IMAGE DENOISING BY NON-LOCAL AVERAGING

Antoni Buades *, *Bartomeu Coll* †

Dpt. Matemàtiques i Informàtica, UIB
Ctra. Valldemossa Km. 7.5,
07122 Palma de Mallorca, Spain

Jean Michel Morel ‡

CMLA, ENS Cachan
61, Av du Président Wilson
94235 Cachan, France

ABSTRACT

In this work, we present and analyze an image denoising method, the NL-means algorithm, based on a non local averaging of all pixels in the image. We also introduce the concept of *method noise*, that is, the difference between the original (always slightly noisy) digital image and its denoised version. Finally we present some experiences comparing the NL-means results with some classical denoising methods.

1. INTRODUCTION

The need for efficient image restoration methods has grown with the massive production of photographs, often taken in poor conditions or with deficient cameras or acquisition systems. Due to the nature of the light, the amount of photons arriving to the camera fluctuates around the true value. These perturbations are called noise. Then, we can write

$$v(i) = u(i) + n(i), \qquad (1)$$

where $v(i)$ is the observed value, $u(i)$ would be the "true" value and $n(i)$ is the noise perturbation at a pixel i. The best simple way to model the effect of noise on a digital image is to add gaussian white noise. In that case, $n(i)$ are i.i.d. gaussian values with zero mean and variance σ^2.

Several methods have been proposed to remove the noise and recover the true image u. Even though they may be very different in tools it must be emphasized that all of them share the same basic remark : denoising is achieved by averaging. This averaging may be performed locally: the Gaussian smoothing model (Gabor [1]), the anisotropic filtering (Perona-Malik [2], Alvarez et al. [3]) and the neighborhood filtering (Yaroslavsky [4]), by the calculus of variations: the Total Variation minimization (Rudin-Osher-Fatemi [5]), or in the frequency domain: the empirical Wiener filters (Yaroslavsky [4]) and wavelet thresholding methods (Coiffman-Donoho [6]).

* This author is supported by a fellowship of the Govern Illes Balears.
† Partially supported by CICYT project, ref. TIC2002-02172.
‡ Partially supported by ONR grant N00014-97-1-0839, CNES, DGA.

Formally we define a denoising method D_h as a decomposition

$$v = D_h v + n(D_h, v),$$

where v is the noisy image and h is a filtering parameter which usually depends on the standard deviation of the noise. Ideally, $D_h v$ is more smooth than v and $n(D_h, v)$ looks like the realization of a white noise. For most denoising methods, $n(D_h, v)$ contains texture and details and therefore these are removed from the estimate $D_h v$.

In order to better understand this removal texture and details, we introduce the *method noise*. That is, the difference between the original (always slightly noisy) image u and its denoised version.

Definition 1 (Method noise) *Let u be an image and D_h a denoising operator depending on a filtering parameter h. Then, we define the* method noise *as the image difference*

$$u - D_h u.$$

This difference measures the degree of preservation of the true image during the denoising process. In order to preserve the original features and fine structure of the image, this method noise should be as small as possible.

We also propose and analyze the *NL-means* algorithm, which is defined by the simple formula, for $x \in \Omega$,

$$NL[u](x) = \frac{1}{C(x)} \int_\Omega e^{-\frac{f(x,y)}{h^2}}\, u(y)\, dy,$$

where $f(x,y) = \int_{\mathbb{R}^2} G_a(t)|u(x+t) - u(y+t)|^2 dt$, G_a is a Gaussian kernel, $C(x) = \int_\Omega e^{-\frac{f(x,z)}{h^2}} dz$ is a normalizing constant and h acts as a filtering parameter. This amounts to say that $NL[u](x)$, the denoised value at x, is a mean of the values of all pixels whose gaussian neighborhood looks like the neighborhood of x. The main difference of the NL-means algorithm with respect to local filters or frequency domain filters is the systematic use of all possible self-predictions the image can provide.

In section 2 we give a discrete and effective procedure for the computation of the NL-means algorithm. In section

3 we show the consistency of the algorithm under stationarity conditions. In section 4 we compute the method noise for the Gaussian filtering and we show how the NL-means algorithm picks up the efficient parts of the afore mentioned methods. Finally, in section 5 we show some experiments comparing the performance of the NL-means algorithm and classical denoising algorithms.

2. THE NL-MEANS ALGORITHM

Given a discrete noisy image $v = \{v(i) \mid i \in I\}$, for a pixel i, the estimated value $NL[v](i)$ is computed as a weighted average of all the pixels in the image,

$$NL[v](i) = \sum_{j \in I} w(i,j)v(j),$$

where the family of weights $\{w(i,j)\}_j$ depend on the similarity between the pixels i and j, and satisfy the usual conditions $0 \leq w(i,j) \leq 1$ and $\sum_j w(i,j) = 1$.

The similarity between two pixels i and j depends on the similarity of the intensity gray level vectors $v(\mathcal{N}_i)$ and $v(\mathcal{N}_j)$, where $\mathcal{N}_k$ denotes a square neighborhood of fixed size and centered at a pixel k. The pixels with a similar grey level neighborhood to $v(\mathcal{N}_i)$ have larger weights in the average, see Fig. 1. The similarity between gray level neighborhoods is computed as a gaussian weighted Euclidean difference, $\|v(\mathcal{N}_i) - v(\mathcal{N}_j)\|_{2,a}^2$, where $a > 0$ is the standard deviation of the Gaussian kernel. The application of the Euclidean distance to the noisy neighborhoods raises the following equality

$$E||v(\mathcal{N}_i) - v(\mathcal{N}_j)||_{2,a}^2 = ||u(\mathcal{N}_i) - u(\mathcal{N}_j)||_{2,a}^2 + 2\sigma^2.$$

This equality shows the robustness of the algorithm since in expectation the Euclidean distance conserves the order of similarity between pixels.

The weights are then defined as,

$$w(i,j) = \frac{1}{Z(i)} e^{-\frac{||v(\mathcal{N}_i) - v(\mathcal{N}_j)||_{2,a}^2}{h^2}},$$

where $Z(i)$ is the normalizing constant

$$Z(i) = \sum_j e^{-\frac{||v(\mathcal{N}_i) - v(\mathcal{N}_i)||_{2,a}^2}{h^2}}$$

and the parameter h acts as a degree of filtering. It controls the decay of the exponential function and therefore the decay of the weights as a function of the Euclidean distances.

3. NL-MEANS CONSISTENCY

We show that under stationarity assumptions, for a pixel i, the NL-means algorithm converges to the conditional expectation of i once observed a neighborhood of it. In this case, the stationarity conditions amount to say that as the size of the image grows we can find many similar patches for all the details of the image.

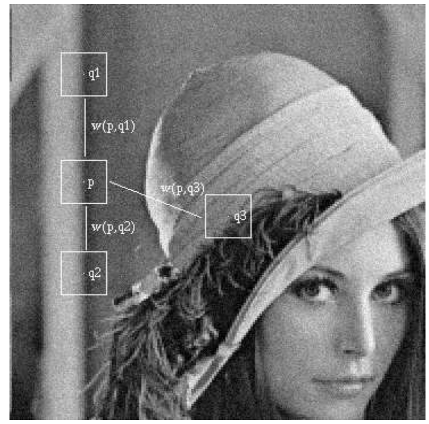

Fig. 1. Scheme of NL-means strategy. Similar pixel neighborhoods give a large weight, w(p,q1) and w(p,q2), while much different neighborhoods give a small weight w(p,q3).

Let V be a random field and suppose that the noisy image v is a realization of V. Let Z denote the sequence of random variables $Z_i = \{Y_i, X_i\}$ where $Y_i = V(i)$ is real valued and $X_i = V(\mathcal{N}_i \backslash \{i\})$ is $\mathbb{R}^p$ valued. The NL-means is an estimator of the conditional expectation $r(i) = E[Y_i \mid X_i = v(\mathcal{N}_i \backslash \{i\})]$.

Theorem 1 (Conditional expectation theorem) *Let $Z = \{V(i), V(\mathcal{N}_i \backslash \{i\})\}$ for $i = 1, 2, \ldots$ be a strictly stationary and mixing process. Let NL_n denote the NL-means algorithm applied to the sequence $Z_n = \{V(i), V(\mathcal{N}_i \backslash \{i\})\}_{i=1}^n$. For $j \in \{1, \ldots, n\}$*

$$|NL_n(j) - r(j)| \to 0 \quad a.s$$

The proof of the previous result can be found in a more general framework in [7].

In the case that the additive white noise model (1) is assumed, the next result shows that the conditional expectation is the function of $V(\mathcal{N}_i \backslash \{i\})$ that minimizes the mean square error with the true image u.

Theorem 2 *Let V, U, N be random fields on I such that $V = U + N$, where N is a signal independent white noise. Then, the following statements are hold.*

(i) $E[V(i) \mid X_i = x] = E[U(i) \mid X_i = x]$ *for all* $i \in I$ *and* $x \in \mathbb{R}^p$.

(ii) The expected random variable $E[U(i) \mid V(\mathcal{N}_i \backslash \{i\})]$ *is the function of* $V(\mathcal{N}_i \backslash \{i\})$ *that minimizes the mean square error*

$$\min_g E[U(i) - g(V(\mathcal{N}_i \backslash \{i\}))]^2$$

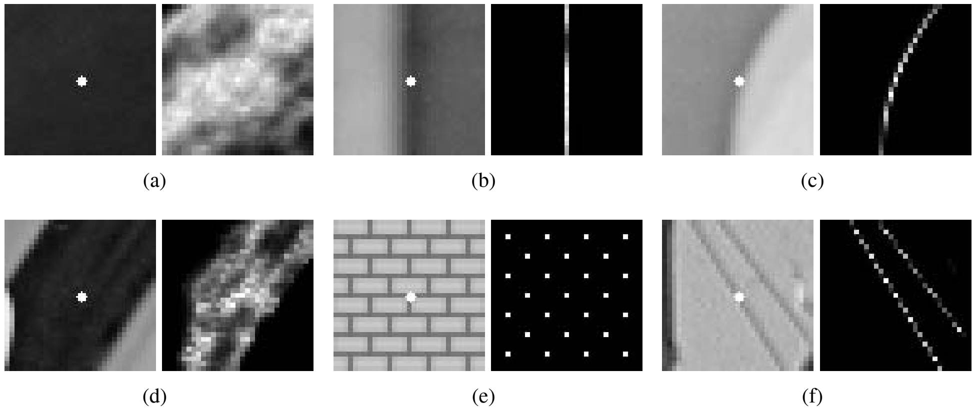

Fig. 2. Display of the NL-means weight distribution used to estimate the central pixel of every image. The weights go from 1(white) to zero(black). The algorithm acts as a convolution filter in float zones (a), as an anisotropic filter in straight edges (b) and as a neighborhood filter in flat neighborhoods (d). On curved edges, the weights favor pixels belonging to the same contour (c) and favors pixels with similar configurations, even they are far away (e) and (f).

Fig. 3. Image method noise. From left to right: Original image, Gaussian filtering and NL-means.

4. THE METHOD NOISE

The method noise results from the application of the denoising algorithm to the non noisy image. As we want not only to reduce the noise but to restore the geometrical and textural features, the application of the algorithm should not alter the original image. Following the above idea we have proposed to analyze the difference $u - D_h u$.

The method noise represents the loss of information from the original image and therefore should be as small as possible. The method noise can be computed for all previously mentioned denoising methods. For simplicity reasons, we only discuss the Gaussian filtering. In this case, the noisy image v is convolved with a Gaussian kernel G_h of standard deviation h.

Theorem 3 (Gabor 1960) *The method noise of the convolution with a gaussian kernel G_h is*

$$u - G_h * u = -\frac{h^2}{2}\Delta u + o(h^2),$$

for h is small enough.

As a practical consequence, edges and texture are not well restored since the Laplacian at these points cannot be small. This method noise helps us to understand the performance and limitations Gaussian filtering.

The denoising methods should adapt to the image in order to preserve it. In this sense, we propose to apply the NL-means algorithm. We visually show how the NL-means algorithm chooses a weighting configuration adapted to the local and non local geometry of the image, see Fig. 2.

Fig. 3 shows a visual experiment of the method noise. The method noise displays the structures which are not well preserved by denoising algorithms and which will be degraded by a further denoising process. As a consequence, this method noise should look as similar as possible to a white noise.

5. EXPERIMENTATION AND DISCUSSION

We display some denoising experiences comparing the NL-means algorithm with classical denoising methods. All experiments have been simulated by adding a gaussian white noise of standard deviation σ to the true image.

For computational purposes, we can restrict the search of similar windows in a larger "search window" of size $S \times S$ pixels. In all the experimentation we have fixed a search window of 21×21 pixels and a similarity square neighborhood $\mathcal{N}_i$ of 7×7 pixels. If N^2 is the number of pixels of the image, then the final complexity of the algorithm is about $49 \times 441 \times N^2$.

The 7×7 similarity window has shown to be large enough to be robust to noise and small enough to take care of details and fine structure. The filtering parameter h has been fixed to $10 * \sigma$. Due to the fast decay of the exponential kernel, large Euclidean distances lead to nearly zero weights acting as an automatic threshold, see Fig. 2.

Due to the nature of the algorithm, the most favorable

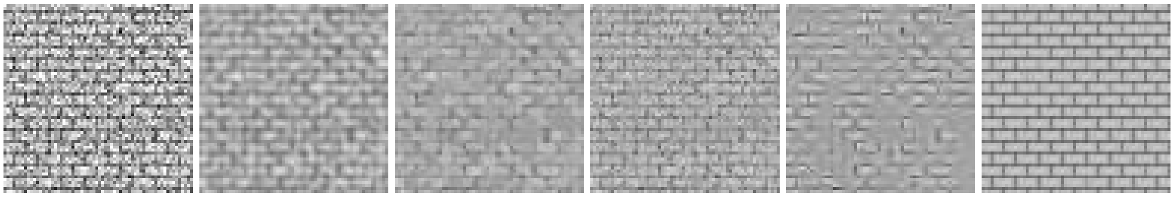

Fig. 4. Denoising experience on a periodic image. From left to right: noisy image (standard deviation 35), Gauss filtering, Total variation, Neighborhood filter, translation invariant wavelet thresholding and NL-means algorithm. It seems that a non-local algorithm is necessary to reconstruct the periodic and fine structure of the wall pattern.

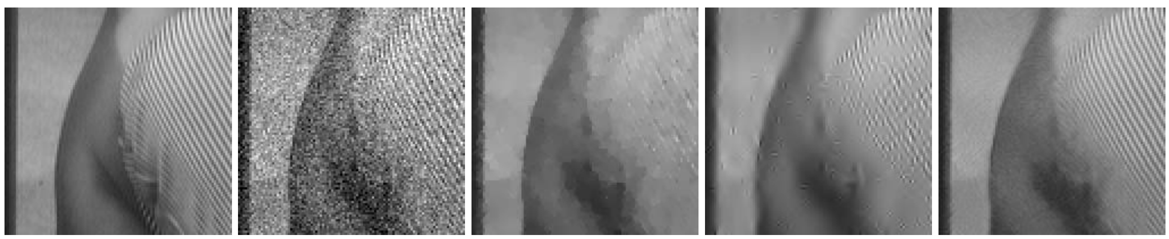

Fig. 5. Denoising experience on a natural image. From left to right: Original image, noisy image (standard deviation 25), Total Variation minimization, translation invariant thresholding and NL-means algorithm. The NL-means is able to reconstruct the fine periodic pattern on the right side of the image while the other methods filter it as noise.

case for the NL-means is the texture or periodic case. In this situation, for every pixel i, we can find a large set of samples with a very similar configuration. Fig. 4 compares the performance of the NL-means with classical local and frequency filters.

Natural images also have enough redundancy to be restored by NL-means. Flat zones present a huge number of similar configurations lying inside the same object, see Fig. 2 (a). Straight or curved edges have a complete line of pixels with similar configurations, see Fig. 2 (b) and (c). In addition, natural images allow us to find many similar configurations in far away pixels, as Fig. 2 (f) shows. Fig. 5 shows how the NL-means algorithm is able to distinguish between white noise and periodic patterns (right side of the image). The algorithm is able to reconstruct this fine periodic pattern while the other methods filter it as noise. Table 1 displays the mean square error of the two previous experiences for the different denoising methods.

Image	GF	AF	TV	NF	EWF	WT	NL
Barbara	220	216	186	176	111	135	72
Wall	580	660	721	598	325	712	59

Table 1. Mean square error table. We compare the Gaussian filtering (GF), the anisotropic filtering (AF), the total variation minimization (TV), the neighborhood filtering (NF), the empirical Wiener filter (EWF), the translation invariant wavelet thresholding (WT) and the NL-means algorithm(NL). A smaller mean square error indicates that the estimate is closer to the original image.

6. REFERENCES

[1] M. Lindenbaum, M. Fischer, and A. M. Bruckstein, "On gabor contribution to image-enhancement," *Pattern Recognition*, vol. 27, pp. 1–8, 1994.

[2] P. Perona and J. Malik, "Scale space and edge detection using anisotropic diffusion," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 12, pp. 629–639, 1990.

[3] L.Alvarez, P-L. Lions, and J-M. Morel, "Image selective smoothing and edge detection by nonlinear diffusion (ii)," *SIAM Journal of numerical analysis*, vol. 29, pp. 845–866, 1992.

[4] L.P. Yaroslavsky and M. Eden, *Fundamentals of Digital Optics*, Birkhauser, 1996.

[5] L. Rudin and S. Osher, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259–268, 1992.

[6] R.R. Coifman and D. Donoho, "Translation-invariant de-noising," *Waveletes and Statistics*, pp. 125–150, 1995.

[7] G.G. Roussas, "Nonparametric regression estimation under mixing conditions," *Stochastic processes and their applications*, vol. 36, pp. 107–116, 1990.