformations (rotation and symmetries) and robust to noise thanks to a polynomial approximation of the patches. Note that the idea of transforming the patches before similarity computation has been proposed for a reduced set of trans-

formations (Sun et al., 2009). *(b)* Moreover, we propose to project, or otherwise stated, combine the overlapping denoised patches accounting for an assessed patch denoising confidence. The notion of denoising confidence for patch combination has also been proposed at the patch level (Dabov et al., 2007) and at the pixel level (Salmon and Strozecki, 2010). The confidence proposed here is patchwise and it is adapted to the chosen patch denoising procedure. These contributions aim at minimizing over-smoothing of textured areas (an effect observed with NL-means), avoiding staircase effects in monotonically varying areas (an effect observed with BM3D), and limiting spurious patterns in areas with virtually no variations (a phenomenon observed in (Brox and Cremers, 2007) when iterating NL-means, as proposed in (Awate and Whitaker, 2006)).

The remainder of the paper is organized as follows: Section 2 reminds some results about a variational interpretation of NL-means based on an energy defined as a conditional entropy; Section 3 describes the proposed confidence-based denoising procedure; Section 4 describes the proposed robust patch similarity; Section 5 presents some experimental results for a synthetic noise on standard benchmark images and a digital camera noise on a professional benchmark photography; and Section 6 provides a summary and discusses some points of the proposed method.

2 DENOISING USING PATCH SIMILARITIES

The inverse problem of image restoration can be formulated as a minimization problem. As mentioned in Section 1, natural images exhibit correlation among the patches which compose them. This correlation should be accounted for in deriving a restoration procedure.

Let us consider the conditional entropy h of patches, i.e. the uncertainty on the color of a pixel when its neighborhood is known. Let us model an image as a random field X. Let T be the set of pixels of the image and let C_t be a neighborhood of pixel $t \in T$. The random vector $Y(t) = \{X(u)\}_{u \in C_t}$ contains the set of intensities or colors of the neighbors C_t of pixel t. The random vector $Z(t) = (X(t), Y(t))$ denotes the corresponding patch, i.e. the pixel color combined with those of its neighborhood. The recovered image ideally satisfies

$$X(t)^* = \arg\min_X h(X|\tilde{Y} = \tilde{Y}(t)) \qquad (1)$$

for all $t \in T$, where $\tilde{Y}$ is the observed noisy neighborhood.

Classically, we propose to use a gradient descent procedure to solve (1). As a consequence, we need to determine the derivative of the conditional entropy of the color of a pixel knowing its neighborhood.

Entropy, whether conditional or not, can be approximated by the Ahmad-Lin estimator (Ahmad and Lin, 1976)

$$h(X|\tilde{Y} = y_i) = -\frac{1}{|T|} \sum_{t_j \in T} \log p(x_j|y_i) \qquad (2)$$

where the color x_j is encountered at pixel t_j, y_i is the set of (noisy) colors in the neighborhood of t_j, and

$$p(s|y_i) = \frac{1}{|T_{y_i}|} \sum_{t_m \in T_{y_i}} K(s - x_m) \qquad (3)$$

is the kernel-based estimate of the underlying conditional probability density function (PDF), with T_{y_i} the set of pixels which have the same neighborhood y_i, K a symmetric kernel, and x_m the color encountered at pixel t_m. The derivative of (2) is

$$\frac{\partial h(X|\tilde{Y} = y_i)}{\partial x_i} = -\frac{1}{|T|} \frac{\nabla p(z_i)}{p(z_i)} \cdot \frac{\partial z_i}{\partial x_i} \qquad (4)$$

where $p(Z)$ is the joint PDF of the high dimensional random vector Z. The solution to (1) can be computed by gradient descent

$$\begin{cases} X^{(0)} = X_0 \\ x_i^{(n+1)} = x_i^{(n)} + \beta \dfrac{\nabla p\big(z_i^{(n)}\big)}{p\big(z_i^{(n)}\big)} \cdot \dfrac{\partial z_i^{(n)}}{\partial x_i} \cdot \end{cases} \qquad (5)$$

3 CONFIDENCE-AWARE DENOISING

3.1 Patch Synchronous Summation

Unlike denoising methods such as (Awate and Whitaker, 2006; Buades et al., 2005) which actually implement pixel-based iterative procedures relying on patches, as done by (5), we propose to convert (5) into a patch-based iterative procedure. This means that in the update equation of (5) we discard the projection term $\partial z_i^{(n)} / \partial x_i$, which projects the *patch space* (Z) gradient vector into the *pixel space* (X) derivative. Thus, we get

$$\begin{cases} X^{(0)} = X_0 \\ z_i^{(n+1)} = z_i^{(n)} + \beta \dfrac{\nabla p\big(z_i^{(n)}\big)}{p\big(z_i^{(n)}\big)} \cdot \end{cases} \qquad (6)$$

In practice, we have noticed that performing only one iteration is sufficient. The normalized derivative

in (6) can be approximated by a mean shift (Fukunaga and Hostetler, 1975; Comaniciu and Meer, 2002) on the high dimensional joint PDF of Z. In the k-th nearest neighbor (kNN) framework, it can be expressed as (Angelino et al., 2008)

$$\frac{\nabla p(z_i)}{p(z_i)} = \frac{d+2}{\rho_k^2(z_i)} \frac{1}{k} \sum_{z_j \in kNN(z_i)} (z_j - z_i) \qquad (7)$$

where d is the dimension of Z, $kNN(z)$ denotes the set of the k closest patches of z and $\rho_k(z_i)$ is the kNN patch distance.

By setting $\beta = \rho_k^2(z_i)/(d+2)$ and using (7), it is clear that one iteration of the procedure (6) is equivalent to

$$z_i^* = \frac{1}{\sum_{z_j \in kNN(z_i)} w_j} \sum_{z_j \in kNN(z_i)} w_j z_j \qquad (8)$$

where the weights w_j account for the fact that, among the patches of $kNN(z_i)$, patches farther away might not correspond as well to noisy versions of z_i as do closer patches (they may also be slightly structurally different).

Similarly to NL-means (Buades et al., 2005),

$$w_j = \exp(-|z_i - z_j|^2/\sigma_w^2), \qquad (9)$$

σ_w being a parameter.

3.2 Patch Combination

Denoised patches obtained in (8) overlap each other. In consequence, there is some redundancy in the denoising process. Indeed, for a given pixel x_i, we obtain several estimators, one for each patch to which x_i belongs (if the patch size is $N_p \times N_p$, we have N_p^2 estimators.).

Clearly, among all the patches containing x_i, some will lead to an accurate denoising at x_i, some might not while leading to an accurate denoising in other pixels.

Since the noise is assumed to be uncorrelated, the synchronous summation (8) reduces the amount of noise (as characterized by its variance) by a factor of $1/c_i$, where

$$c_i = \frac{\left(\sum_{z_j \in kNN(z_i)} w_j\right)^2}{\sum_{z_j \in kNN(z_i)} w_j^2}. \qquad (10)$$

We call c_i the patch denoising confidence.

In summary, for each patch z_i of the noisy image, we are able to compute with the confidence c_i (see (10)) a denoised patch z_i^* (see (8)). These patches are then combined (or aggregated) according to their confidence term. Among a plethora of methods for

combining estimators, we used, as in (Dabov et al., 2007), a linear combination of denoised patches.

The denoised patches z_i^* are then aggregated as follows: starting from an aggregation image of zeros and a confidence map of zeros, a denoised patch z_i^* is added, after weighting by c_i, to the aggregation image at its original location, and a constant patch equal to c_i is added to the confidence map at the same location. After dealing with all the patches, the denoised image is defined as the pointwise division between the aggregation image and the confidence map.

For a given pixel i, the aggregation writes

$$x_i = \frac{1}{\sum_{q=1}^{N_p^2} c_{i,q}} \sum_{q=1}^{N_p^2} c_{i,q} x_{i,q}^*, \qquad (11)$$

where $c_{i,q}$ is the confidence of patch q among the N_p^2 patches which contain pixel i.

4 ROBUST PATCH SIMILARITY

The key idea of patch-based denoising algorithm is image self similarity. On an image, one can find several small areas or patches that are *similar* to each other. Suppose that we have several noisy realizations of the same patch. Assuming an additive zero mean noise, we can denoise such a patch by taking the average of the noisy patches. The (dis)similarity function between two patches plays an important role in choosing the best *candidates* (in our case the k nearest neighbors) for the denoising task. The most popular choice is the $\mathcal{L}^2$-distance between the patches. However, when the noise is high, such a distance becomes unreliable. Moreover, two patches can be similar only up to some transformation such as rotation. Therefore, we propose to define a patch similarity invariant to some patch transformations and applied to polynomial approximations of the patches. This represents an improvement over (Angelino et al., 2010).

4.1 Robustness to Noise

Before trying to look for similar patches, the patches are modified in order to ensure some level of robustness with respect to noise. Given a noisy patch considered as a surface, a polynomial surface approximation is computed with independent degrees in both directions. Each degree can range from 1 to 10. The best approximation has been defined as the one producing a zero-mean noise between the noisy patch and its approximation. The approximations of all the patches are then used, possibly after geometrical

transformations (see Section 4.2), to look for self similarities.

4.2 Transformation Invariance

Standard patch-based denoising algorithms compute the $\mathcal{L}^2$-distance blockwise. This means that the similar patch search is performed allowing only for patch translation. However, more similar patches can often be found if allowing for rotation or symmetry of the patches (Sun et al., 2009). Building on this idea, we propose an extended set of patch transformations:

- symmetries with respect to the horizontal and vertical axes and the diagonals;

- rotations by $\pi/2$, π, and $3\pi/2$;

- rotations between $-\alpha$ and α with a step of $d\alpha$.

Note that when comparing two patches, the rotations $\pi/2$, π, and $3\pi/2$ are often not the transformation leading to the best matching (contrary to the rotations between $-\alpha$ and α with α around $\pi/6$). However, they are computed cheaply and exactly (no interpolation required).

In summary, let S be the set of patches in the image and $\mathcal{T}$ the set of transformations (symmetries and rotations), then the set in which we look for similar patches is the augmented set $S \cup \mathcal{T}(S)$.

5 EXPERIMENTAL RESULTS

5.1 Effect of the Transformations: A Toy Example

A white Gaussian noise was added to a synthetic image of a disk. The PSNR of the noisy image was 14.9. This image was denoised without and with transformation invariant patch similarity (see Section 4.2). The positive effect of using patch transformations can be seen on Figure 1. Objectively, the denoised image obtained with transformations has a PSNR of 29.1 while the image obtained without transformations has a PSNR of 27.7.

5.2 Synthetic Noise

We tested our Patch Denoising Confidence with Robust Similarity (PDC-RS) algorithm on several images and compared to BM3D and NL means algorithms. Each image was corrupted with four levels of AWGN noise. We set BM3D in normal profile with standard parameters provided by authors. We chose for both PDC-RS and NL means a patch radius of 7
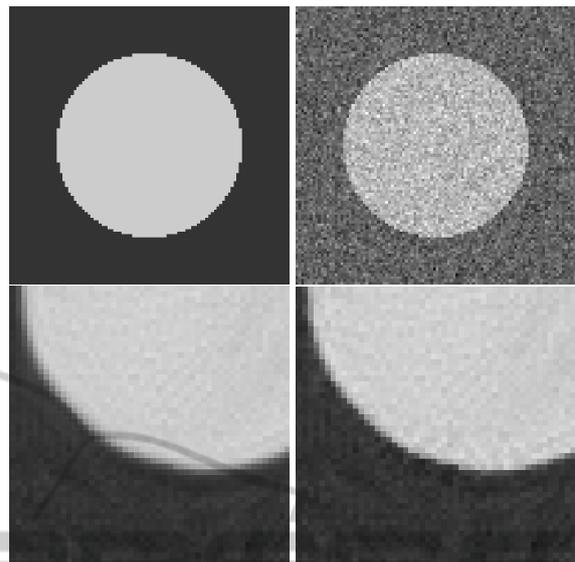

Figure 1: Effect of the transformations: A toy example. In lexicographic order: Original, Noisy, Denoised with no transformations, Denoised with transformations.
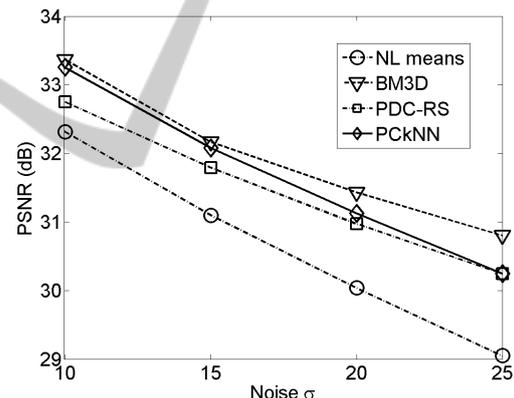

Figure 2: PSNR plot for the image Elaine.

and a search radius of 15. Furthermore, as suggested from authors, we set $h = \sigma$, h being the filtering parameter of NL means and σ the noise standard deviation. Figure 2 shows the Peak Signal to Nose Ratio (PSNR) as function of the noise standard deviation for the three compared algorithms. PDC-RS outperforms NL-means and is close in terms of PSNR to both our previous method PCkNN (Angelino et al., 2010) and BM3D. However, as shown in Figure 3, the image obtained with PDC-RS looks more natural than the ones obtained with the other algorithms. In particular, the residual noise is quite natural and does not exhibit spurious patterns, thus leading to a denoised image with a very natural appearance. NL-means is clearly oversmoothed and BM3D presents many flattened regions in smoothly varying areas (see Fig. 4),

Figure 3: Visual comparison on the image Elaine for $\sigma = 25$. In lexicographic order: Noisy, Original, BM3D, PCkNN, NL-means, and PDC-RS.

giving a somewhat unnatural, cartoon effect to the denoised image. This is the major drawback of this algorithm for high noise levels that might be due to the thresholding in the wavelet domain. To illustrate the cartoon effect, Figure 4 shows a close-up on the image Elaine of Figure 3 and the corresponding isolevel lines. The orientation and density of these lines provide an indication on the direction and the norm of the gray level gradient. PDC-RS preserved very well the original isolevel line configuration while BM3D created a "patchwork" of flattened regions.

Figure 5 clearly shows how the denoised images well preserve textures and edges, avoiding any cartoon effect for all the noise levels.

5.3 Digital Camera Noise

Digital camera noise is very well approximated with an independent Gaussian stochastic process. However, contrary to the classical additive white Gaussian noise model with constant variance (used in Section 5.2 for comparison purposes), the variance of the digital camera noise can be modeled as an affine func-
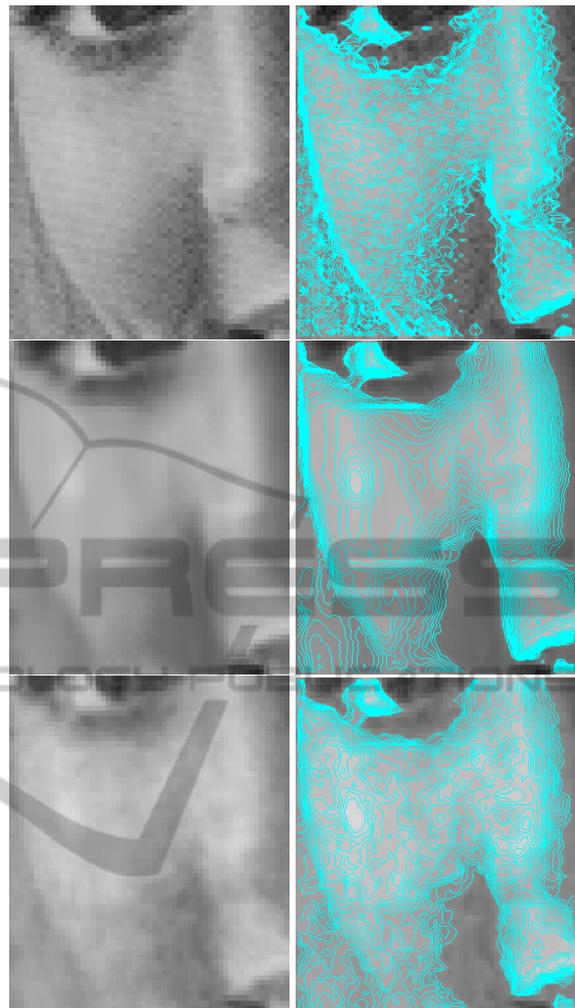
Figure 4: A close-up on the image Elaine of Figure 3. From top to bottom: Original, BM3D, and PDC-RS. Left: image alone; right: isolevel lines superimposed on the image.

tion of the signal intensity x: $\sigma^2(x) = \gamma x + \delta$. Moreover, most digital cameras acquire images using a single image sensor overlaid with a color filter array (CFA) which produces images with a single red, blue, or green component per pixel. The process of computing the missing 2 color components at each pixel is called demosaicing. Unfortunately, demosaicing introduces correlation among neighborhood pixels and hence correlates the noise. The result is a "structured noise" which is not Gaussian and not independent anymore. Removing this noise is a harder task since algorithms usually rely on a hypothesis of independence. Thus, we applied it to each color channel of the raw image, i.e., before demosaicing. Furthermore, we adapted our denoising algorithm to the varying variance model by making σ_w^2 in (9) equal to $\gamma \bar{x} + \delta$, $\bar{x}$ being the average patch intensity. After denoising the

Figure 5: Denoising of the image Elaine using the proposed method, PDC-RS. The image was corrupted by an additive white Gaussian noise with successive standard deviations σ of 10, 15, 20, and 25. In lexicographic order: Noisy ($\sigma = 25$), Original, denoised image for $\sigma = 10$, 15, 20, and 25.
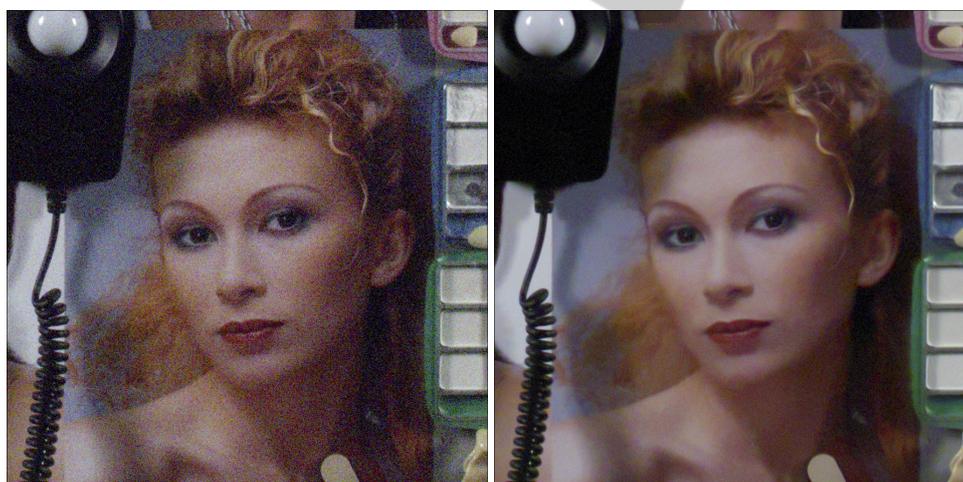
Figure 6: Professional benchmark image (Courtesy of DxO Labs). RAW image (left) and denoised image (right).

image has been demosaiced. Figure 6 shows a result on a 1024x1024-crop out of a 4416x3312-DxO Labs benchmark image taken with a Canon G10 at ISO 1600. Again, the denoised image has a very natural appearance without flattened regions.

6 DISCUSSION

This paper presented PDC-RS a fully patch-based denoising algorithm. Denoised patches are provided with a confidence term which measures the quality of denoising. The denoised patches are then combined together based on their respective confidence terms in

order to reconstruct the denoised image. Moreover, we introduced a new patch similarity measure invariant to transformation and robust to noise. Experimental results show the good quality of PDC-RS denoised images w.r.t. state of the art denoising techniques especially for high level of noise. In particular, denoised images have a very natural appearance. Image details are well preserved and there is no cartoon effect even in high levels of noise. PDC-RS has also been tested on a professional DxO Labs benchmark image giving very good and promising result. As mentioned in Section 5.3, digital cameras image denoising is a challenging task since noise variance is function of the signal.

Concerning the future works, let us mention the measure used to search for similar patches in the image. Currently, this search is done in the $\mathcal{L}^2$ sense. Certainly, a better notion of visual similarity such as the SSIM (Wang et al., 2004) could be used instead.

ACKNOWLEDGEMENTS

REFERENCES

Ahmad, I. A. and Lin, P. (1976). A nonparametric estimation of the entropy for absolutely continuous distributions. *IEEE Transactions On Information Theory*, pages 372–375.

Angelino, C. V., Debreuve, E., and Barlaud, M. (2008). Image restoration using a knn-variant of the mean-shift. In *IEEE ICIP*, San Diego, CA, USA.

Angelino, C. V., Debreuve, E., and Barlaud, M. (2010). Patch confidence k-nearest neighbors denoising. In *IEEE ICIP*, Hong Kong.

Awate, S. P. and Whitaker, R. T. (2006). Unsupervised, information-theoretic, adaptive image filtering for image restoration. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(3):364–376.

Boulanger, J., Kervrann, C., and Bouthemy, P. (2007). Space-time adaptation for patch-based image sequence restoration. *IEEE Trans. On Pattern Analysis And Machine Intell.*, 29:1096–1102.

Brox, T. and Cremers, D. (2007). Iterated nonlocal means for texture restoration. In *Scale Space and Variational Methods in Computer Vision, First International Conference, SSVM 2007, Ischia, Italy, May 30 - June 2, 2007, Proceedings*, pages 13–24.

Buades, A., Coll, B., and Morel, J. (2005). A non-local algorithm for image denoising. In *IEEE CVPR*, pages 60–65, Washington, DC, USA.

Carlsson, G., Ishkhanov, T., de Silva, V., and Zomorodian, A. (2008). On the local behavior of spaces of natural images. *Int. J. Comput. Vision*, 76:1–12.

Comaniciu, D. and Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24:603–619.

Criminisi, A., Pérez, P., and Toyama, K. (2004). Region filling and object removal by exemplar-based image inpainting. *IEEE Trans. on Image Process.*, 13:1200–1212.

Dabov, K., Foi, A., K., V., and Egiazarian, K. (2007). Image denoising by sparse 3d transform-domain collaborative filtering. *IEEE Trans. Image Process.*, 16:2007.

Fukunaga, K. and Hostetler, L. D. (1975). The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Trans. On Inf. Theory*, 21:32–40.

Huang, J. and Mumford, D. (1999). Statistics of natural images and models. In *IEEE CVPR*, pages 541–547.

Lee, A. B., Pedersen, K. S., and Mumford, D. (2003). The nonlinear statistics of high-contrast patches in natural images. *Int. J. Comput. Vision*, 54(1-3):83–103.

Orchard, J., Ebrahimi, M., and Wong, A. (2008). Efficient nonlocal-means denoising using the svd. In *IEEE International Conference on Image Processing*, pages 1732–1735, San Diego (CA), USA.

Salmon, J. and Strozecki, Y. (2010). From patches to pixels in non-local methods: Weighted-average reprojection. In *IEEE International Conference on Image Processing*, Hong Kong, China.

Sun, W., Peng, Y., and Hwang, W. (2009). Modified similarity metric for non-local means algorithm. *Electronics Letters*, 45(25):1307–1309.

Tasdizen, T. (2008). Principal components for non-local means image denoising. In *IEEE International Conference on Image Processing*, pages 1728–1731, San Diego (CA), USA. Electr. & Comput. Eng. Dept., Univ. of Utah, UT.

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