Analysis of Non Local Image Denoising Methods

Álvaro Pardo

Department of Electrical Engineering, Faculty of Engineering and Technologies, Universidad Católica del Uruguay
apardo@ucu.edu.uy

Abstract. Image denoising is probably one of the most studied problems in the image processing community. Recently a new paradigm on non local denoising was introduced. The Non Local Means method proposed by Buades, Morel and Coll attracted the attention of other researches who proposed improvements and modifications to their proposal. In this work we analyze those methods trying to understand their properties while connecting them to segmentation based on spectral graph properties. We also propose some improvements to automatically estimate the parameters used on these methods.

1 Introduction

Image deonising is probably one of the most studied problems in image processing. The main goal of denoising is to remove undesired components from the image. These undesired components, usually defined as noise, can be of different nature: random noise introduced at acquisition time, noise introduced during transmission, noise due to degradation such in films, etc. In this work we assume that the observed image, x, is the result of adding a random noise component n to the original noiseless image z. Therefore, the relationship between those images at pixel i becomes: $x_i = z_i + n_i$.

The problem of image denoising then is to estimate z while preserving its features such as edges and texture. There is usually a tradeoff between noise reduction and feature preservation. Since image features usually involve high frequencies linear low pass filters usually produce poor results regarding feature preservation. For this reason several non linear or locally adapted methods have been developed. As examples we mention median filters, anisotropic diffusion and wavelet thresholding. More recently non local methods attracted the attention of the image processing community. Starting from the pioneering work of Efros and Leung [7] several non local methods have been introduced for image denoising. In [5] Buades, Morel and Coll presented the Non Local Means (NLM) denoising method. The underlying idea of this method is to estimate, z_i, using a weighted average of all pixels in the image. Given the pixel to be denoised, i, the weights w_{ij} measure the similarity between neighborhoods centered at i and j. The trick is that corresponding neighborhoods are found all along the image imposing a non local nature to the method. Similar methods can be found in [1,3,8]. A review of several denoising strategies and its comparison against non local means can be found in [6] and [4].

In this work we study the behavior of non local denoising methods. First we show the connection of non local means to graph clustering algorithms and use it to study the

E. Bayro-Corrochano and J.-O. Eklundh (Eds.): CIARP 2009, LNCS 5856, pp. 103–110, 2009.

denoising performance. Using synthetic images we will show the limitations of standard non local means and propose an improvement to automatically estimate the parameters of NLM based on noise variance estimation.

2 Non Local Means Denoising

The NLM algorithm [5] estimates the denoised value at pixel i using a weighted average of all pixels in the image:

$$\hat{x}_i = \sum_j \bar{w}_{ij} x_j$$

The weights $\bar{w}_{ij}$ reflect the similarity between pixels i and j based on the distance between neighborhoods around them (see equations (1) and (2)).

Ideally, due to the non local nature of the algorithm, similar neighbors are found across the whole image. This has two drawbacks. The first one is the computational complexity of searching similar neighborhoods across the whole image. The second one is related with the fact that taking weighted averages for all pixels in the image does not achieve the best MSE score for this algorithm. This issue was addressed in [2] and [6] noted the problems with edge pixels. The problem is that in some cases the weights w_{ij} are not able to discriminate between different neighborhoods classes. This is especially the case along edges since pixels along them have less corresponding neighborhoods in the image. Other authors that addressed the computational complexity of NLM encountered this trade off, for instance see [8]. Based on these considerations we can see that a better solution is obtained via averaging only pixels within the same class of neighborhoods. Therefore, the denoising performance depends in a good neighborhood classification. In what follows we will review NLM and show its connection with segmentation based on spectral clustering.

To conclude this discussion we point out that the performance of NLM depends on the selection of the parameter σ. Although in [6] the authors provide some guidance on how to select its value, we will show that the selection of σ has a great impact on the results.

2.1 Graph Formulation of NLM

Let x_i be the original noisy image value at pixel i. Its denoised version using NLM can be obtained as [5]:

$$\hat{x}_i = \frac{\sum_j w_{ij} x_j}{\sum_j w_{ij}} \tag{1}$$

where the weights w_{ij}[1] are computed using a gaussian kernel,

$$w_{ij} = \exp(-||N_i - N_j||^2/\sigma^2) \tag{2}$$

and N_i, N_j are image patches of size $(2K+1) \times (2K+1)$ centered at pixels i and j.

[1] $\bar{w}_{ij} = \frac{w_{ij}}{\sum_j w_{ij}}$.

The equation (1) can be rewritten in matrix notation as follows. Let the matrix W be the one with entries w_{ij}, and D the diagonal matrix with entries $d_{ii} = \sum_j w_{ij}$. If we consider $\mathbf{x}$ as the vectorial version of the image, scanned in lexicographic order, equation (1) can be rewritten as:

$$\hat{\mathbf{x}} = D^{-1}W\mathbf{x} \tag{3}$$

The matrix $L = D^{-1}W$ defines an operator which filters the image, $\mathbf{x}$, to obtain a denoised version $\hat{\mathbf{x}}$. This denoising filter is an image adapted lowpass filter since the operator depends on the image itself. Therefore, the properties of the matrix L determine the denoising result. If we are interested in the properties of a denoising algorithm it is of common use to study the properties of the residual after denoising, $\mathbf{r} = \mathbf{x} - \hat{\mathbf{x}}$. If we write the residual using equation (3) we obtain: $\mathbf{r} = \mathbf{x} - \hat{\mathbf{x}} = \mathbf{x} - D^{-1}W\mathbf{x} = (Id - D^{-1}W)\mathbf{x}$. The matrix $H = (Id - D^{-1}W)$ is the highpass operator associated with the lowpass operator defined by matrix L.

If we view pixels x_i as nodes of a graph connected with weights w_{ij} the matrix H is the normalized Laplacian of the graph which is used in Normalized Cuts (NC). In [10] Malik and Shi presented a relaxed version of the normalized cut which solution is the second eigenvector of H. In this way we show the connection between NLM and segmentation based on NC.

Matrices L and H share the same eigenvectors; if φ_k is an eigenvector of L with eigenvalue λ_k then φ_k is an eigenvector of H with eigenvalue $1 - \lambda_k$. From these considerations we conclude that the eigenvectors and eigenvalues of L and H play an important role in the denoising process.

It can be shown that the multiplicity of the eigenvalue with value one of L corresponds to the number of connected components in the graph [11]. These connected components correspond in our case to the neighborhood classes. So, since an ideal denoising method should average only points in the same classes, the spectrum of the graph related to L is important to measure the performance of the algorithm. We will use the multiplicity of the eigenvalue one to judge the performance of our proposal and compare it with traditional NLM.

3 Experiments with Synthetic Images

To study the denoising performance of NLM we will use a synthetic image with three regions with values 1, 3 and 1 plus Gaussian noise with variance 0.3 (see Figure 2). We consider patches of size 3×3 which gives us six different noiseless neighborhood configurations as show in Figure 2.

Following the same idea proposed in [2] we applied NLM together with a restriction on the number of neighboring patches used for the denoising process. That is, for each pixel to be denoised we considered only pixels with neighborhood similarity greater than ε, that is $\exp(-||N_i - N_j||^2/\sigma^2) \geq \varepsilon$, and computed the error for different values of ε. For this experiment we set $\sigma = 12\sigma_n$ as suggested in [6]. We also computed the error over each region of the image. That is, based on the local configurations show in Figure 2, we segmented the image in six regions and computed the denoising error for each one of them. The results of these simulations are show in Figure 1. As we can see

Table 1. Minimum errors for regions from image in Figure 2

Region	ε	Error NLM	Type
1	0.60	0.0011	non-boundary
2	0.80	0.0036	boundary
3	0.80	0.0099	boundary
4	0.60	0.0005	non-boundary
5	0.75	0.0187	boundary
6	0.75	0.0069	boundary

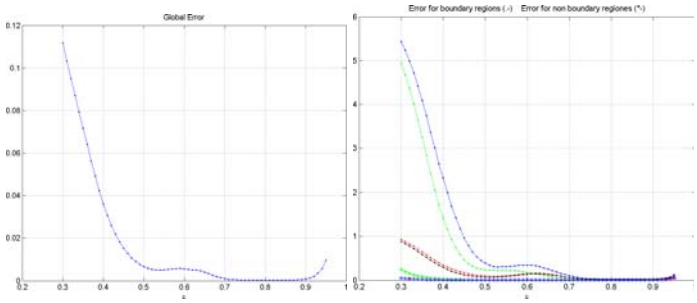


Fig. 1. Left: Global error evolution. Right: Region error evolution.

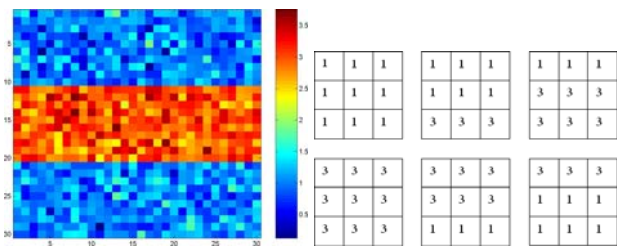


Fig. 2. Left: Noisy image. Right: Neighborhood configurations.

the global error has a U shaped curve. The error decreases as ε increases which means that the error improves while we restrict the set of neighborhoods used. Also as ε goes to one the error increases as the number of points used for the estimation decreases. In the middle we obtain the minimum global error which is quite stable. This means that considering all neighborhoods for the denoising process is clearly not the best option. To understand the reasons of this behavior we computed the errors per region shown in Figure 1. It is clear that boundary regions (neighborhoods with pixels of two regions) perform differently than non boundary regions (neighborhoods with pixels of the same region). Non boundary regions have an almost constant error while boundary regions show a stronger dependence on ε. This explains the obtained global error. In Table 1 we show the minimum errors per regions and the values of ε where these minima are achieved. In next section we will use these results to design an improved NLM.

4 Modified NLM

In this section we address the automatic estimation of the parameters σ of NLM and ε as discussed earlier and present a modified NLM (MNLM).

4.1 Parameter Estimation

Noise variance estimation The estimation of σ will be based on the noise variance. For Gaussian noise the estimation of its variance can be done applying methods as the ones proposed in [9].

Estimation of $\boldsymbol{\sigma}$. Following [6] we set σ proportional to the noise variance: $\sigma = h\sigma_n$. We propose to choose the value of h looking at the expected distances for neighborhoods inside the same class. The expected squared distance for two identical neighborhoods corrupted by Gaussian noise with zero mean and variance σ_n is:

$$\bar{d}^2 = E\{||N_i - N_j||^2\} = E\left\{\sum_{k=1}^{(2K+1)^2} (x_i^k - x_j^k)^2\right\} \tag{4}$$

$$= \sum_{k=1}^{(2K+1)^2} E\left\{(n_i^k - n_j^k)^2\right\} = 2(2K+1)^2\sigma_n^2. \tag{5}$$

We set the value of h in order to obtain weights greater than γ for similar neighborhoods. In this way the value of h is defined as the one that satisfies the following equation:

$$\exp\left(\frac{-\bar{d}^2}{h^2\sigma_n^2}\right) = \gamma.$$

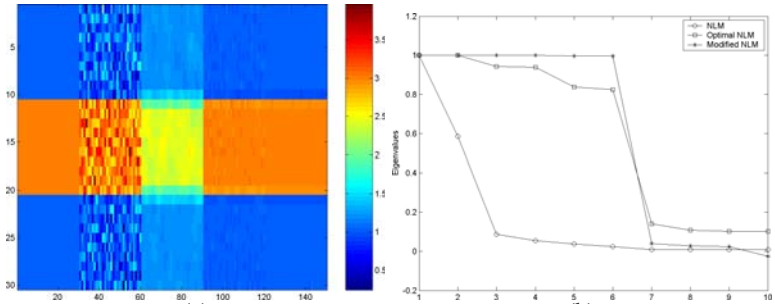
If we substitute $\bar{d}^2$ in previous equation we obtain:

$$h = \sqrt{\frac{2(2K+1)^2}{\log(1/\gamma)}}$$

Finally we have to select the value for ε. As we said before better results are obtained when only neighborhoods with similarities greater than ε are considered. Therefore, we let $\gamma = \varepsilon$. So, instead of parameter σ we have a new parameter ε which controls the neighborhoods considered in the estimation and the value of σ. Also this parameter does not depend on the input image but only on the noise level estimation. In order to consider only neighborhoods of the same class ε must take values close to one. In following sections we will analyze this proposal at the light of the relationship between NLM and segmentation and show that taking $\varepsilon = 0.8$ gives excellent results for a set of real images.

Table 2. Global MSE scores

	NLM	Best NLM	Modified NLM
MSE	0.2046	0.0012	0.0006



(a) (b)

Fig. 3. (a) Fron left to right: noiseless image, noisy image, result of NLM, result of best NLM and result of modified NLM. (b) Eigenvalues.

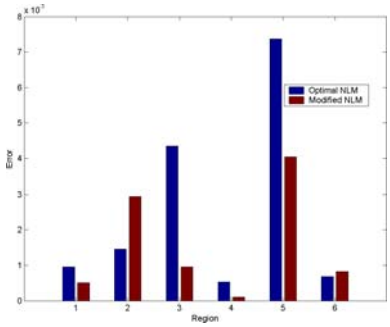


Fig. 4. NLM against MNLM: Errors per region

4.2 Modified NLM and Graph Cuts

In this section we will compare the performance of MNLM against the original NLM using the results from section 3 and the image showed in Figure 2. The image in Figure 2 was filtered with three algorithms: the original NLM with $\sigma = 12\sigma_n$, MNLM with $\varepsilon = 0.8$ and the best NLM in which case we selected the parameter $\sigma = 5\sigma_n$ that gives the smallest global MSE. The obtained MSE errors are presented in Table 2. In Figure 3(a) we show the original noiseless image, the noisy image and the images corresponding to the methods in evaluation. As we can see the original NLM gives the worst result in terms of MSE and visual quality while the modified NLM obtains the best overall performance (see MSE scores in Table 2). If we look at MSE per region we can see in Figure 4 that MNLM performs better than NLM in four out of six of the regions.

Finally we present the eigenvalues of the corresponding matrices L for each method. In Figure 3(b) we show the first ten eigenvalues for each method. It is clear that MNLM has better performance since it has six eigenvalues of value one corresponding to the six regions present in the image. We recall that the multiplicity of the eigenvalue one corresponds to the number of connected components in the graph, i.e. the number of neighborhood classes which in this case is six.

4.3 Results for Real Images

Here we compare the best performances of NLM and our modified NLM (MNLM). Each of the images in Table 3 was contaminated with independent and additive Gaussian noise with $\sigma_n = 10$. We used neighborhoods were of 3×3 and to reduce the computational complexity we used a search window of 21×21. For the evaluation of the results we use Mean Square Error (MSE) and the Structural Similarity Index (SSIM) proposed in [12] which compares the similarity between images using perceptual factors.

With this results we confirm that NLM attains the best result at $h = 3$ in all cases. This contrast with the values of h selected by Buades, Morel and Coll in [5,6] where they suggest $h \in [10, 15]$. Clearly with their selection for h the results are not the best possible. We confirm this based on MSE and SSIM. We must stress that in all cases the optimum is achieved with the same value of h. On the other hand, the results obtained with our modified NLM method present similar results as the ones given by NLM. Therefore based only on MSE and SSIM we cannot say which method is better. As for MNLM the best score values are obtained with $\epsilon = 0.8$ in all case but one.

Result Analysis. To conclude the evaluation we give an explanation on why the best results of NLM are similar to the ones of MNLM. In previous experiments the parameter h which gives the best results for NLM is in all cases 3. The difference between both methods is the width of the Gaussian kernel. For modified NLM the width is $\sigma^2_{MNLM} = \frac{2(2K+1)^2}{\log(1/\gamma)}$ and for NLM the width which produces the best results is $\sigma^2_{NLM} = 3^2\sigma^2_n$. The other difference is that for MNLM we consider only weights above $\varepsilon = \gamma$ and for NLM we consider all weights. The distances for which MNLM gives weights γ are $d^2_\gamma = 2(2K+1)^2\sigma^2_n$. If we substitute this distance in the NLM kernel we get: $\exp(-2(2K+1)^2/3^2) \approx 0.13$. Therefore the corresponding weights for NLM are small and explain the similarity between results of NLM and MNLM. We confirmed the same results using neighborhoods of size 5×5 but due to the lack of space we can not report them here.

Table 3. Minimum errors for regions from image in Figure 2

	NLM			MNLM		
Image	h*	MSE*	SSIM*	ε*	MSE*	SSIM*
Barbara	3	30.11	0.913	0.75	28.93	0.922
Baboon	3	63.45	0.895	0.80	70.67	0.890
Couple	3	35.13	0.883	0.80	35.97	0.884
Einstein	3	35.03	0.859	0.80	36.01	0.858
Goldhill	3	33.78	0.868	0.80	34.61	0.869

5 Conclusions

In this work we study the relationship between non local denoising methods and spectral graph properties. Based on these results we proposed a modification of NLM which automatically estimates the parameter σ. We justified the proposed algorithm using the connection between NLM and graph clustering. Based on simulations we showed that this approach outperforms the NLM with the parameters suggested by Buades and colleagues in [5]. Furthermore we showed that this parameter setting is the best one for all images tested when comparing the MSE scores.

References

1. Awate, S.P., Whitaker, R.T.: Unsupervised, information-theoretic, adaptive image filtering for image restoration. IEEE Trans. Pattern Anal. Mach. Intell. 28(3), 364–376 (2006)
2. Bertalmio, M., Caselles, V., Pardo, A.: Movie denoising by average of warped lines. IEEE Transactions on Image Processing 16(9), 2333–2347 (2007)
3. Boulanger, J., Kervrann, Ch., Bouthemy, P.: Adaptive space-time patch-based method for image sequence restoration. In: Workshop on Statistical Methods in Multi-Image and Video Processing (SMVP 2006) (May 2006)
4. Buades, A., Coll, B., Morel, J.M.: The staircasing effect in neighborhood filters and its solution. IEEE Transactions on Image Processing 15(6), 1499–1505 (2006)
5. Buades, A., Coll, B., Morel, J.M.: Denoising image sequences does not require motion estimation. In: Proc. IEEE Conf. on Advanced Video and Signal Based Surveillance, pp. 70–74 (2005)
6. Buades, A., Coll, B., Morel, J.M.: A review of image denoising algorithms, with a new one. SIAM Multiscale Modeling and Simulation 4(2), 490–530 (2005)
7. Efros, A., Leung, T.: Texture synthesis by non-parametric sampling. In: IEEE Int. Conf. on Computer Vision, ICCV 1999, pp. 1033–1038 (1999)
8. Mahmoudi, M., Sapiro, G.: Fast image and video denoising via nonlocal means of similar neighborhodds. IEEE Signal Processing Letters 12(12), 839–842 (2005)
9. Olsen, S.I.: Noise variance estimation in images. In: Proc. 8th SCIA, pp. 25–28 (1993)
10. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(8), 888–905 (2000)
11. von Luxburg, U.: A tutorial on spectral clustering. Statistics and Computing 17(4) (2007)
12. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: From error visibility to structural similarity. IEEE Trans. Image Processing 13(4), 600–612 (2004)