Denoising image sequences does not require motion estimation

A. Buades, B. Coll

Dpt. Matemàtiques i Informàtica
Universitat Illes Balears
07122 Palma de Mallorca, Spain

J.M. Morel

CMLA
ENS Cachan
94235 Cachan, France

Abstract

State of the art movie restoration methods either estimate motion and filter out the trajectories, or compensate the motion by an optical flow estimate and then filter out the compensated movie. Now, the motion estimation problem is ill-posed. This fact is known as the aperture problem*: trajectories are ambiguous since they could coincide with any promenade in the space-time isophote surface. In this paper, we try to show that, for denoising, the aperture problem can be taken advantage of. Indeed, by the aperture problem, many pixels in the neighboring frames are similar to the current pixel one wishes to denoise. Thus, denoising by an averaging process can use many more pixels than just the ones on a single trajectory. This observation leads to use for movies a recently introduced image denoising method, the NL-means algorithm. This static 3D algorithm outperforms motion compensated algorithms, as it does not lose movie details. It involves the whole movie isophote and not just a trajectory.*

1. Introduction

What makes particularly difficult the sequence restoration is the movement. If an image sequence is completely static, then a simple temporal averaging would be an excellent estimation. Unfortunately, if some region of the scene moves, then the temporal averaging will blur it. For this reason, most filtering techniques are adapted to the dynamic character of image sequences, see [1] for a complete review.

Nearly all the recently proposed image sequence filters are motion compensated. Samy [14] and Sezan et al. [15] proposed the spatiotemporal linear minimum mean square error filter (LMMSE) which is a motion compensation of the filter proposed by Lee [9]. Ozkan et al [13] proposed the Adaptive Weighted Averaging (AWA) filter which is in fact a motion compensation of a neighborhood or sigma filter [10, 17]. Motion compensated Wiener filters were also proposed by Kokaram [8]. Huang [7] and Martinez [12] implemented motion compensated median filters. It is assumed that, in order to deal with the dynamic character of sequences and to obtain high quality results, a motion estimation is necessary. However, the motion estimation is a very difficult problem itself and there is no algorithm able to give a final solution. In fact, the motion compensation can propose inaccurate trajectories and a filtering along these trajectories can lead to blur and information loss.

We shall sustain the position that, in fact, motion estimation is not only unnecessary, but probably counterproductive. The aperture problem, viewed as a general phenomenon in movies, can be positively interpreted in the following way : there are many pixels in the next or previous frame which can match the current pixel, see Figure 2. Now by the law of large numbers, a denoising method works so much the better if the estimation of the current pixel is based on many, and not just a few pixels. Thus, it seems sound to use not just one trajectory, but rather all possible trajectories to estimate the value of a current pixel. We shall actually not push in the direction of a multiplicity of trajectories. We shall give up any such notion and just talk about similar pixels to the current pixel across time and space.

We shall experimentally prove that a recently introduced nontrivial extension of the sigma-filter, the NL-means algorithm [2, 3], defines a generalized isophote on which noise reduction by averaging gives near optimal results. In particular this *filter is no more improvable by motion compensation*.

The plan of this paper derives from the above remarks. In section 2 we introduce the NL-means algorithm and its extension to image sequence filtering. In section 3 we experimentally show that the NL-means algorithm averages the more similar pixels even when they move from a frame to another. Finally, in section 4 we compare the NL-means with motion compensated algorithms.

2. The NL-means algorithm

The NL-means is a recently introduced image denoising algorithm [2, 3, 4]. This algorithm tries to take advantage of the high degree of redundancy of any natural image. By this, we simply mean that every small window in a natural image has many similar windows in the same image. Now,

0-7803-9385-6/05/$20.00 ©2005 IEEE

Figure 1: Three consecutive frames of a degraded image sequence. The sparse time sampling in film sequences makes restoration more difficult than in 3D images.

in a very general sense inspired by the neighborhood filters [10, 17], one can define as "neighborhood of a pixel i" any set of pixels j in the image such that a window around j looks like a window around i. All pixels in that neighborhood can be used for predicting the value at i, as was first shown in Efros et al. [5].

Let us take first u to be a single image defined on a bounded domain $\Omega \subset \mathbb{R}^2$. The NL-means algorithm is defined as

$$NLu(\mathbf{x}) = \frac{1}{C(\mathbf{x})} \int_\Omega e^{-\frac{(G_a * |u(\mathbf{x}+.)-u(\mathbf{y}+.)|^2)(0)}{h^2}} u(\mathbf{y})\, d\mathbf{y},$$

where $\mathbf{x} \in \Omega$, G_a is a Gaussian kernel of standard deviation a, h acts as a filtering parameter and $C(\mathbf{x}) = \int_\Omega e^{-\frac{(G_a * |u(\mathbf{x}+.)-u(\mathbf{z}+.)|^2)(0)}{h^2}} d\mathbf{z}$ is the normalizing factor. This amounts to say that $NLu(\mathbf{x})$, the denoised value at $\mathbf{x}$, is a weighted average of the values of all pixels whose Gaussian neighborhood looks like the neighborhood of $\mathbf{x}$.

The NL-means is a non-local algorithm since all pixels in the image are used for the estimation at a pixel $\mathbf{x}$. NL-means take advantage of the huge redundancy present in natural images. As this redundancy is even larger in image sequences, it seems obvious to directly extend the $2D$ support to a $3D$ spatiotemporal one.

2.1. Sequence filtering algorithm

Every detail or small window has many similar windows through the sequence. However, the comparison of a window with all possible windows of the sequence is a prohibitive amount of computation. For this reason, we reduce the support of the NL-means algorithm, at a pixel $\mathbf{x} = (i_0, j_0, t_0)$, to

$$S_\mathbf{x} = \{(i,j,t) \mid |i - i_0| \leq \delta_i,\ |j - j_0| \leq \delta_j,\ |t - t_0| \leq \delta_t\},$$

where $\delta_i, \delta_j, \delta_t > 0$. The estimated value $NL(u)(\mathbf{x})$ is computed as a weighted average of all the pixels in the support of $\mathbf{x}$,

$$NLu(\mathbf{x}) = \frac{1}{C(\mathbf{x})} \sum_{\mathbf{y} \in S_\mathbf{x}} w(\mathbf{x},\mathbf{y}) u(\mathbf{y}), \qquad (1)$$

where the weights $w(\mathbf{x},\mathbf{y}) \geq 0$ depend on the similarity between the pixels $\mathbf{x}$ and $\mathbf{y}$ and $C(\mathbf{x})$ is the normalizing factor.

The similarity between pixels $\mathbf{x}$ and $\mathbf{y}$ depends upon the similarity of the intensity gray level vectors $u(\mathcal{N}_\mathbf{x})$ and $u(\mathcal{N}_\mathbf{y})$, where $\mathcal{N}_\mathbf{x}$ denotes a two dimensional square neighborhood of fixed size and centered at the pixel $\mathbf{x}$.

In order to compute the similarity of the intensity gray level vectors $u(\mathcal{N}_\mathbf{x})$ and $u(\mathcal{N}_\mathbf{y})$, we compute a Gaussian weighted Euclidean distance, $\|u(\mathcal{N}_\mathbf{x}) - u(\mathcal{N}_\mathbf{y})\|^2_{2,a}$,

$$\|u(\mathcal{N}_\mathbf{x}) - u(\mathcal{N}_\mathbf{y})\|^2_{2,a} = \sum_{\mathbf{z} \in Q} G_a(\mathbf{z}) |u(\mathbf{x}+\mathbf{z}) - u(\mathbf{y}+\mathbf{z})|^2,$$

where Q denotes a square of fixed size centered in $(0,0)$ and G_a a two dimensional Gaussian kernel of standard deviation a. Efros and Leung [5] showed that the L^2 distance is a reliable measure for the comparison of image windows in a texture patch. Now, this measure is so much the more adapted to any additive white noise as such a noise alters the distance between windows in a uniform way. Indeed,

$$E||u(\mathcal{N}_\mathbf{x}) - u(\mathcal{N}_\mathbf{y})||^2_{2,a} = ||u_0(\mathcal{N}_\mathbf{x}) - u_0(\mathcal{N}_\mathbf{y})||^2_{2,a} + 2\sigma^2$$

where the observed sequence u is supposed to be obtained by the addition of a signal independent white noise of standard deviation σ to the true sequence u_0. This equality shows that, in expectation, the Euclidean distance preserves the order of similarity between pixels. So the most similar pixels to $\mathbf{x}$ in u also are expected to be the most similar pixels to $\mathbf{x}$ in u_0. The weights associated with the quadratic distances are defined by

$$w(\mathbf{x},\mathbf{y}) = e^{-\frac{||u(\mathcal{N}_\mathbf{x}) - u(\mathcal{N}_\mathbf{y})||^2_{2,a}}{h^2}},$$

where h controls the decay of the exponential function and therefore the decay of the weights as a function of the Euclidean distances. Notice that the weights $w(\mathbf{x},\mathbf{y})/C(\mathbf{x})$ can be interpreted as a probability distribution as their sum is 1 and they are positive. The number $w(\mathbf{x},\mathbf{y})/C(\mathbf{x})$ can be named probability that $\mathbf{y}$ looks like $\mathbf{x}$.

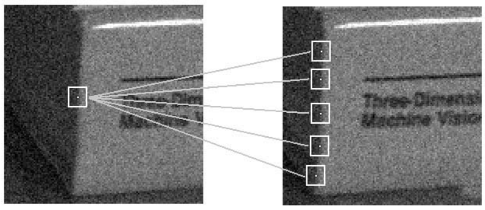

Figure 2: Aperture problem: The ambiguity of trajectories is the most difficult problem in motion estimation. Many good candidates are possible and the motion estimation algorithms must choose one.

3. Motion adaptation and the aperture problem.

The NL-means algorithm applied to movies is a static filter, that is, a straightforward extension of a two dimensional filter. It does not directly take into account the dynamic character of image sequences. The aim of this section is to experimentally show that motion compensation is not necessary and even counterproductive.

Motion estimation algorithms try to solve the aperture problem. The block matching algorithms choose the pixel with the more similar configuration, thus losing many other interesting possibilities, as displayed in Figure 2. Algorithms based on the "Optical Flow Constraint" [6, 16] must impose a regularity condition of the flow field in order to choose a single trajectory. Thus, the motion estimation algorithms are forced to choose a candidate among all possible equally good choices. However, when dealing with sequence restoration, the redundancy is not a problem but an advantage. Figure 2 shows that we could choose anyone of the possible candidates for the averaging, so why not take them all.

In Figure 3, we display the probability distributions computed by the NL-means. We display the support and the probability distribution $w(\mathbf{x},\mathbf{y})/C(\mathbf{x})$ used to estimate the central pixel $\mathbf{x}$ of the middle frame (in white). The algorithm favors pixels with a similar local configuration even if they are far away from the reference pixel. As the similar configurations move, so do the weights. Thus, the algorithm is able to follow the similar configurations when they move but without an explicit motion computation. So, there is no need to solve any aperture problem, we just average all pixels with a similar local configuration. Notice that, in contrast to motion compensated methods, other pixels of the same frame can be equally used to denoise a pixel of the current frame.

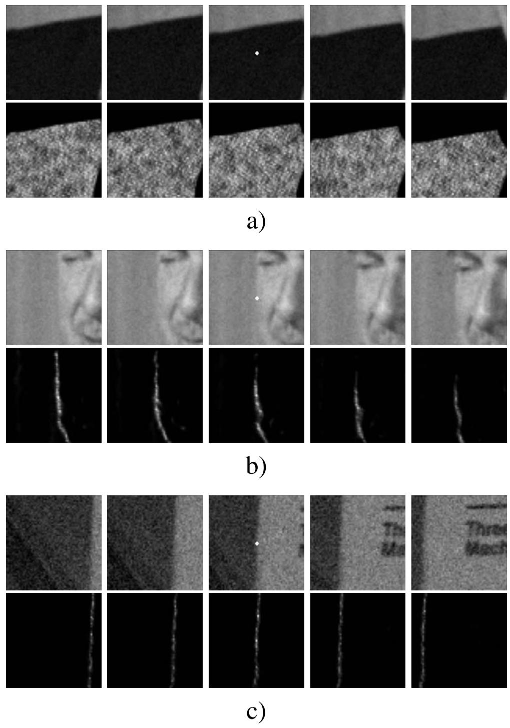

Figure 3: Display of the probability distributions given by the NL-means algorithm. We display the original noisy values of the support and the weight distribution $w(\mathbf{x},\mathbf{y})/C(\mathbf{x})$ used to estimate the central pixel $\mathbf{x}$ (in white) of the middle frame. The weight distribution varies from zero (black) to one (white). The weight configuration is spatially adapted to the local configuration of each frame. The algorithm looks for the pixels with a more similar configuration even they have moved. This algorithm is adapted to moving pictures without the need of an explicit motion estimation.

4. Discussion and comparison

We compare the NL-means algorithm with two motion compensated algorithms: the LMMSE [14, 15] and the

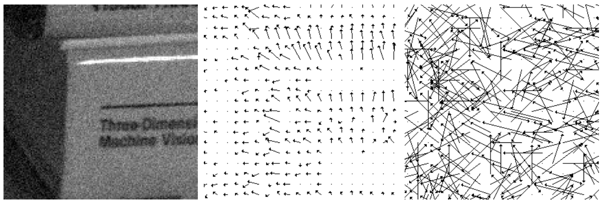

Figure 4: Comparison of the optical flow. Left: Small part of a frame of the sequence in Figure 1. Middle: Weickert and Schnorr algorithm [16]. Right: block matching algorithm. The regularity condition of the OFC based methods yields a smooth optical flow. The non-presence of this regularity term in the block matching can lead to a chaotic optical flow field. Indeed, by the aperture problem, trajectories on isophotes are essentially ambiguous and therefore the choice performed by block matching close to a random walk.

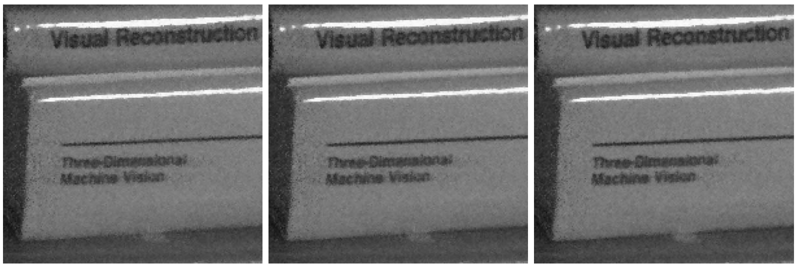

Figure 5: Comparison experiment. One frame extracted from the sequence filtered by the neighborhood filter (left image), the motion compensated neighborhood filter by using WS (middle image) and the motion compensated neighborhood filter by using BM (right image). Both compensated motion estimations improve the static version of the filter. The BM algorithm better preserves the details and creates less blur.

AWA [13]. The motion estimation is obtained by using a block matching algorithm and the OFC based method, the Weickert and Schnorr algorithm [16].

In Figure 5 we compare the neighborhood filter and its motion compensated version (AWA) by both motion estimation algorithms. Both compensated motion estimations improve the static version of the filter. In spite of its chaotic optical flow (Figure 4), the BM algorithm better preserves the details and creates less blur than the filtered sequence using the WS algorithm. When estimating the motion on a noisy sequence, the original grey level values have been modified and therefore the (OFC) is no more valid. The block comparison is more robust and able to look for the more similar configurations even though the sequence is noisy.

In Figure 6 we compare the NL-means algorithm with the motion compensated algorithms. The motion estimate has been obtained using the block matching algorithm. We display a filtered image of the sequence and the noise removed by the three algorithms. Ideally, the removed noise should not contain any noticeable structure and it should look like the realization of a white noise. The LMMSE and the AWA present many structures on their residual noise. This implies that these structures have been removed from the original sequence. The NL-means residual noise does not present any noticeable structure. As a consequence, the filtered image has kept more details and it is less blurred.

5. Conclusion

In this paper, an extension of the NL-means algorithm for image sequence filtering, was proposed. This approach avoids any motion estimation computation, which is ill-posed.

The experimentation showed that the algorithm is able to denoise image sequences while preserving the main features and the details. The algorithm improves previous methods as the removed noise does not present any noticeable structure.

References

[1] J. C. Brailean, R. P. Kleihorst, S. Efsratiadis, A. K. Katsaggelos, and R. L. Lagendijk, "Noise reduction filters for dynamic image sequences: a review," *Proceedings of the IEEE*, vol. 83, pp. 1272-1292, 1995.

[2] A. Buades, B. Coll and J.M. Morel, "A review of image denosing methods, with a new one", to appear in *Multiscale Modeling and Simulation (SIAM)*. Preprint avalaible at http://www.cmla.ens-cachan.fr/Cmla/Publications/2004.

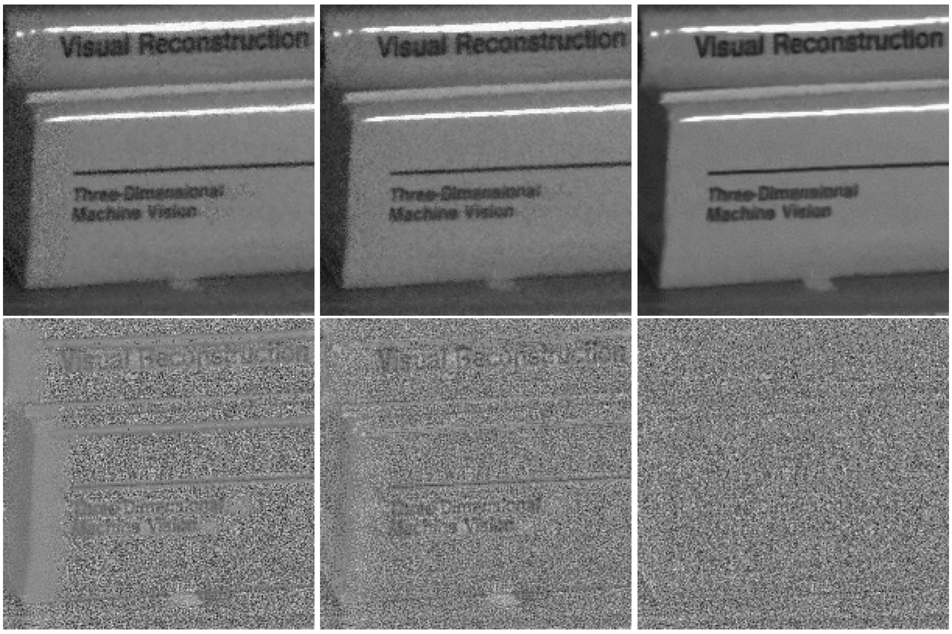

Figure 6: Comparison experiment with the sequence of Figure 1. Top: the same filtered frame obtained by the LMMSE (left), the AWA (middle) and the NL-means (right). The motion estimation has been obtained by the block matching algorithm. Bottom: the residual noise from the three algorithms above. The LMMSE residual noise is nearly zero on the strong boundaries. Now, these boundaries are kept noisy on the filtered sequence. We can read the titles of the books on the residual noise of the AWA. Therefore, that much information has been removed from the original. Finally, the NL-means algorithm does not have any noticeable structure in its residual noise. As a consequence, the filtered sequence has kept more details and is less blurred.

[3] A. Buades, B. Coll and J.M. Morel, "A non-local algorithm for image denoising", *IEEE International Conference on Computer Vision and Pattern Recognition*, 2005.

[4] A. Buades, B. Coll and J.-M. Morel, "Image data process by image noise reduction and camera integrating the means for implementing this process", European patent pending, May 5, 2004.

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

[6] B. Horn and B. Schunck, "Determining optical flow," *Artif. Intell.*, Vol. 17, pp. 185-203, 1981.

[7] T. Huang, *Image Sequence Analysis*, Springer-Verlag, 1981.

[8] A. C. Kokaram, "Motion Picture Restoration", PhD thesis, Cambridge University, 1993.

[9] J.S. Lee, "Digital image enhancement and noise filtering by use of local statistics", *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 2, pp. 165-168, 1980.

[10] J.S. Lee, "Digital image smoothing and the sigma filter", *Computer Vision, Graphics and Image Processing*, vol. 24, pp. 255-269, 1983.

[11] E. Levina, *Statistical Issues in Texture Analysis.* PhD dissertation, UC Berkeley, 2002.

[12] D. M. Martinez, "Model-based motion estimation and its application to restoration and interpolation of motion pictures", PhD thesis, Massachusetts Institute of Technology, 1986.

[13] M.K. Ozkan, M.I. Sezan, and A.M. Tekalp, "Adaptive motion compensated filtering of noisy image sequences", *IEEE Trans. Circuits and Systems for Video Technology*, vol. 3, pp. 277-290, 1993.

[14] R. Samy, "An adaptive image sequence filtering scheme based on motion detection", *SPIE*, Vol. 596, pp. 135-144, 1985.

[15] M.I. Sezan, M.K. Ozkan and S.V. Fogel, "Temporally adaptive filtering of noisy sequences using a robust motion estimation algorithm", *Proceedings of the Int. Conf. Acoustics, Speech, Signal Processing 91*, pp. 2429-2432, 1991.

[16] J. Weickert and C. Schnörr, "Variational optic flow computation with a spatio-temporal smoothness constraint", *Journal of Mathematical Imaging and Vision*, vol. 14, pp. 245-255, 2001.

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