Adaptive regularization of the NL-means for video denoising

Camille Sutour, Jean-François Aujol, Charles-Alban Deledalle, Jean-Philippe Domenger

ADAPTIVE REGULARIZATION OF THE NL-MEANS FOR VIDEO DENOISING

Camille Sutour[1,2], Jean-François Aujol[1], Charles-Alban Deledalle[1], Jean-Philippe Domenger[2]

[1]Université de Bordeaux, IMB, UMR 5251 CNRS, 33405 Talence, France
[2]Université de Bordeaux, LABRI, UMR 5800, 33405 Talence, France

ABSTRACT

We derive a denoising method based on an adaptive regularization of the non-local means. The NL-means reduce noise by using the redundancy in natural images. They compute a weighted average of pixels whose surroundings are close. This method performs well but it suffers from residual noise on singular structures. We use the weights computed in the NL-means as a measure of performance of the denoising process. These weights balance the data-fidelity term in an adapted ROF model, in order to locally perform adaptive TV regularization. Besides, this model can be adapted to different noise statistics and a fast resolution can be computed in the general case of the exponential family. We adapt this model to video denoising by using spatio-temporal patches. Compared to spatial patches, they offer better temporal stability, while the adaptive TV regularization corrects the residual noise observed around moving structures.

Index Terms— Video denoising, NL-means, variational methods, total variation

1. INTRODUCTION

Image and video processing is a wide domain that includes reconstruction, segmentation, super-resolution, etc. Most of these tasks require a high quality input signal, so denoising is a fundamental issue. A wide range of image denoising techniques have been studied, among which the variational methods and the non-local models.

The variational methods consist in minimizing an energy in order to force certain properties on the unknown solution. The ROF model [1] minimizes the total variation, forcing the restored image to be smooth while preserving the edges. The solution u_{TV} is obtained by minimizing the following energy:

$$u_{\mathrm{TV}} = \underset{u}{\operatorname{argmin}} \; -\log p(g|u) + \lambda \,\mathrm{TV}(u). \qquad (1)$$

The term $-\log p(g|u)$ is a data fidelity term based on the log-likelihood, that adapts to the noise statistic. $\mathrm{TV}(u) = \sum \|(\nabla u)_i\|$ is a regularization term and $\lambda > 0$ is the parameter that sets the compromise between data fidelity and smoothness. A regularization parameter that is set too high will overly smooth textures, but if set too low it will not allow a recovery of flat areas, creating a *staircasing effect*.

Another family of denoising techniques are the non-local methods. Among them, the NL-means algorithm [2] exploits the spatial redundancy in natural scenes. It performs a weighted average of pixels whose surroundings are similar, by comparing patches, ie small windows extracted around each pixel. For each pixel i in the image domain Ω, the solution of the NL-means is the following :

$$(u_{\mathrm{NL}})_i = \frac{1}{Z_i} \sum_{j \in \Omega} w_{i,j} g_j \quad \text{where} \quad Z_i = \sum_{j \in \Omega} w_{i,j}. \qquad (2)$$

The weights $w_{i,j} \in [0,1]$ reflect the likeliness of two noisy pixels to belong to the same structure, ie to have the same true gray level. They are defined to select the pixels j whose surrounding patches are similar to the one extracted around the interest pixel i:

$$w_{i,j} = \varphi\left(-\frac{d(g(P_i), g(P_j))}{2|P|h^2}\right) \qquad (3)$$

where φ is a kernel decay function, h is a filtering parameter and d a distance function that measures the similarity between the two patches P_i and P_j of size $|P|$ extracted around the pixels i and j. More explanations regarding the interpretation of this kernel can be found in [2]. The distance d can be adapted to the statistics of the noise, following [3], making the NL-means algorithm robust to different noise statistics. The non-local methods have been widely used and improved [4, 5, 6, 7]. The NL-means perform well on flat areas and repetitive structures such as textures, but they suffer from the lack of similarity on small or singular structures, resulting in insufficient denoising referred to as the *rare patch effect*. Besides, the NL-means have been naturally adapted to video denoising [8]. Extending the non-local search for candidates to the temporal dimension strengthens the similarity, hence allowing better denoising without prior motion compensation. Further improvements have been studied regarding the shape of the patches. As suggested in [9] in an inpainting context, using spatio-temporal patches enforces temporal consistency between frames. This prevents a glittering effect that occurs on static areas. However, the 3-dimensional patches make the

C. Sutour would like to thank the DGA and the Aquitaine region for funding her PhD. J.-F. Aujol acknowledges the support of the Institut Universitaire de France. This study has been carried out with financial support from the French State, managed by the French National Research Agency (ANR) in the frame of the "Investments for the future" Programme IdEx Bordeaux - CPU (ANR-10-IDEX-03-02).

search for candidates more selective, resulting in an increased residual noise around objects in movement.

Several approaches have combined the non-local means with variational methods. They are often based on a non-local regularization term. The model studied here is based on a non-local data-fidelity term that results in a locally adaptive TV regularization, allowing the NL-means and the TV regularization to complete each other.

The goal of this paper is to reduce the *rare patch effect* observed on images and videos as a result of the NL-means algorithm. We perform a local regularization of the NL-means, using a non-local data-fidelity term combined with total variation. We propose an intuitive model that is adapted to different noise models and we derive a simple resolution scheme in the general case of the exponential family, that includes a large range of noises encountered in imaging devices. Besides, our model offers a natural extension to video denoising. We use the temporal NL-means with 3-dimensional patches, that we regularize using our adaptive spatial TV regularization. This allows us to guarantee temporal consistency without suffering from residual noise. The main contributions of our work are the simple model and its intuitive interpretation, the ability to deal with different noise statistics thanks to a general model, and the natural extension to video denoising that provides thanks to the spatio-temporal patches efficient denoising along with a better temporal consistency.

2. R-NL

2.1. Proposed model

We propose here to combine the NL-means with TV regularization in order to reduce both the *rare patch effect* and the *staircasing effect*. We perform a TV minimization with a non-local data-fidelity term, using the following equation :

$$u_{\text{R-NL}} = \underset{u}{\operatorname{argmin}} -\sum_{i\in\Omega}\sum_{j\in\Omega} w_{i,j} \log p(g_j|u_i) + \lambda \operatorname{TV}(u) \quad (4)$$

When the noise belongs to the exponential family, we can show that this is equivalent to the following :

$$u_{\text{R-NL}} = \underset{u}{\operatorname{argmin}} -\sum_{i\in\Omega} Z_i \log p((u_{NL})_i|u_i) + \lambda \operatorname{TV}(u) \quad (5)$$

This is interpreted as a fidelity term based on the solution of the NL-means, weighted by the sum of the weights computed in the NL-means. This allows both the NL-means and the TV regularization to complete each other: on areas where the redundancy is high, the NL-means select many candidates so the sum of the weights is high. In the energy to minimize, the data fidelity term is then prominent over the regularization term, so the solution is close to the NL-means. This provides good smoothing and prevents the *staircasing effect* observed on smooth areas when treated with TV minimization. In small structures and edges where the redundancy is low, the NL-means select fewer candidates so the sum of the weights is low. The regularization term becomes prominent over the data fidelity term, so it will cost less to minimize the total variation of the image. The solution tends to a TV solution, preserving edges while reducing the *rare patch effect*.

Other approaches have combined the NL-means with total variation. Gilboa and Osher have developed a non-local regularization based on the non-local gradient that allows to smooth the flat areas while preserving the textures. Another non-local regularization based on the similarity of patches has been introduced in [10]. Louchet and Moisan [11] also combine the NL-means with total variation, by turning the total variation into a local filter. Their TV-means algorithm performs local TV regularization where the NL-means do not manage to find enough candidates, and they reduce both the *staircasing effect* and the *rare patch effect* with an iterative scheme. We use a non-local data-fidelity term instead of a non-local regularization. This type of approach has been used in a super-resolution context by Protter *et al.* [12] and d'Angelo and Vandergheynst [13]. They use the normalized weights issued from the NL-means to define a non-local data-fidelity term. In a denoising context, this results in computing the solution of the NL-means, then applying TV regularization. In the proposed model, the non-local weights are not normalized so they weigh the data-fidelity term, which makes the TV regularization adaptive. These non-normalized weights act as a measure of performance of the NL-means: on constant areas where the similarity is high the NL-means are able to find many candidates, so the denoising is efficient and free of the *staircasing effect*. On small structures and edges where the similarity is insufficient, the adaptive TV regularization locally reduces the *rare patch effect*.

2.2. Resolution for the exponential family

This model can be solved efficiently in the general case of the exponential family. It includes additive white Gaussian noise, Poisson noise and some multiplicative noises that are encountered in image processing problems such as medical imaging, astronomy, etc. A probability law belongs to the exponential family if it can be written under the following form :

$$p(g|u) = h(g)\exp(\eta(u)T(g) - A(u)) \quad (6)$$

where h, T, η and A are known functions. Based on equation (5), the extended model is the following :

$$\begin{aligned} u_{\text{R-NL}} &= \underset{u}{\operatorname{argmin}} \textstyle\sum_{i\in\Omega} Z_i A(u_i) - \eta(u_i)\mu_i + \lambda \operatorname{TV}(u) \\ \text{with } Z_i &= \textstyle\sum_{j\in\Omega} w_{i,j} \quad \text{and} \quad \mu_i = \textstyle\sum_{j\in\Omega} w_{i,j} T(g_j). \end{aligned} \quad (7)$$

μ and Z can be calculated with a quick implementation of the NL-means. We refer the interested reader to [14] for a more complete description of a fast way to compute the weights. Then the minimization step is achieved thanks to standard

minimization algorithms, depending on the type of noise involved. In the Gaussian case or the Poisson case we can use the primal-dual methods given in Chambolle and Pock's algorithm [15], and its adapted version to Poisson noise described in [16]. In the gamma case, the functional is not convex, so there is no guarantee as to the existence of a unique minimizer, but we can show that a minimization algorithm will converge towards a stationary point [17]. Besides, the data fidelity term is differentiable so we can use the forward-backward algorithm [18]. A general implementation of the R-NL algorithm is given in algorithm 1. More details and some results regarding the denoising process in the case of non-Gaussian noise are given in [19].

Algorithm 1 R-NL

Require: g : noisy input image,
h : filtering parameter
$|P|$: patch size
$\mathcal{N}$: size and shape of the search neighborhood
λ : regularization parameter

NL-means step - computation of the weights

for $i \in \Omega$, $j \in \mathcal{N}(i)$ **do**

Compute $w_{i,j} = \varphi\left(-\frac{d(g(P_i),g(P_j))}{2|P|h^2}\right)$

end for

return

$$Z = \left(\sum_{j\in\mathcal{N}(i)} w_{i,j}\right)_{i\in\Omega} \qquad \mu = \left(\sum_{j\in\mathcal{N}(i)} w_{i,j}T(g_j)\right)_{i\in\Omega}$$

Minimization step

$$u_{\text{R-NL}} = \underset{u}{\operatorname{argmin}} \sum_{i\in\Omega} Z_i A(u_i) - \eta(u_i)\mu_i + \lambda\,\mathrm{TV}(u)$$

return $u_{\text{R-NL}}$

Based on the NL-means for video denoising, we can adapt the R-NL algorithm to image sequences: the NL-means step benefits from the temporal information while the TV regularization is applied spatially to reduce the residual noise.

3. ADAPTATION TO VIDEO DENOISING

Video denoising benefits from the temporal information throughout the frames, provided that the frames can be put in correspondence with each other. The NL-means have been adapted to video denoising in [20] and they achieve spatio-temporal filtering without prior motion compensation. In fact, the authors have even shown that motion-compensation is counter-productive, since it reduces the number of candidates and it is often inaccurate. We follow the idea of the NL-means and adapt our R-NL algorithm to video denoising. We perform the NL-means using spatio-temporal search windows, then we apply the adaptive TV regularization spatially on each frame to reduce the residual noise.

However, Liu et al. advocate in [21] that motion compensation is in fact essential, even in a non-local context. They integrate optical flow estimation in the NL-means framework in order to perform more efficient denoising while guaranteeing better temporal stability. In both NL-means and R-NL, if a small number of frames are used for the search window (which is often the case in order to lower the computational costs), no temporal regularity is guaranteed. Indeed, here contrary to motion-compensation problems no assumption is made regarding the gray level of a pixel during its trajectory, so it is allowed to vary with time. This results in a *glittering effect* when looking at static objects on the video. To reduce this effect, we use 3D patches instead of 2D patches, as suggested in an inpainting context in [9]. We compute spatio-temporal patches that compare neighborhoods using the temporal information as well as the spatial one, in order to force temporal consistency. In the 3D-NL-means alone, the use of three-dimensional patches favors the rare patch effect, since candidates are made harder to find. Using the adaptive TV regularization afterward balances this drawback, resulting in a more stable video without residual noise or glittering effect.

4. RESULTS AND DISCUSSIONS

We focus in this section on video denoising, where our regularized model takes all its meaning. We demonstrate the results on three image sequences, *target*, *tennis* and *bicycle*, that are available for download[1]. We compare the performance of the standard NL-means algorithm adapted to video denoising using 2D patches and 3D search window (NL-means-2D), the R-NL algorithm based on the NL2D combined with spatial adaptive TV regularization (R-NL-2D), the NL-means algorithm using three-dimensional patches (NL-means-3D) the R-NL based on the NL3D (R-NL-3D), and the video denoising algorithms V-BM3D [22] and BM4D [23] that respectively use 2D and 3D patches. Figure 1 shows a frame of the video *tennis* denoised by the algorithms described above. Fig. 1-a is the original frame, fig. 1-b is the noisy image degraded by Gaussian noise with standard deviation $\sigma = 20$, and fig. 1-c to h are the denoised versions using respectively NL-means-2D, NL-means-3D, R-NL-2D, R-NL-3D, V-BM3D and BM4D. Table 1 displays the mean PSNR of the denoised videos using the above methods. Our results with the proposed R-NL-3D method are quite competitive, and they offer a good compromise between preservation of the small structures and efficient denoising, without suffering from residual noise (on the NL-means for example, fig. 1-d) or excessive smoothing (with BM4D, fig. 1-h). We also measure the temporal consistency ensured by the different algorithm, that we obviously cannot illustrate on paper. We compute temporal variance on areas that are static through time : on areas that do not move during a part of the sequence, the pixel value should be unchanged from one frame to another, so the temporal standard devia-

[1] http://image.math.u-bordeaux1.fr/RNL

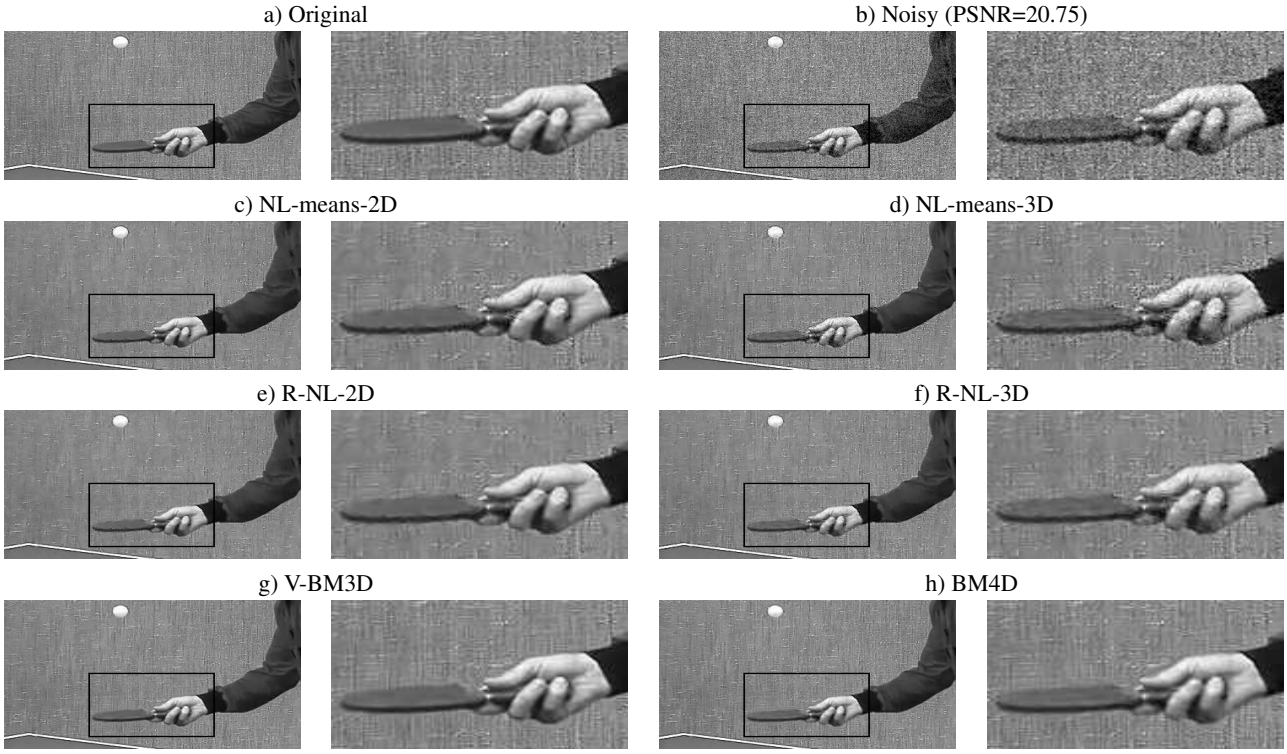

Fig. 1. Denoising of image sequences with 2D-patches using NL-means-2D, R-NL-2D and V-BM3D, and with 3D-patches using NL-means-3D,R-NL-3D and BM4D. We observe that the background of the image tends to be lost with NL-means-2D and R-NL-2D. NL-means-3D preserves the background but suffers from residual noise around the moving arm, which is corrected on R-NL-3D thanks to the adaptive TV regularization. R-NL-3D provides efficient denoising of textures and moving structures as opposed to traditional NL-means, without over-smoothing the background as in BM4D. More results are available at http://image.math.u-bordeaux1.fr/RNL.

	Target	Tennis	Bicycle
NL-means 2D [20]	29.91	29.31	30.77
R-NL-2D	30.00	29.65	31.72
V-BM3D [22]	30.21	29.79	32.92
NL-means 3D	32.02	29.06	29.30
R-NL-3D	*32.12*	*29.93*	*31.06*
BM4D [23]	**34.53**	**31.06**	**33.37**

Table 1. Mean PSNRs on denoised sequences with NL-means-2D, R-NL-2D and V-BM3D (2D patches), and NL-means-3D R-NL-3D and BM4D (3D patches).

	Target	Tennis (1-24)	Tennis (90-148)	Bicycle
NL-means 2D [20]	5.46	4.27	4.69	2.02
R-NL-2D	4.94	3.65	4.11	1.89
V-BM3D [22]	5.25	3.66	4.60	1.76
NL-means 3D	4.50	5.13	4.82	1.39
R-NL-3D	*3.98*	*3.56*	*3.91*	*1.29*
BM4D [23]	**3.67**	**2.78**	**3.14**	**0.94**

Table 2. Temporal standard deviation on denoised sequences with NL-means and R-NL using 2D patches, V-BM3D, NL-means and R-NL using 3D patches, and BM4D. Thanks to the use of 3D patches, R-NL-3D and BM4D provide the best temporal stability, and visual comfort.

tion should be close to zero. Table 2 displays the standard deviation computed on such constant areas on the denoised versions of the same three image sequences. R-NL3D offers a good performance regarding the quality and temporal stability of the denoised image sequences.

Conclusion

The proposed model offers efficient regularization of the NL-means thanks to the measure of confidence provided by the non-local weights. The adaptive regularization provides a good compromise between reduction of the residual noise and preservation of textures. We also show the superiority of 3-dimensional patches that guarantee better temporal consistency. The model has an intuitive interpretation and a fast implementation with linear complexity for the exponential family, and it could be adapted to different problems using other regularization terms. We have presented this model as a regularization of the NL-means but the philosophy of this method could apply to any non-local functional such as the improved versions of the NL-means or even BM3D.

5. REFERENCES

[1] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60(1):259-268, 1992.

[2] A. Buades, B. Coll, and J.-M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Modeling and Simulation*, vol. 4(2), pp. 490–530, Sep 2005.

[3] C.-A. Deledalle, L. Denis, and F. Tupin, "How to compare noisy patches? Patch similarity beyond Gaussian noise," *International Journal of Computer Vision*, vol. 99(1):86-102, 2012.

[4] M. Mahmoudi and G. Sapiro, "Fast image and video denoising via non-local means of similar neighborhoods," *IEEE Signal Processing Letters*, vol. 12(12), 2005.

[5] C.Kervrann and J. Boulanger, "Optimal spatial adaptation for patch-based image denoising," *IEEE Trans. Image Processing*, vol. 15(10):2866-2878, 2006.

[6] M. Van de Ville and D. Kocher, "Non-local means with dimensionality reduction and SURE-based parameter selection," *IEEE Transactions on image processing*, vol. 20(9):2683-2690, 2011.

[7] H. Talebi, X. Xhu, and P. Milanfar, "How to SAIF-ly boost denoising performance," *IEEE Transactions on image processing*, vol. 22(4):1470-1485, 2013.

[8] A. Buades, B. Coll, and J.-M. Morel, "Denoising image sequences does not require motion compensation," *Proc. IEEE Conf. Advanced Video and Signal based Surveillance*, pp. 70–74, 2005.

[9] Y. Wexler, E. Shechtman, and M. Irani, "Space-time completion of video," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 29(3):463-476, March 2007.

[10] S. Kindermann, S. Osher, and P. Jones, "Deblurring and denoising of images by nonlocal functionals," *SIAM Journal Multiscale Model. Simul.*, vol. 4(4):1091-1115, 2005.

[11] C. Louchet and L. Moisan, "Total variation as a local filter," *SIAM Journal on Imaging Sciences*, vol. 4(2):651-694, 2011.

[12] M. Protter, M. Elad, H. Takeda, and P. Milanfar, "Generalizing the non-local means to super-resolution reconstruction," *IEEE Transactions on image processing*, vol. 18(1):36-51, 2009.

[13] E. d'Angelo and P. Vandergheynst, "Fully non-local super-resolution via spectral hashing," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1137–1140, 2011.

[14] J. Darbon, A.Cunha, T.F. Chan, S. Osher, and G.J. Jensen, "Fast non-local filtering applied to electron cryomicroscopy," *IEEE Int. Symposium on Biomedical Imaging (ISBI)*, pp. 1331–1334, 2008.

[15] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision*, vol. 40(1):120-145, 2011.

[16] S. Anthoine, J.-F. Aujol, Y. Boursier, and C. Mélot, "Some proximal methods for poisson intensity CBCT and PET," *Inverse Problems and Imaging*, vol. 6(4):565-598, 2012.

[17] G. Aubert and J.-F. Aujol, "A variational approcah to remove multiplicative noise," *SIAM Journal on Applied Mathematics*, vol. 68(4):925-946, 2008.

[18] P.L. Combettes and V.R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Modeling and Simulation*, vol. 4(4):1168-1200, 2005.

[19] C. Sutour, C.-A. Deledalle, and J.-F. Aujol, "Adaptive regularization of the NL-means - application to image and video denoising," *IEEE Transactions on image processing, to appear*, 2013.

[20] A. Buades, B. Coll, and J.-M. Morel, "Nonlocal image and movie denoising," *International Journal of Computer Vision*, vol. 76(2), pp. 123–139, 2008.

[21] C. Liu and W. Freeman, "A high-quality video denoising algorithm based on reliable motion estimation," *European Conference on Computer Vision (ECCV)*, pp. 706–719, 2010.

[22] K. Dabov, A. Foi, and K. Egiazarian, "Video denoising by sparse 3D transform-domain collaborative filtering," in *Proc. 15th European Signal Processing Conference*. EUSIPCO 2007, Poznan, Poland, 2007.

[23] M. Maggioni, G. Boracchi, A. Foi, and K. Egiazarian, "Video denoising, deblocking and enhancement through separable 4-D nonlocal spatiotemporal transforms," *IEEE Transactions on image processing*, vol. 21(9):3952-3966, 2012.