

Video Denoising and Simplification via Discrete Regularization on Graphs

Mahmoud Ghoniem, Youssef Chahir, Abderrahim Elmoataz

► **To cite this version:**

Mahmoud Ghoniem, Youssef Chahir, Abderrahim Elmoataz. Video Denoising and Simplification via Discrete Regularization on Graphs. Proceeding of International Conference on Advanced Concepts for Intelligent Vision System, Oct 2008, Juan-les-Pins, France. pp.380-389, 10.1007/978-3-540-88458-3_34 . hal-00820663

HAL Id: hal-00820663

<https://hal.archives-ouvertes.fr/hal-00820663>

Submitted on 6 May 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Video Denoising and Simplification Via Discrete Regularization on Graphs

Mahmoud Ghoniem, Youssef Chahir, and Abderrahim Elmoataz

GREYC - CNRS UMR 6072

6, Boulevard du Marechal Juin 14050 CAEN CEDEX

{mghoniem, ychahir, aelmoataz}@greyc.ensicaen.fr

Abstract. In this paper, we present local and nonlocal algorithms for video denoising and simplification based on discrete regularization on graphs. The main difference between video and image denoising is the temporal redundancy in video sequences. Recent works in the literature showed that motion compensation is counter-productive for video denoising. Our algorithms do not require any motion estimation. In this paper, we consider a video sequence as a volume and not as a sequence of frames. Hence, we combine the contribution of temporal and spatial redundancies in order to obtain high quality results for videos. To enhance the denoising quality, we develop a robust method that benefits from local and nonlocal regularities within the video. We propose an optimized method that is faster than the nonlocal approach, while producing equally attractive results. The experimental results show the efficiency of our algorithms in terms of both Peak Signal to Noise Ratio and subjective visual quality.

1 Introduction

The last decade has witnessed an overwhelming proliferation of video sequences due to technological progress. These sequences are often of poor quality, because of the acquisition system, the compression algorithm or the transmission process that may have been utilized. For instance, low quality videos that have been captured with webcams or mobile phones are now ubiquitous. Video denoising aims at enhancing the quality of the video by reducing as much as possible the noise. The main challenge is to achieve the best trade-off between reducing noise and preserving significant structural elements.

Many methods have been proposed to denoise video sequences, and may be classified primarily based on the importance of motion estimation. Indeed, motion constitutes the fundamental challenge in video denoising. In this perspective, some methods involve a preliminary phase for motion estimation that is followed by the denoising scheme, whereas other approaches incorporate the motion estimation into the denoising algorithm with or without smoothing constraints.

However, recent works have shown that resolving the motion estimation not only is useless for the denoising problem but it also induces negative effects. Among these works, Protter et al. [1] based their approach on the diffusion of a dictionary. Also, Buades et al. [2] worked on nonlocal filtering and demonstrated how the aperture problem can be taken advantage of. A variant of nonlocal filtering was presented by Boulanger et al. [3] using patches of variable sizes.

These nonlocal filters can be viewed as a regularization based on nonlocal functionals. Kindermann et al. [4] were the first to interpret nonlocal means and neighborhood filters as regularization based on nonlocal functionals. Later, Gilboa and Osher have proposed a quadratic functional of weighted differences for image regularization and semi-supervised segmentation. Elmoataz et al. [5] recently introduced a nonlocal discrete regularization framework, which is the discrete analogue of the continuous Euclidean nonlocal regularization functionals by Gilboa and Osher [6]. This method is applicable to the images, meshes, manifold and data processing using weighted graphs of arbitrary topologies.

In this paper, we extend the nonlocal discrete regularization framework of Elmoataz et al. to the arena of video denoising.

In Section 2, we state the denoising problem in terms of regularization on graphs. Thereafter, the adaptation to videos is detailed in Section 3 where we propose a local / nonlocal algorithm and its optimized variant. We report and discuss in Section 4 the empirical results for our proposed methods for video denoising and simplification. In addition to the visual quality of our results, we utilize the Peak Signal to Noise Ratio (PSNR) as a performance measure.

2 Regularization on Weighted Graphs

2.1 Graph Derivatives

Let $G = (V, E)$ be an undirected graph, where $V = \{v_1, v_2, \dots, v_n\}$ is a finite set of vertices and $E \subseteq V \times V$ is a finite set of edges. Two vertices u and v are said to be adjacent if the edge $(u, v) \in E$. A graph is weighted if we associate to it a weight function $w : V \times V \rightarrow \mathbb{R}^+$, that satisfies the following conditions:

$$\forall u, v \in E, w(u, v) = w(v, u), w(u, v) > 0, \text{ if } u \neq v, w(u, u) = 0, \text{ otherwise.}$$

We consider now the derivative operators on graphs we need for regularization. Let $\mathcal{H}(V)$ be a Hilbert space of real-valued functions on vertices. A function $f : V \rightarrow \mathbb{R}$ in $\mathcal{H}(V)$ assigns a vector f_v to each vertex v in V . The *local variation of the weighted gradient operator* $\|\nabla_f\|$ of a function $f \in \mathcal{H}(V)$ at a vertex v is defined by:

$$\|\nabla_f(v)\| = \sqrt{\sum_{u \sim v} w(u, v)(f(v) - f(u))^2}.$$

This can be viewed as a measure of the regularity of a function around a vertex.

The *weighted p -Laplace operator*, with $p \in]0, +\infty[$, at a vertex v is defined on $\mathcal{H}(V)$ by:

$$(\Delta_p f)(v) = \frac{1}{p} \sum_{u \sim v} \gamma(u, v) (f(v) - f(u)) \text{ , where,}$$

$$\gamma(u, v) = w(u, v) (\|\nabla_f(v)\|^{p-2} + \|\nabla_f(u)\|^{p-2})$$

2.2 p -Laplace Regularization on Weighted Graphs

Consider a function f^0 that could be an image, a video or any discrete data set. This function is defined over the vertices V of a weighted graph $G_w = (V, E, w)$ by $f^0 : V \rightarrow \mathbb{R}$. f^0 is an observation of an original function f corrupted by a noise $n : f^0 = f + n$.

The discrete regularization of $f^0 \in \mathcal{H}(V)$ using the weighted p -Laplace operator consists in seeking a function $f^* \in \mathcal{H}(V)$ that is not only smooth enough on G_w , but also close enough to f^0 . It can be formalized by the minimization of two energy terms:

$$f^* = \min_{f \in \mathcal{H}(V)} \left\{ \frac{1}{p} \sum_{v \in V} \|\nabla f(v)\|^p + \frac{\lambda}{2} \|f - f^0\|_{\mathcal{H}(V)}^2 \right\} \quad (1)$$

where $p \in]0, +\infty[$ is the smoothness degree, λ is the fidelity parameter, called the Lagrange multiplier, which specifies the trade-off between the two competing terms, and ∇f represents the weighted gradient of the function f over the graph. The solution of problem (1) leads to a family of nonlinear filters, parameterized by the weight function, the degree of smoothness, and the fidelity parameter.

The first energy in (1) is the smoothness term or regularizer, whereas the second is the fitting term. Both energy functions in E_p are strictly convex functions of f . In particular, by standard arguments in convex analysis, Problem (1) has a unique solution, for $p \geq 1$, which satisfies:

$$(\Delta_p f(v)) + \lambda(f(v) - f^0(v)) = 0, \forall v \in V.$$

3 Video Processing Based on Video Regularization

To denoise video sequences, one can consider a video sequence as a simple sequence of independent frames and process each frame separately. Evidently, this would yield poor results from viewpoints of performance and denoising quality. A video sequence mainly differs from a sequence of images in that the consecutive frames of a video are usually related due to the temporal redundancy from one frame to another. Therefore, the temporal dimension is an essential feature that should be integrated in the denoising algorithm itself. The extension to video sequences is based on the integration of time into the regularization process. We will take advantage of the high temporal redundancy of the data due to the high frame rates, which enhances the quality of our regularization. Hence, we develop a spatiotemporal regularization on weighted graphs.

3.1 Proposed Algorithm

We consider the video sequence as a function f defined over the vertices of a weighted graph $G_{k_1, k_2, k_3} = (V, E, w)$, where $k_1, k_2, k_3 \in \mathbb{N}^3$. A vertex u is defined by a triplet (i, j, t) where (i, j) indicates the spatial position of the vertex and t , which is a frame number, indicates the temporal position of the vertex within the video sequence. We denote by $u \sim v$ a vertex u that belongs to the neighborhood of v which is defined as follows:

$$N_{k_1, k_2, k_3}(v) = \{u = (i', j', t') \in V : |i - i'| \leq k_1, |j - j'| \leq k_2, |t - t'| \leq k_3\}.$$

Similarly, we extend the definition of the patch to videos to obtain 3D patches. A patch around a vertex v is a box of size $r_x \times r_y \times r_t$, denoted by $B(v)$. Then, we associate to this patch a feature vector defined by:

$$F(f^0, v) = f^0(u), u \in B(v).$$

The weight function w associated to G_{k_1, k_2, k_3} provides a measure of the distance between its vertices that can simply incorporate local, semi-local or nonlocal features according to the topology of the graph and the image.

We consider the following two general weight functions:

$$w_L(u, v) = \exp\left(-\frac{|f(u) - f(v)|^2}{2\sigma_d^2}\right)$$

$$w_{NL}(u, v) = w_L(u, v) \cdot \exp\left(-\frac{\|F(f^0, u) - F(f^0, v)\|^2}{h^2}\right),$$

where σ_d^2 depends on the variations of $|f(u) - f(v)|$ over the graph. h can be estimated using the standard deviation depending on the variations of $\|F(f^0, u) - F(f^0, v)\|$ over the graph.

$w_L(u, v)$ is a measure of the difference between $f(u)$ and $f(v)$ values, and is used in the local approach of denoising. In addition to the difference between values, $w_{NL}(u, v)$ includes a similarity estimation of the compared features by measuring a \mathcal{L}^2 distance between the patches around u and v . It is the nonlocal approach.

To denoise video sequences, we use the Gauss-Jacobi iterative algorithm presented in [5]. For all (u, v) in E :

$$f^{(0)} = f^0$$

$$\gamma^{(k)}(u, v) = w(u, v) (\|\nabla f^{(k)}(v)\|^{p-2} + \|\nabla f^{(k)}(u)\|^{p-2}) \tag{2}$$

$$f^{(k+1)}(v) = \frac{p\lambda f^0(v) + \sum_{u \sim v} \gamma^{(k)}(u, v) f^{(k)}(u)}{p\lambda + \sum_{u \sim v} \gamma^{(k)}(u, v)}$$

where $\gamma^{(k)}$ is the function γ at the step k . The weights $w(u, v)$ are computed from f^0 , or can be given as an input.

At each iteration, the new value $f^{(k+1)}$, at a vertex v , depends on two quantities, the original value $f^0(v)$ and a weighted average of the existing values in a neighborhood of v . This shows that the proposed filter, obtained by iterating, is a low-pass filter which can accommodate many graph structures and weight functions. For more details on the Gauss-Jacobi iterative algorithm, the reader is referred to [5].

Observe that for the specific parameters $p = 2, \lambda = 0$ and after one iteration we retrieve the nonlocal means solution of Buades et al. [2].

3.2 Algorithm Optimization

The nonlocal method compares all the patches within a volume of size $k_1 \times k_2 \times k_3$ with the central patch of this volume, and is observed to be slow. In the sequel, we reduced

the processed data to $X\%$ and made an optimized nonlocal version. In our work, for each vertex v , we randomly select vertices within the 3D window, $W(v)$, centered in v . Hence, the optimized nonlocal algorithm is as follows:

Algorithm. For each vertex v of the graph:

1. **Construct $\mathcal{N}_W(v)$, the nonlocal neighborhood of v :** Select randomly $X\%$ of the vertices in $W(v)$ and out of the patch $B(v)$.
2. **Compute the patch similarity** between $B(v)$ and $B(u)$ for all $u \in \mathcal{N}_W$. All the patches $B(u)$ must be included in $W(v)$.
3. **Update the value $f(v)$** according to (2)

The procedure described above is iterated N times. Alternatively, we utilize a termination criterion and the algorithm converges in fewer iterations. As a consequence, we accelerate the nonlocal method, and obtain equally attractive results as demonstrated in Section 4.

4 Experimental Results

To test our algorithms, we considered the following parameters : $3 \times 3 \times 3$ patches, $7 \times 7 \times 3$ windows, $p = 2$, and $\lambda = 0$. We corrupted sequences with a synthetic zero-mean additive white Gaussian noise n having a variance σ^2 .

4.1 Video Denoising

As shown in Figure 1, the nonlocal method provides the best visual result. Noise is highly reduced while thin structures such as textures and fine details are well preserved.

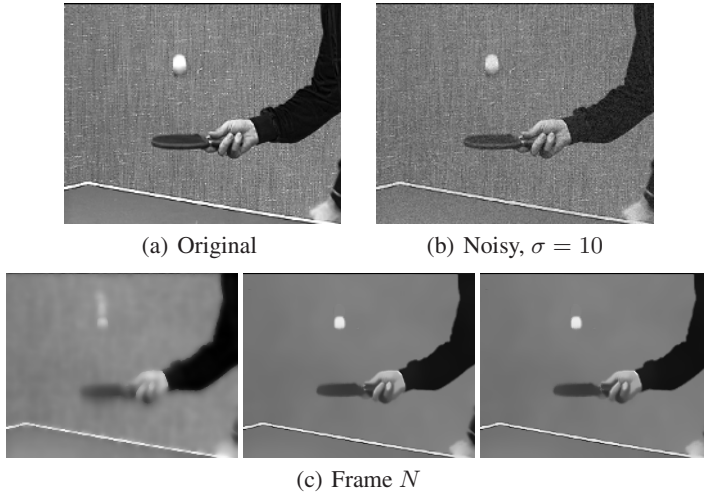
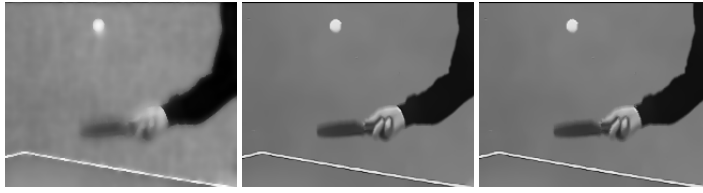


Fig. 1. Illustration of our three denoising methods with parameters $p = 2$, $\lambda = 0$ and $h = 80$ after ten iterations. (a) original image; (b) noisy image; (c), (d) and (e) : from left to right: local, nonlocal and optimized nonlocal results.

(d) Frame $N + 4$ (e) Frame $N + 8$ **Fig. 1.** (continued)**Table 1.** Comparison of the PSNR, expressed in dB, of noisy sequences (noise level $\sigma = 10$) denoised by our methods after one iteration

Sequence	Size	Input	Local	Nonlocal	Optimized nonlocal
Flower	$180 \times 144 \times 126$	22.96	23.82	25.15	25.87
Tennis	$216 \times 172 \times 126$	24.68	27.12	28.65	28.18
Football	$180 \times 144 \times 105$	24.68	25.14	26.02	26.68
Mobile	$180 \times 144 \times 251$	21.24	24.31	27.08	26.22



(a) Original

(b) Noisy, $\sigma = 10$ **Fig. 2.** Illustration of our three denoising methods with parameters $p = 2$, $\lambda = 0$ and $h = 30$ after ten iterations. (a) original image; (b) noisy image; from left to right: local, nonlocal and optimized nonlocal results.

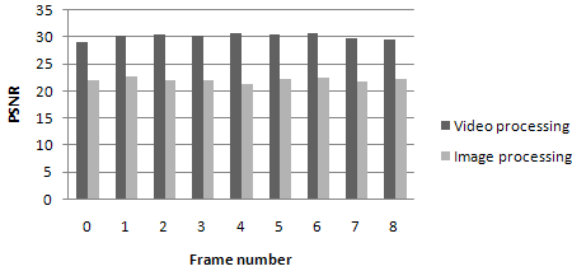


Fig. 3. PSNR comparison between video and image denoising results

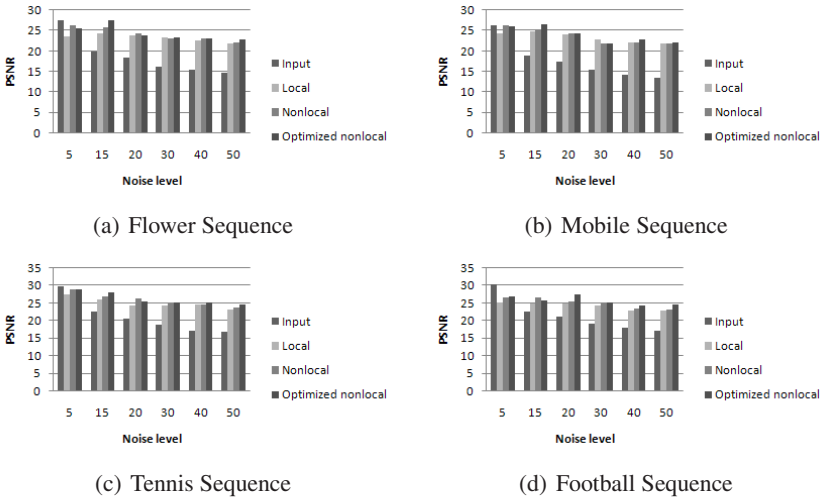


Fig. 4. PSNR comparison between our different denoising results and the input PSNR

The local method produces somewhat good results but some details are lost. Particularly, mobile objects such as the table tennis ball become very blurry with the local method whereas they are clear-cut in the results of the non local method. This test underlines the relevance of the nonlocal approach for denoising objects in motion. The optimized nonlocal method is very appealing as it preserves details and is faster than the nonlocal one.

The visual observations are further confirmed by the PSNR measures reported in Table 1. Although these results are slightly in favor of the nonlocal method, the proposed optimized variant produces very competitive results, using only 30% of the data, thereby achieving a faster computational performance. In fact, computational time is reduced by 70%. Nevertheless, to reduce the little quality loss, we work on ordering the patches within a window and selecting the best patches instead of a random selection. The first results are encouraging.

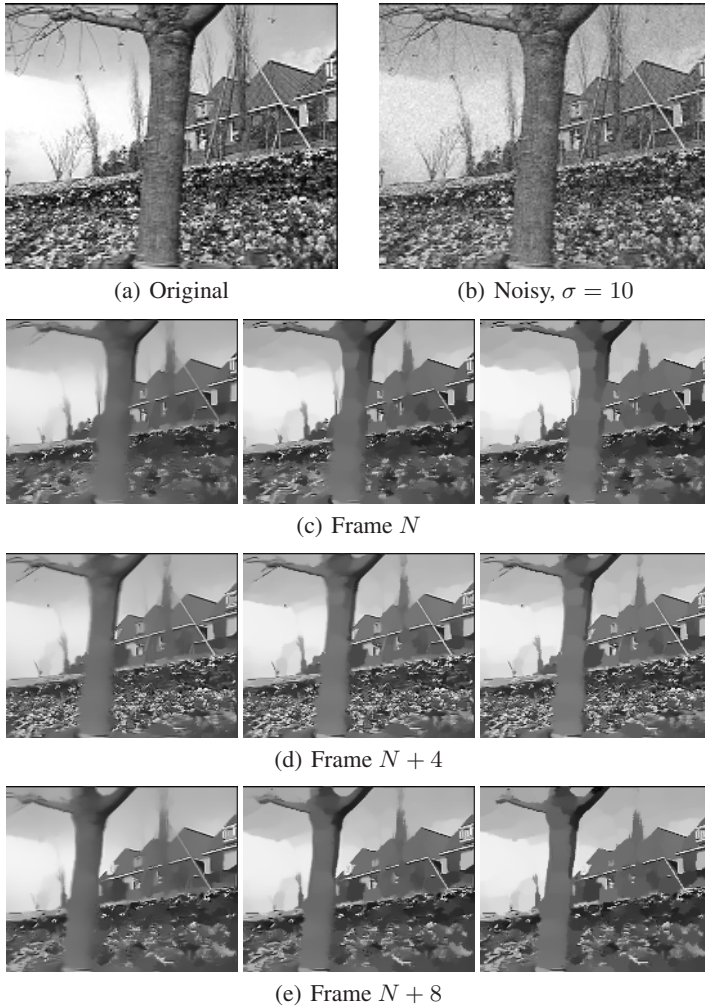


Fig. 5. Video simplification with the optimized nonlocal method, $\lambda = 0$ and $h = 80$ after ten iterations. From left to right: $p = 2$, $p = 0.5$ and $p = 0.1$.

To evaluate the quality of our video algorithm, we implemented a 2D version of our approach which corresponds to the algorithm by Elmoataz et al. [5], and applied it to denoise the frames of a video as single images.

Then we applied our video algorithm to the same video considered this time as a volume. We compared the 2D and 3D results. The PSNR of the video results are consistently higher than the PSNR of the 2D results as shown in Figure 3. It is observed that video processing with the optimized nonlocal method provides a higher quality denoising than image processing. This experiment evaluates and confirms the important contribution of temporal redundancy for video denoising. It is also interesting to evaluate the efficiency of our methods for different noise levels. We run a bunch of tests

on sequences corrupted by noises having σ from 5 to 50. The results are regrouped in Figure 4. These results show how the input PSNR decreases significantly when the noise level increases, contrary to the results of our three methods. The higher the noise level is the biggest the difference is between the input PSNR and our results' PSNR.

These measures also confirm that the optimized nonlocal method gives a similar PSNR and even a higher PSNR than the nonlocal one. In consequence, we consider that the optimized nonlocal method is a good trade-off between the fully nonlocal method and the local one from performance and quality viewpoints.

4.2 Video Simplification

Video simplification can be interpreted as a particular case of our algorithm for $p < 1$. We obtain homogeneous partitions of the video content. Figure 5 displays the results of the smoothing with $p = 0.5$ and $p = 0.1$. In comparison with the result with $p = 2$, we observe that similar regions join together and form bigger blocks. Moreover, we obtain a coarser simplification of the video when parameter p decreases as we can see for $p = 0.1$ and $p = 0.5$ (see Fig. 5). These results can be taken advantage of for video segmentation and visual object detection.

5 Conclusion

In this paper, a new algorithm has been introduced for the denoising and simplification of video sequences. The algorithm does not require any motion estimation and is based on discrete regularization on graphs. We view a video sequence as a volume to benefit from temporal redundancy, which enhances the denoising results as demonstrated by our experiments.

We take advantage of local and nonlocal regularities in our methods to reduce noise while preserving significant features and details in the video. As the nonlocal method is time-consuming, we reduced the amount of processed data (by 70% in our tests) without compromising the denoising quality.

Our future research will investigate the ordering of patches within a window in order to choose the most similar patches in lieu of randomly selected ones. As a direct application of our algorithm, we shall also employ this work for inpainting purposes.

References

- [1] Protter, M., Elad, M.: Sparse and redundant representations and motion-estimation-free algorithm for video denoising. In: Wavelets XII. Proceedings of the SPIE, vol. 6701, p. 43 (2007)
- [2] Buades, A., Coll, B., Morel, J.: Denoising image sequences does not require motion estimation. In: Advanced Video and Signal Based Surveillance, 2005. IEEE Conference on AVSS 2005, pp. 70–74 (2005)
- [3] Boulanger, J., Kervrann, C., Bouthemy, P.: Space-time adaptation for patch-based image sequence restoration. IEEE Trans. PAMI 29, 1096–1102 (2007)

- [4] Kindermann, S., Osher, S., Jones, P.W.: Deblurring and denoising of images by nonlocal functionals. *Multiscale Modeling & Simulation* 4, 1091–1115 (2005)
- [5] Elmoataz, A., Lezoray, O., Bogleux, S.: Nonlocal discrete regularization on weighted graphs: a framework for image and manifold processing. *IEEE Trans. Image Processing* (accepted) (to appear, 2008)
- [6] Gilboa, G., Osher, S.: Nonlocal linear image regularization and supervised segmentation. *Multiscale Modeling & Simulation* 6, 595–630 (2007)