Adapted Non-Local Means Filter using Variable Window Size

Ammar M. Alhosainy and Ehab F. Badran
Arab Academy for Science and Technology, Alexandria, Egypt
eammar@aast.edu, ebadran@aast.edu

ABSTRACT

In this paper, a novel improvement idea in image restoration by Non-Local means (NL-means) filter is proposed. Most of special image denoising techniques require a region size called neighborhood window for denoising, window-size selection is usually done by visual inspection based upon experience or hit and trial. It is shown in this paper that varying the window-size for each pixel according to window variance obtains better results. Also, an investigation on choosing the optimum filtering parameter for each window size is presented. Tests on standard images are carried out and presented. The results show that the performance is very close and in some cases even surpasses state-of-art denoising techniques.

Key Words: Image denoising, Non-Local Means.

1. Introduction

Image denoising is the operation of removing unwanted noise, from a noisy image. Denoising is still an open problem in image processing. Its great challenge is dealing with rich content like texture. A recent outstanding review of many image denoising methods can be found in [1]. Gaussian filters have been largely used in some applications, but they have the disadvantage of blurring edges due to averaging of non similar patterns. In order to avoid this problem many edge preserving filters have been proposed. Probably the best known is the Anisotropic Diffusion Filter [2], Such filters respect edges by averaging pixels in the orthogonal direction of the local gradient. However such filtering can erase small features and may change image statistics. Also the total variation minimization was introduced by Rudin and Osher [3], this filter maintains straight edges but other details and texture can be over smoothed.

As an example of filters in the other domain, Wavelet based filters [4,5] have been used but such filters tend to introduce unwanted artifacts in the image.

Many filtering methods are based on the signal averaging principle which uses the spatial redundancy in the image. Generally speaking, the information in a natural image is redundant to some extent. Based on this observation, Buades [1] developed a Non-Local means (NL means) image denoising algorithm which takes full advantage of image redundancy. The basic idea is that images contain repeated structures, and averaging them will reduce the (random) noise.

In this paper, we propose an improved filtering method, which helps to calculate each pixel's weight in a better way, using Non-Local means filter. This paper is organized as follows: Section II presents a short overview of the Non-Local means algorithm, sections III and IV we suggest new ideas for NL-means filter improvement, section V shows the results, and finally we end with our conclusions.

2. The Non Local Means Algorithm

The NL-means filter [1] is an evolution of the Yaroslavsky filter [6] which averages similar image pixels defined according to their local intensity similarity. The main difference between the NL-means and this filter is that the similarity between pixels has been made more robust to noise by using a region comparison, rather than pixel comparison and also that matching patterns are not restricted to be local. That is, pixels far from the pixel being filtered are not penalized.

Given an image Y the filtered value at a point i using the NL-means method is computed as a weighted average of neighboring pixels N_i in the image following this formula [7],

$$NL(Y(i)) = \sum_{j \in N_i} w(i,j)\, Y(j) \qquad (1)$$

with $0 \le w(i,j) \ge 1$ and $\sum_{j \in N_i} w(i,j) = 1$.

where i is point being filtering and j represents any other image pixel. The weights $w(i, j)$ are based on the similarity between the neighborhoods N_i and N_j of pixels i and j. N_i is defined as a square neighborhood window indices cantered around pixel i. Theoretically, noise filtering performed must be considered as an estimation task. Therefore, the process of linear weighting and the weight factors *w(i, j)* can be regarded as the calculation which computes the most likely noise free grey-level value of the selected pixel, on the basis of the measured evidence. A simple example which leads to this form of solution can be derived using Likelihood, on the assumption that the measurements form each image (*Y(i)* and *Y(j)*) can be taken as an independent estimate of the noise free value, drawn from a Gaussian distribution with variance *1/w(i, j)*. The process of selecting the most effective noise filtering algorithm can be considered as a way of optimizing the match between the assumed computational form and the statistical distributions in the data.

In the case of NL-means, *w(i, j)* is calculated as:

$$w(i,j) = \frac{1}{W(i)} exp(-d(i,j)/h^2) \quad (2)$$

where h is the filtering parameter (will be discussed shortly), $d(i, j) = \| v(N_i) - v(N_j) \|^2_a$ is a Gaussian weighted Euclidean distance −Efros and Leung [8] showed that the L^2 distance is a reliable measure for the comparison of image windows in a texture patch− weighted by a Gaussian kernel of zero mean and variance a, N_i and N_j are the neighborhood pixels indices for pixels i and j respectively, so $v(N_i)$ and $v(N_j)$ are windows centered at pixels i and j respectively with a user-defined radius n, $W(i)$ is the normalization factor and defined as,

$$W(i) = \sum_{j \in S} exp(-d(i,j)/h^2) \quad (3)$$

where S is the search window, the original definition of the NL-means algorithm considers that each pixel can be linked to all other pixels in the image, but practically the number of pixels taken into account in the weighted average can be restricted in a neighborhood that is called "search window" S_i of size $(2s+1)^2$, centered at the current pixel i.

Several accelerated versions of this filter have been proposed. In [1] Buades *et al.* recommended the vectorial (or block-based) NL-means filter which amounts to simultaneously restore pixels of a whole patch $v(N_i)$ from nearby patches. The restored value at a pixel i is finally obtained by averaging the different estimators available at that position. Also, other recent accelerated versions of NL-means [9-12] use two filters to pre-classify the image patches according to fundamental characteristics such as their average gray values and gradient orientation [9], or their first and second moments [10], so only blocks with similar characteristics are used to compute the weights. Another accelerated version is proposed in [12] suggesting that arranging the data in a cluster tree, the structuring of data allows for fast an accurate preselection of similar patches.

3. Variable Window Size

Antoni Buades in [1] stated that, a similarity window of size 7×7 or 9×9 can be taken for grey level images with little noise, but these fixed size windows will not yield good results for all kinds of images.

The problem with the fixed size similarity window is that, in case of large window, some details could be removed from the image, blurring singular points (i.e. pixels with no similar patches, like image corners and peaks or valleys) by averaging them with non similar patches, otherwise, in case of small window, there will be a lot of patches similar to the current patch, resulting in non accurate estimation. This means that, in flat regions (low variance regions), large similarity window is needed, in other regions containing a lot of details (high variance), small window size is needed in order to find similar patches and to estimate the current pixel more accurately.

In order to adapt the similarity window size to any kind of images, it has to be changed gradually according to the patch's variance. Initially, the patches' variances must be calculated in a pre-processing stage in the NL-means algorithm, using a chosen fixed window size for the patches, the resulting patch-variance will be stored in a variance matrix.

The proposed variable similarity window size is chosen to be inversely proportional to patches variance, this relation can be formulated as follows,

$$f(i) = f_{\max} - \frac{f_{\max} - f_{\min}}{\sigma^2_{N\max} + \sigma^2_{N\min}} (\sigma^2_N(i) + \sigma^2_{N\min}) \quad (4)$$

where f_{max} and f_{min} are the maximum and minimum similarity window half width, $\sigma^2_{N\max}$ and $\sigma^2_{N\min}$ are the maximum and minimum variances from the pre-calculated patch-variance matrix, Fig. 1 shows the relation.

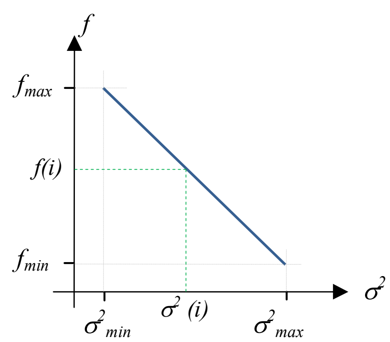

Fig. 1 The relation between patches variance and the similarity window size half width.

Using a first order model in calculating the similarity window size according to windows variance, will give a range of window sizes, which have the same distribution of the image window variances. The value of f_{max} and f_{min} can be fixed for any kind of images, the adaptive similarity window size will obtain nearly optimum results without needing to find the single optimum value of the similarity window size for every noisy image.

The calculation step of the patch-variance matrix will delay the algorithm for only one second on a normal computer with Pentium IV 2.1 GHz CPU and 2GB RAM. The patch-variance matrix is also used by the accelerated version of the algorithm presented in [10] to reduce its complexity, so, the adaptive similarity window size technique will not affect the complexity of the algorithm.

4. Filtering Parameter

The filtering parameter *h* controls the decay of the exponential expression in the weighting scheme (see equation (3)). Choosing a very small *h* parameter tends to produce noisy results similar to the input, while very large *h* gives a very smoothed image, this means that, *h* controls the smoothing degree of the filtered image. First it has been shown in [1] that the optimal filtering parameter *h* is proportional to the standard deviation of the noise, also the statistical derivation in [13] enables the automatic choice of *h* once the noise variance is estimated, namely $h \approx 12\sigma$, where σ is the noise standard deviation. As in equation (3), the power of the decaying function varies according to the window size for the same pixel, this means that, changing the window size will change *h* parameter indirectly. The author of [14] stated that *h* must be independent of the choice of the window size, to do that, the Euclidean distance $\|\,.\,\|^2$ must be normalized by the number of elements $|N|$, as if it is said that, the filtering parameter will be $h^2 = 2\beta\sigma^2|N|$, where β is a constant.

This proportionality will deviate *h* from the optimum value, because small similarity window size hardly contains image details, so the signal to noise ratio is very low, in this case, *h* needs to be high to do a hard smoothing of noise and to estimate the correct value of the pixel, but in a large similarity window size, the signal to noise ratio is high, because it contains a lot of image details, so *h* needs to be small to preserve image details. Figure 2 shows the signal to noise ratio of well-known standard test images, as it is seen, SNRs of all test images are monotonically increasing with the similarity window size.

This explanation leads to one conclusion, which is, the parameter *h* must be inversely proportional to the similarity window size in order to obtain the best results of any similarity window size used. This conclusion is confirmed by the following curves shown in Fig. 3, each point on every curve relates the similarity window size half width to its optimum *h* value, also Fig. 4 confirms this proportionality for different values of the noise standard deviation.

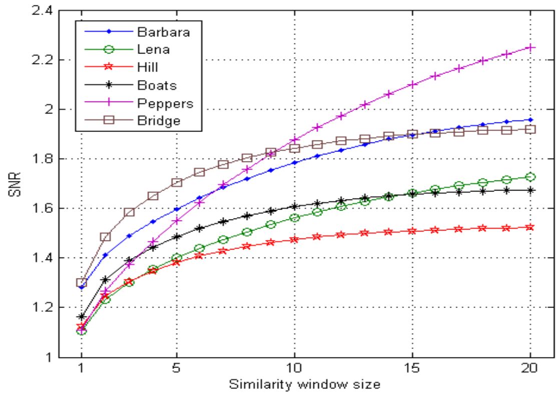

Fig. 2 The SNR of Barbara, Lena, Hill, Boats, Peppers, and Bridge, all with AWGN of zero mean and 20 standard deviation.

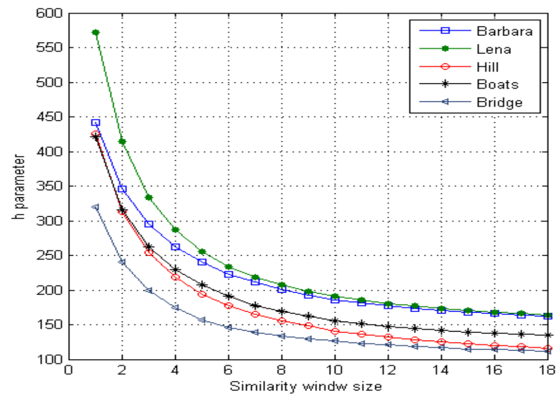

Fig. 3 The inverse proportionality between optimum *h* parameter and the similarity window size half width, for Barbara, Lena, Hill, Boats, and Bridge images, all of them with additive white Gaussian noise of zero mean and 20 standard deviation.

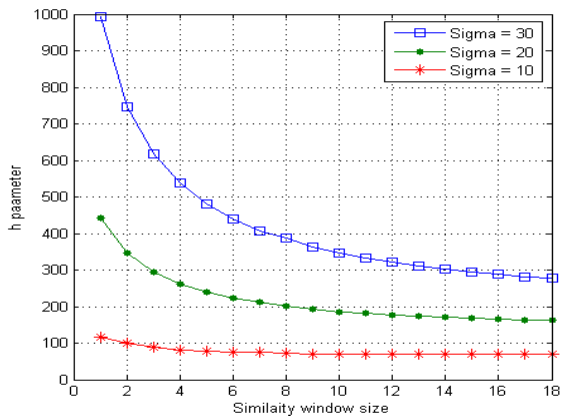

Fig. 4 The inverse proportionality between optimum *h* parameter and the similarity window size half width, for Barbara image, with additive white Gaussian noise of zero mean and standard deviation of 10, 20, and 30.

5. Results

The proposed adaptive similarity window sizes has been tested on an 8-bit standard grayscale test images corrupted by an additive white Gaussian noise (AWGN) (*PSNR = 22.13 db*, $\sigma = 20$), a finite range of integer half width values is assumed as,

$$[C\text{-}m\ ,\ C\text{+}m] = [f_{min}\ ,f_{max}]$$

where m is half the range from f_{min} to f_{max} ,$0 \le m \le C\text{-}1$, and C is the center half width, for this experiment, the patch variance matrix was calculated using window size half width equal to 7. The proposed algorithm has been applied to five standard test images, Table 1 shows the PSNR resulted from applying the variable window sizes NL-means algorithm with $C = 7$ and $m = 5$, also Fig. 5 shows the improvement in the visual quality.

Table 1 The PSNR output, using adaptive NL-means and original NL-means.

	Barbara	Lena	Boats	Hill	Peppers
NL-means with Adaptive similarity window size	**30.757**	**32.273**	**29.860**	**29.666**	**30.213**
Original NL-means	30.660	32.000	29.515	29.437	30.050

6. Conclusion

The proposed framework yields an improved NL-means filter, varying the (patch) window size according to the patch variance will produce results better than the fixed widow size. Also it is shown that the optimum h parameter is proportional to the noise standard deviation and inversely proportional to the window size.

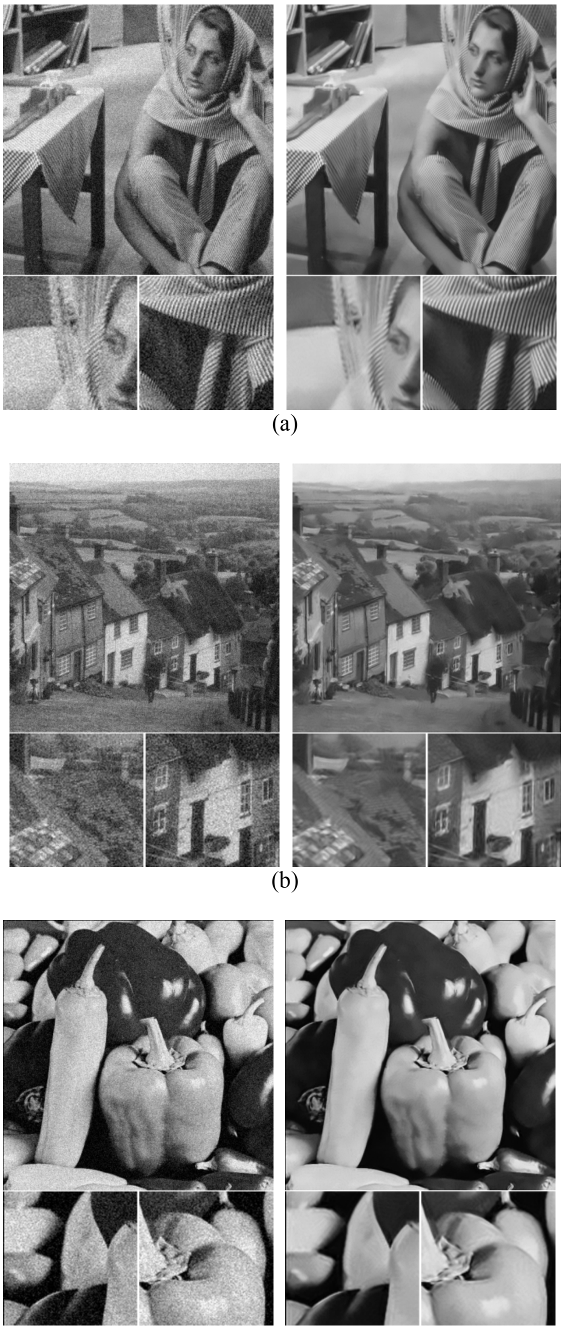

Fig. 5 Noisy and restored images with two enlarged fragments from them.

References:

[1] A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Modeling and Simulation (SIAM Interdisciplinary Journal),* Vol. 4, No. 2, 2005, pp 490-530.

[2] P. Perona and J. Malik, "Scale space and edge detection using anisotropic diffusion," *IEEE T. Patt. Anal. ach. Intell. 12*, 1990, pp. 629-239.

[3] L. Rudin, S. Osher, and E. Fatemi "Nonlinear total ariation based noise removal algorithms," *Physica D (2992) 60*, 1992, pp. 259-268.

[4] J. C. Wood and K. M. Johnson, "Wavelet packet denoising of magnetic resonance images: importance of Rician noise at low SNR," *Magnetic Resonance in Medicine*, Vol. 41, 1999, pp. 631-635.

[5] R. D. Nowak, "Wavelet-based Rician noise removal for magnetic resonance imaging," *IEEE Transactions on Image Processing*, 8:10, 1999, pp. 1408 -1419.

[6] L. P. Yaroslavsky, *Digital Picture Processing - An Introduction*, Springer Verlag, 1985.

[7] J. V. Manjon, M. Robles, and N. A. Thacker, "Multispectral MRI de-noising using non-local means," *MIUA*, 2007.

[8] A. Efros and T. Leung, "Texture synthesis by non parametric sampling," *Proc. Int. Conf. Computer Vision (ICCV 99)*, Vol. 2, 1999, pp. 1033-1038.

[9] M. Mahmoudi and G. Sapiro, "Fast image and video denoising via nonlocal means of similar neighborhoods," *IEEE Signal Processing Letters*, Vol. 12, 2005, pp. 839-842.

[10] P. Coupe, P. Yger, and C. Barillot, "Fast non-local means denoising for 3D MR images," *MICCAI'06, Copenhagen*, 2006.

[11] Y.-Li Liu, J. Wang, X. Chen, Y.-W. Guo, and Q.-S. Peng, "A robust and fast non-local means algorithm for image denoising," *Journal of Computer Science And Technology*, March 2008.

[12] T. Brox, O. Kleinschmidt, and D. Cremers, "Efficient nonlocal means for denoising of textural patterns," *IEEE Transactions on Image Processing*, August 2008.

[13] A. Kervrann, J. Boulanger, and P. Coup´e, "Bayesian non-local means filter, image redundancy and adaptive dictionaries for noise removal," *In F. Sgallari, A. Murli, and N. Paragios, editors, Scale Space and Variational Methods in Computer Vision*, Vol. 4485 of LNCS, Springer 2007, pp. 520–532.

[14] P. Coupe, P. Hellier, S. Prima, C. Kervrann, and C. Barillot1, "3D wavelet subbands mixing for image denoising," *International Journal of Biomedical imaging*, 2008.