Space-Time Adaptation for Patch-Based Image Sequence Restoration

Jérôme Boulanger, Charles Kervrann, Patrick Bouthemy

I R I S A

INSTITUT DE RECHERCHE EN INFORMATIQUE ET SYSTÈMES ALÉATOIRES

PUBLICATION INTERNE

N^o 1798

SPACE-TIME ADAPTATION FOR PATCH-BASED IMAGE SEQUENCE RESTORATION

JÉRÔME BOULANGER , CHARLES KERVRANN AND PATRICK BOUTHEMY

ISSN 1166-8687

IRISA

CAMPUS UNIVERSITAIRE DE BEAULIEU - 35042 RENNES CEDEX - FRANCE

INSTITUT DE RECHERCHE EN INFORMATIQUE ET SYSTÈMES ALÉATOIRES
Campus de Beaulieu – 35042 Rennes Cedex – France
Tél. : (33) 02 99 84 71 00 – Fax : (33) 02 99 84 71 71
http://www.irisa.fr

Space-Time Adaptation for Patch-Based Image Sequence Restoration

Jérôme Boulanger *, Charles Kervrann ** and Patrick Bouthemy ***

Systèmes cognitifs
Projet Vista

Publication interne n° 1798 — Mai 2006 — 17 pages

Abstract: We present a novel space-time patch-based method for image sequence restoration. We propose an adaptive statistical estimation framework based on the local analysis of the bias-variance trade-off. At each pixel, the space-time neighborhood is adapted to improve the performance of the proposed patch-based estimator. The proposed method is unsupervised and requires no motion estimation. Nevertheless, it can also be combined with motion estimation to cope with very large displacements due to camera motion. Experiments show that this method is able to drastically improve the quality of highly corrupted image sequences. Quantitative evaluations on standard artificially noise-corrupted image sequences demonstrate that our method outperforms other recent competitive methods. We also report convincing results on real noisy image sequences.

Key-words: Image sequence restoration, denoising, non parametric estimation, non linear filtering, bias-variance trade-off

(Résumé : tsvp)

The authors would thank the project ACI IMPBio for its support.

* jerome.boulanger@irisa.fr
** charles.kervrann@irisa.fr
*** patrick.bouthemy@irisa.fr

Centre National de la Recherche Scientifique
(UMR 6074) Université de Rennes 1 – Insa de Rennes

Institut National de Recherche en Informatique
et en Automatique – unité de recherche de Rennes

Une méthode adaptive spatio-temporelle pour la restauration de séquences d'images video

Résumé : Nous présentons une nouvelle méthode pour la restauration de séquences d'images exploitant des "patches". Nous proposons une approche d'estimation adaptative s'appuyant sur l'analyse locale du compromis biais-variance. Pour chaque pixel, le voisinage spatio-temporel considéré est adapté afin d'améliorer les performances de l'estimateur proposé. Cette méthode est non-supervisée et ne requiert pas d'estimation du mouvement. Cependant, elle peut aussi être combinée avec une estimation du mouvement afin de faire face, par exemple, aux grands déplacements dus aux mouvements de la caméra. Les expériences montrent que cette méthode est capable d'améliorer significativement la qualité de séquences d'image fortement bruitées. Des évaluations quantitatives sur des séquences artificiellement bruitées selon un protocole standard démontrent que notre méthode surpasse d'autres méthodes récentes. Nous apportons également des résultats convaincants sur des séquences d'images réelles.

Mots clés : Restauration de séquences d'images, débruitage, estimation non paramétrique, filtrage non linéaire, compromis biais-variance

1 Introduction

Image sequence restoration takes a crucial place in several important application domains. Infra-red imaging, confocal microscopy, ultra-sound and X-ray imaging are known to be noisy modalities. Restoring old films or videos is also of key importance for cultural heritage preservation. Image sequence restoration is then widely studied and still an active field of research. The main difficulties arise from non-stationarities observed in the spatio-temporal signals. Denoising or restoration methods must preserve space-time discontinuities while minimizing the error between the unknown original noise-free image sequence and the denoised sequence.

A review of image sequence restoration methods can be found in [4]. These methods, especially designed for image sequences, take into account temporal correlation between images, and some of them involve a motion compensation/detection stage [7, 10, 28]. Other image sequence restoration methods can be exported from the still image denoising domain (see [6] for a recent review). Thus, wavelet shrinkage [13, 24], Wiener filtering [9] or PDE-based methods [18] have been extended to process image sequences. However, other methods like Total Variation minimization [25] cannot be easily extended to space-time domain. Recently an extension of the non-local mean filter [6] also related to the universal denoising (DUDE) algorithm [21], has been proposed to process image sequences and relies on the principle that the image sequence contains repeated patterns [5] by averaging. The detection of repeated patterns can be used to reduce the noise in images. Such an approach is already popular in texture synthesis [11], inpainting [8], video completion [26] and has also been explored for image restoration [27]. Nevertheless, searching similar examples in the whole image sequence is infeasible in practice. Accordingly, a variant of the non-local mean filter has been recently proposed in [20] and use a pre-classification of the pixels in the sequence in order to speed up the denoising procedure.

The original method we propose is a space-time patch-based adaptive statistical method for image sequence restoration. A preliminary version has been described in [3]. It is related to the statistical framework described for still images [15–17, 23] and image sequences [12]. Unlike robust anisotropic diffusion [2] and non-linear Gaussian filtering [1], our approach supplies scale selection by estimating the appropriate space-time filtering window at each pixel. Moreover, the proposed method differentiates the space and time dimensions unlike other methods that consider the image sequence as an isotropic $3D$ volume [18, 24]. As a matter of fact, a naive approach can introduce motion blur and artifacts if the time dimension is merged with spatial dimension. In contrast to [12, 15], our approach is not based on a geometrical partition of the neighborhood in sectors. It uses a fixed neighborhood geometry but involves an appropriate and more flexible weighted sum of data points in an adaptive neighborhood which is far more flexible and efficient. The weights are defined by computing a distance between a patch centered at the considered pixel x_i and patches taken in an adapted space-time neighborhood. Additionally, a confidence level (i.e. , estimate variance) attached to each restored pixel is provided.

The remainder of the paper is organized as follows. Section 2 describes our general framework for image sequence restoration. In Section 2.2, the adaptive choice of the local

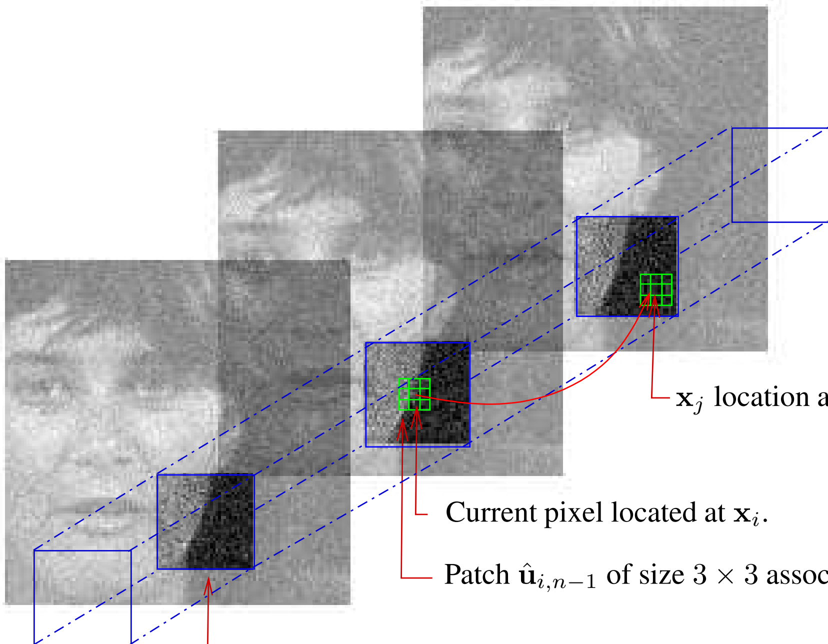


Figure 1: Patch-based space-time framework. To each point of the image sequence is associated an estimated space-time neighborhood $W_{i,n}$. To each point $\mathbf{x}_j$ of this neighborhood is associated a 3×3 patch. Every weight ω_{ij} is defined as a function of the distance between the patch centered in point $\mathbf{x}_i$ and the patch centered in point $\mathbf{x}_j \in W_{i,n}$.

space-time neighborhood is introduced. Section 2.3 deals with the similarity measure involved in the selection of patches in the space-time neighborhood, and Section 3 gives details of the algorithm implementation. In Section 4, we report an important set of experimental results on artificially noise-corrupted video sequences as well as on a real noisy IR image sequence. Intensive comparison with other recent methods has been carried out, demonstrating that our method outperforms them. We also present how our denoising method can be combined with a motion estimation method if required. Finally, Section 5 contains concluding remarks.

2 Patch-based space-time approach

2.1 Model description

We consider the following statistical image model:

$$Y_i = u(\mathbf{x}_i) + \xi_i, \tag{1}$$

where $\mathbf{x}_i \in \Omega$ denotes the pixel location in the space-time volume $\Omega \subset \mathbb{R}^3$. The function $u_i = u(\mathbf{x}_i)$ is the ideal image to be recovered from observations Y_i. The errors ξ_i are assumed to be independent zero-mean Gaussian variables with unknown variance τ^2.

We need minimal assumptions on the structure of the image for recovering u. In what follows, we assume that there exists repeated small image patches of the patch centered at $\mathbf{x}_i$ in the space-time neighborhood of pixel $\mathbf{x}_i$. However, the size and shape of this neighborhood will vary in the image sequence because of non-stationarities and presence of discontinuities. If we can determine the adequate neighborhood for each pixel, then the regression function u can be estimated by optimizing a local maximum likelihood (ML) criterion. The proposed method addresses these two issues as described below.

We design a sequence of increasing nested space-time neighborhoods $\{W_{i,n}\}_{n\in[0:N]}$, centered at each point $\mathbf{x}_i$, *i.e.*, $W_{i,n} \subset W_{i,n+1}$, with N denoting the largest neighborhood. Additional details about the neighborhood design are given in Section 2.2. As for initialization, we choose the smallest neighborhood (the 8 nearest neighbors in the $2D$ space domain) as the *pilot* (starting) neighborhood $W_{i,0}$ at $\mathbf{x}_i$. Then, we compute an initial estimate $\hat{u}_{i,0}$ of $u(\mathbf{x}_i)$ and its associated variance $\hat{\sigma}^2_{i,0}$ as

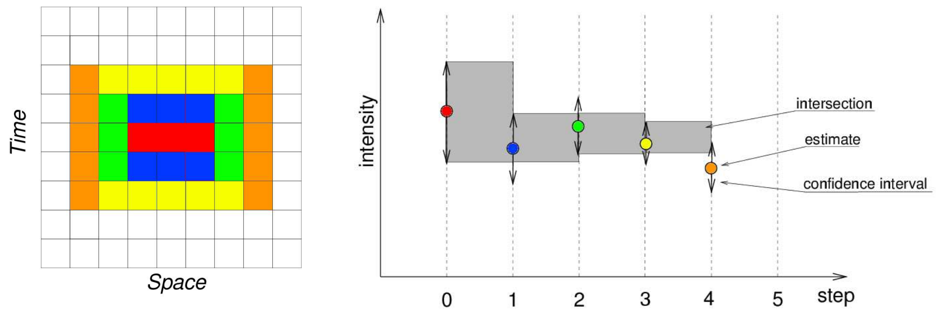
$$\hat{u}_{i,0} = \sum_{\mathbf{x}_j \in W_{i,0}} \omega_{ij} Y_j \quad \text{and} \quad \hat{\sigma}^2_{i,0} = \hat{\tau}^2 \sum_{\mathbf{x}_j \in W_{i,0}} \omega^2_{ij} \tag{2}$$

where $\hat{\tau}^2$ is an empirical estimate of the noise variance τ^2 as described in Section 3. At the initialization step, the weights ω_{ij} are defined as a function of the distance between two spatial $p \times p$ image patches or space-time $p \times p \times q$ image patches centered at $\mathbf{x}_i$ and $\mathbf{x}_j$ respectively. There is no real difference for the proposed method between spatial only and space-time patch since the spatial only patch can be considered as a space-time patch with the temporal dimension q equal to one.

At the first iteration, we consider a larger space-time neighborhood $W_{i,1}$ such that $W_{i,0} \subset W_{i,1}$ and calculate new estimates $\hat{u}_{i,1}$ and $\hat{\sigma}^2_{i,1}$ over $W_{i,1}$. We continue this way, and at iteration n, we define the estimator as

$$\hat{u}_{i,n} = \sum_{\mathbf{x}_j \in W_{i,n}} \omega_{ij} Y_j \quad \text{and} \quad \hat{\sigma}^2_{i,n} = \hat{\tau}^2 \sum_{\mathbf{x}_j \in W_{i,n}} \omega^2_{ij} \tag{3}$$

where the estimator $\hat{u}_{i,n}$ corresponds to a weighted average of the intensities located in the space-time neighborhood. We propose to define weights ω_{ij} as a function of the distance between two image patches $\hat{\mathbf{u}}_{i,n-1}$ and $\hat{\mathbf{u}}_{j,n-1}$ estimated at iteration $n-1$ and centered in



(a) (b)

Figure 2: (a) Spatio-temporal neighborhood: colors correspond to iterations plotted in (b); (b) confidence intervals: circles represent estimates $\hat{u}_{i,n}$ obtained at each iteration n. The gray rectangles represent the intersection between the current confidence interval and the previous one. As long as the estimate belongs to this intersection, the estimation process is updated.

$\mathbf{x}_i$ and $\mathbf{x}_j$ respectively as shown in Figure 1. In the two next sub-sections, we specify the sequence of neighborhoods $\{W_{i,n}\}$ and we propose a suitable distance to compare image patches.

2.2 Space-time neighborhood adaptation

2.2.1 Space-time neighborhood geometry

One important contribution of this work is the on-line construction of the space-time neighborhood sequence $\{W_{i,n}\}_{n\in[0:N]}$. First, we consider a simple hyper-cube space-time volume as neighborhood shape. Also, we separate the space dimensions and the time dimension. Consequently, space-time neighborhoods are parametrized by their extent in the space domain and their extent in the time domain. The use of two distinct extents (one is for the space dimensions and the other one is for the time dimension) allows us to respect space-time discontinuities and the image sequence is not considered as an isotropic $3D$ volume. In Fig 2(a), the increasing neighborhood sequence is illustrated. At each iteration, the spatial extent and the temporal extent are alternatively increased until a stopping rule is satisfied. Then, the growth of the neighborhood is stopped for this direction (*e.g.*, time) and continues in the other direction (*e.g.*, space) until the stopping rule is again satisfied. The next paragraph will explicitly define the considered stopping rule.

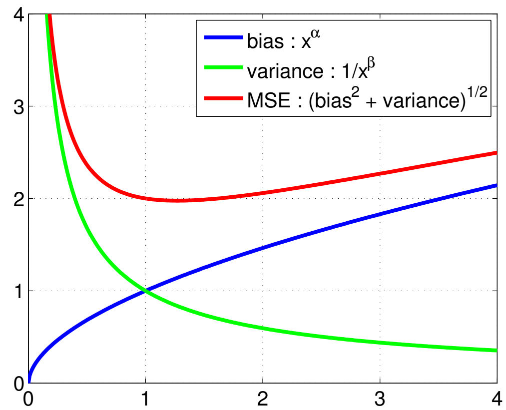


Figure 3: Illustration of the bias-variance trade-off principle. When the size of the neighborhood increases, the bias of the estimator increases while taking into account more and more intensity sample, the variance decrease. The parameters α and β are two unknown constants.

2.2.2 Space-time neighborhood selection

A point-wise rule is used to guide the space-time neighborhood selection process. This rule aims at selecting the optimal neighborhood at $\mathbf{x}_i$ and is based on the measure of the closeness of the estimator $\hat{u}_i$ to the unknown function u_i given by the local L_2 risk. This local measure of performance, used to take into account the non-stationarities in the image sequence, can be decomposed in two terms, that is the squared bias and the variance of the estimate as follows

$$\mathbb{E}\left[\hat{u}_{i,n} - u_i\right]^2 \;=\; \left[\text{bias}\left(\hat{u}_{i,n}\right)\right]^2 + \hat{\sigma}^2_{i,n}, \tag{4}$$

where $\mathbb{E}[.]$ denotes the mathematical expectation. In the sequel, we can reasonably assume that the squared bias is an increasing function of the neighborhood size and the variance is a decreasing function of the neighborhood size [15, 16, 19, 23]. Figure 3 illustrates the bias-variance trade-off principle. Then, the optimal neighborhood will be the one that achieves an optimal compromise between these two terms. A closed-form optimal solution for the ideal neighborhood is not available for such a non-linear estimator. However, we can assume that the optimal neighborhood $W^*_{i,n}$ is the one for which the squared bias and the variance are nearly the same [19]: $\mathbb{E}\left[u^*_{i,n} - u_i\right]^2 \;\approx\; 2\sigma^{2*}_{i,n}$.

A practical rule corresponding to the optimal compromise and based on a pairwise comparison of successive estimates, can be derived to select the optimal neighborhood [16]. It amounts to define the largest neighborhood satisfying the point-wise statistical rule

[16, 19, 23]

$$|\hat{u}_{i,n} - \hat{u}_{i,n'}| \;<\; \eta\, \hat{\sigma}_{i,n'},\ \forall n' < n, \tag{5}$$

as the optimal neighborhood. This rule can be interpreted as follows. While the successive estimates $\hat{u}_{i,n}$ are sufficiently close to each other, we continue the estimation process. More specifically, the estimation process is continued, while new estimates belong to the intersection of previously estimated confidence intervals $[\hat{u}_{i,n} - \eta\hat{\sigma}_{i,n}, \hat{u}_{i,n} + \eta\hat{\sigma}_{i,n}]$ (see Fig. 2). Besides, let us point out that we do not need to store all the previous estimates $\{\hat{u}_{i,n'}\}_{n' \leq n}$ but only the current intersection of confidence intervals, the last estimate and its variance for each pixel. Finally, the factor η can be easily chosen in the range $[2, 4]$ as justified with statistical arguments in [15, 16, 23].

2.3 Similarity measure for patch selection

In contrast to geometry-based approaches [12, 15], we use weights that allow us to select the correct data points in the neighborhood for averaging. This selection is based on the similarity between a given spatial $p \times p$ or a space-time $p \times p \times q$ image patch at $\mathbf{x}_i$ and $p \times p$ (or a $p \times p \times q$) image patches at $\mathbf{x}_j$ belonging to the space-time neighborhood $W_{i,n}$. Such patches are able to capture texels and local geometry in images. The L_2 distance is widely used for similarity measure between image patches. However, in order to take into account the local variance of the estimator, we introduce the following symmetric distance between image patches $\hat{\mathbf{u}}_{i,n-1}$ and $\hat{\mathbf{u}}_{j,n-1}$:

$$\Delta_{ij} = \frac{1}{2}\left[(\hat{\mathbf{u}}_{i,n-1} - \hat{\mathbf{u}}_{j,n-1})^\top \hat{V}_{i,n-1}^{-1}(\hat{\mathbf{u}}_{i,n-1} - \hat{\mathbf{u}}_{j,n-1}) + (\hat{\mathbf{u}}_{i,n-1} - \hat{\mathbf{u}}_{j,n-1})^\top \hat{V}_{j,n-1}^{-1}(\hat{\mathbf{u}}_{i,n-1} - \hat{\mathbf{u}}_{j,n-1})\right] \tag{6}$$

where the two vectors $\hat{\mathbf{u}}_{i,n-1}$ and $\hat{\mathbf{u}}_{j,n-1}$ denote the spatial $p \times p$ (or space-time $p \times p \times q$) image patches respectively centered at $\mathbf{x}_i$ and $\mathbf{x}_j$. The two matrices $\hat{V}_{i,n-1}$ and $\hat{V}_{j,n-1}$ are diagonal with the diagonal elements equal to $\hat{\sigma}^2_{i,n}$ and $\hat{\sigma}^2_{j,n}$ respectively. We decide that the two vectors $\hat{\mathbf{u}}_{i,n-1}$ and $\hat{\mathbf{u}}_{j,n-1}$ are similar with a probability of false alarm $1 - \alpha$, under the hypothesis that they are Gaussian distributed, using a classical χ^2 test with p^2 degrees of freedom. In other words, if $\Delta_{ij}/\lambda < 1$, with λ chosen as a quantile of a $\chi^2_{p^2,1-\alpha}$ distribution, we can decide that the two patches are similar. In our experiments, we use a confidence level of 99% and set α to 0.01 .

The distance Δ_{ij} is transformed into a similarity measure using the exponential kernel. We compute the similarity measure for all the points of the neighborhood and normalize it to obtain the following expression for the weights:

$$\omega_{ij} = \frac{e^{-\frac{\Delta_{ij}}{2\lambda}}}{\displaystyle\sum_{\mathbf{x}_j \in W_{i,n}} e^{-\frac{\Delta_{ij}}{2\lambda}}}. \tag{7}$$

If the distance Δ_{ij} between two patches is large then the weight ω_{ij} associated to pixel $\mathbf{x}_j$ is small and the pixel will not participate in the estimation at point $\mathbf{x}_i$. Consequently, this weighting provides an efficient and flexible way to retain the appropriate pixels contributing to the estimation of u_i in the adaptive space-time neighborhood while effectively preserving space-time discontinuities. Let us note that the process is entirely data-driven and does not require any particular geometry adapted to image contents as proposed in [12].

2.4 Motion compensation

The motion of the image sequence can be taken into account in order to apply the proposed method along the direction of the motion. However, dense motion estimation is known to be a difficult task in noisy contexts [5]. Then, we propose to robustly estimate a global parametric motion model only, which is able to capture the dominant image motion due to the camera movement. This is achieved using the multi-resolution robust method described in [22]. A similar exploitation of a parametric motion compensation was proposed in [9] and associated with a $3D$ Wiener filtering technique.

Once the dominant motion has been estimated, the filtering along the motion direction can be considered in three ways. A naive one would be to warp all the frames in the referential of the first frame. Because of the accumulation of errors and because of interpolations, it turns out that such a scheme does not improve the performance of the denoising process. The second way is to compensate the motion into the space-time neighborhood by warping the frames into the referential of the current frame, but it also involves interpolations. Then, we propose to avoid interpolation by adapting the shape of the space-time neighborhood according to the estimated dominant motion. This is achieved by translating the patch at point $\mathbf{x}_j = (x_j, y_j, t_j)$ with displacement given by the estimated parametric motion model at the center (x_i, y_i, t_j) of the neighborhood $W_{i,n}$. Moreover, when using $p \times p \times q$ space-time image patches, the motion have also to be compensated into the patch. Experiments show that this third way is able to improve the performance of the proposed method.

3 Algorithm implementation

At the initialization, the noise variance τ^2 has first to be estimated. It can be robustly estimated by calculating pseudo-residuals ε_i as described in [14]. We consider four spatial neighbors and two spatial ones, and pseudo-residuals are compactly represented by $\varepsilon_i = (8Y_i - \Delta Y_i)/\sqrt{42}$ where ΔY_i is the discrete Laplacian of Y_i at $\mathbf{x}_i$ (See Eq. (1)) and the constant $\sqrt{42}$ is introduced to ensure that $\mathbb{E}[\varepsilon_i^2] = \tau^2$. Given the residuals ε_i, the noise variance τ^2 can then be robustly estimated as:

$$\tau = 1.4826 \operatorname{med}_i \left(| \ \varepsilon_i - \operatorname{med}_j |\varepsilon_j| \ | \right) \tag{8}$$

Let us recall that parameter λ is set to the 0.99 quantile of the $\chi^2_{p^2,0.99}$ distribution with a size of patch p chosen within $\{3, 5, 7, 9\}$. The last parameter η is set to $2\sqrt{2}$ to ensure a good accuracy of the estimation [16, 19, 23]. During the estimation, spatial and temporal extents of the space-time neighborhoods are alternatively increased (see Section 2.2).

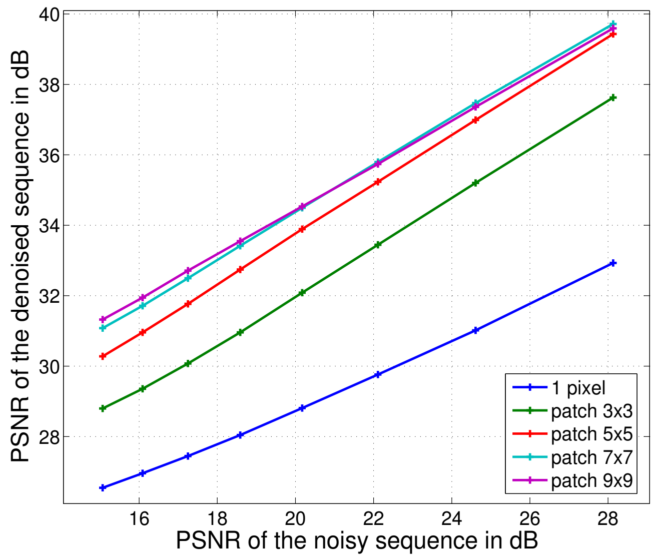


Figure 4: Performance of our denoising method for several noise levels and for several patch sizes including a no-patch version (pixel-wise). The $PSNR$ is used to measure the overall performance of the filtering and the test sequence is Akyio ($176 \times 144 \times 300$). This sequence mainly exhibits low motion. Experiments show that the introduction of patches improves the $PSNR$ of at least $2dB$. We can also point out that the results for patches 7×7 and 9×9 are similar.

Furthermore, the algorithm can be easily parallelized. Indeed, the estimation steps use only local information and thus we have distributed the computation load over eight CPUs dividing the computation time by eight. Finally, another possibility to speed up the algorithm is to use a dyadic scheme when increasing the extent of the neighborhood. The proposed method is very simple to implement and does not require the fine adjustment of parameters λ and η which control the estimation process.

4 Experimental results

In this section, a large set of experiments are reported to validate our patch-based space-time adaptive estimation method. We first consider real image sequences with artificially added Gaussian white noise. Using this usual protocol, we compare our method to other recent methods for denoising image sequences. For an objective performance evaluation, we consider the global measure given by the Peak-Signal-to-Noise-Ratio defined as $PSNR = 20\log_{10}(255/mse)$ (mse denotes the mean squared error between the original noise-free image sequence and the denoised image sequence). The visual quality of the image sequence is also taken into account. We then discuss the usefulness of motion compensation and we comment the respective performance of $2D$ spatial patches and $3D$ space-time patches. Finally, the proposed method is applied to a real noisy Infra-Red image sequence.

Performance assessment We first report experiments to evaluate the influence of the noise level and the spatial patch size on the overall method performance. Figure 4 plots $PSNR$ values obtained for eight noise levels and five patch sizes. First, we can note that the patch-based method performance is smoothly affected with the increase of the noise level. The improvement gained by introducing patches (to be compared to no-patch version) is clearly demonstrated. As it could be expected, it is useless to consider too large patches. When the size of the patch increases, the $PSNR$ increases too, however, results for sizes 7×7 and 9×9 are quite similar while the computation time is proportional to the number of pixels in the patch.

The well known sequence "*Flower garden*" is shown on Fig. 5 to illustrate the visual quality of the denoised image sequence in very noisy conditions. In order to give insights into the spatio-temporal behavior of the denoising method, we have displayed XT slices of the image sequence. The reported results demonstrate that our method can cope with the presence of motion while preserving temporal discontinuities and reaches a $PSNR$ of 23.59 using 7×7 patches and 6 iterations. By applying our own implementation of the *non-local mean* algorithm [5] on the "Flower garden" sequence using a 7×7 patch in a $21 \times 21 \times 3$ neighborhood and choosing a bandwidth equal to $h = 12 \times \tau^2$, we get a $PSNR$ equal to $21.14dB$ only.

Comparison with other recent methods We have compared our method with four other recent methods: a combination of a spatial Wiener filter with a motion-compensated temporal Kalman filter [10], a space-time non linear adaptive K-NN filter [28], a 3D wavelet-based method [24] and a 3D point-wise adaptive estimate using different neighborhood geometries [12]. For these experiments, the motion-compensation stage is not applied.

Eight test sequences corrupted with noise of different levels are used. For a fair evaluation, we have considered the results supplied by the authors themselves in the referenced papers. Therefore, we cannot provide the $PSNR$ measures for all the test sequences and Table 1 contains all the available results. Our method clearly outperforms all the other methods since it supplies the best $PSNR$ results for all the tested sequences, sometimes with a quite significative improvement (up to $4dB$). Let us also stress that the implementation of our method is straightforward and our method involves no parameter tuning.

Motion compensation In this section, we aim at evaluating the impact of a motion compensation stage on the performance of our method. We have applied this version to the "*Avenger*" sequence which contains two moving cars tracked by the camera mounted

Sequence name	Size	$PSNR$	(a)	(b)	(c)	(d)	(e)
Akiyo	$176 \times 144 \times 300$	22	–	–	–	33.86	**34.31**
Salesman	$176 \times 144 \times 449$	28	34.4	32.5	–	–	**35.13**
		24	31.1		–	–	**32.60**
Garden	$352 \times 240 \times 115$	28	–	28.2	–	–	**31.33**
Miss America	$176 \times 144 \times 108$	28	–	35.3	–	–	**39.39**
Miss America	$128 \times 128 \times 128$	7	–	–	26.36	–	**26.69**
		12	–	–	28.16	–	**29.63**
		17	–	–	30.46	–	**32.05**
		22	–	–	32.66	–	**34.20**
Suzie	$176 \times 144 \times 150$	28	34.8	–	–	–	**37.07**
		24	32.0	–	–	–	**35.11**
Trevor	$176 \times 144 \times 90$	28	33.9	34.1	–	–	**36.68**
		24	31.3	–	–	–	**34.79**
Foreman	$176 \times 144 \times 300$	28	33.9	–	–	–	**34.94**
		24	31.1	–	–	–	**32.90**

Table 1: PSNR results for eight test sequences and five denoising methods. (a) Join Kalman and Wiener denoising with motion compensation using dense motion field [10], (b) Adaptive K-NN space-time filter [28],(c)Wavelet-based method for image sequence denoising: TIWP3D [24], (d) 3D non-parametric regression approach [12], (e) the proposed adaptive method with 7×7 patches and 6 iterations without motion compensation. Numerical results for the other methods are taken from the related publications. The symbol – means that results were not provided by the authors for the corresponding test sequence.

in an helicopter. Fig. 6(a) and (b) respectively contain one image of the original image sequence and of the noisy image sequence with an additive Gaussian white noise of standard deviation $\tau = 20$. Fig. 6(c) displays the result of the version of our method without motion-compensated while Fig. 6(e) shows the improvement supplied by the motion-compensation stage. The $PSNR$ difference between the two image sequence is about $1dB$ and the visual quality is also improved. We can then conclude that, when the global motion the image sequence is well described by a 2D parametric model, the proposed motion compensation scheme improves the quality of the denoising process. A quadratic motion model was used. Let us add that the motion of the two cars is not handled by the dominant motion model. However, since our method involves a data-driven adaptation scheme, the neighborhoods for the pixels belonging to these two cars essentially reduce to 2D spatial ones.

Space-time patches Fig. 6(d) shows one image of the "*Avenger*" sequence denoised using the proposed method with space-time $3 \times 3 \times 3$ patches. It can be compared to the result obtained with spatial 5×5 patches and shown in Fig. 6(c). In the two cases, the number of intensity values used to compute the similarity measure (see Eq. 7) is approximatively

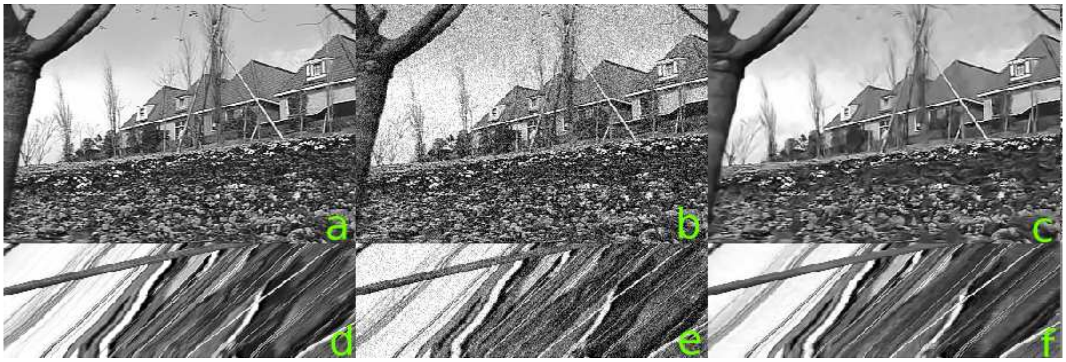


Figure 5: Sequence "*Flower garden*". (a) one image of the original sequence, (b) the same image of the noisy sequence with an additive Gaussian white noise of standard deviation $\tau = 30$ ($PSNR = 18.58dB$) and (c) the corresponding image of the denoised sequence, $PSNR = 23.59dB$. (d), (e) and (f) represent corresponding XT slices in the space-time domain. The camera motion appear as lines in the XT slices.

the same. The $PSNR$ difference is negligible. Nevertheless, the use of space-time patches increases the temporal stability of the reconstructed structures along the image sequence which is an important point for the visual quality.

Experiment on a real image sequence Fig. 7 reports an experiment on a real infra-red sequence which is naturally noisy. It is taken from a plane approaching an harbor with boats and a moving vehicle on the pier. The noise standard deviation is estimated to 7.3. The contrast of some structures on the ground is very low and the sequence is shaking due to the plane vibrations. We use once again a quadratic motion model to estimate the dominant motion. Let us recall that this model is exact in the case of a rigid motion and a planar scene. The proposed method (Fig. 7c.) can be favorably compared to our own implementation of the *non-local means* algorithm (Fig. 7b.) [5]. The details of the images like the small vehicle are better restored while the noise has been well removed. Finally, temporal discontinuities of the sequence due to the vibrations are also preserved.

5 Conclusion

We have described a novel and efficient unsupervised method for denoising image sequences. The proposed method is based upon an adaptive estimation statistical framework. It can specify, in a simple data-driven way, the most appropriate space-time neighborhood and associate weights to select the data points involved in the intensity estimation process at each pixel. Moreover, it involves a patch-based approach extended to the space-time domain.


(a)


(b) $PSNR = 22dB$


(c) $PSNR = 31.53dB$


(d) $PSNR = 31.47dB$


(e) $PSNR = 32.50dB$


(f) $PSNR = 32.40dB$

Figure 6: Sequence "Avenger". (a) $384 \times 288 \times 12$ original sequence (b) noisy sequence with an additive Gaussian white noise of standard deviation $\tau = 20$, denoised sequence with the proposed method using : (c) 5×5 patches and no motion compensation stage (5 min), (d) $3 \times 3 \times 3$ patches and no motion compensation stage (6 min), (e) 5×5 patches with the motion compensation stage using a quadratic motion model (9 min), (d) $3 \times 3 \times 3$ patches with the motion compensation stage using a quadratic motion model (20 min). The computation time is indicated for a Linux PC with $8 \times 3Ghz$ CPU.

All the parameters of the algorithm are well calibrated and our method does not require any fine parameter tuning. Quite satisfactory results have been obtained on several image sequences. Furthermore, it was experimentally demonstrated that our method outperforms other recent methods. The visual quality improvement of the denoised image sequences is noticeable since noise is well smoothed out while spatial and temporal discontinuities are well preserved. Finally, some improvements are proposed to incorporate a motion-compensation stage improving the performance of the proposed method.

References

[1] Danny Barash. A fundamental relationship between bilateral filtering, adaptive smoothing, and the nonlinear diffusion equation. *IEEE Trans. on Pattern Analysis and Ma-*

Figure 7: Infra-red sequence. (a) one image of the original sequence, (b) the corresponding image of the denoised sequence with the *non-local* filter applied with a patch of size 5×5 and the bandwidth equal to four times the noise standard deviation, (c) denoised image with the proposed motion compensated filter, (b1) Cropped region of (b), (c1) Cropped region of (c).

chine Intelligence, 24(6):844–847, 2002.

[2] M. Black, G. Sapiro, D. Marimont, and D. Heeger. Robust anisotropic diffusion. *IEEE Trans. on Image Processing*, 7(3):421–432, 1998.

[3] J. Boulanger, C. Kervrann, and P. Bouthemy. Adaptive spatio-temporal restoration for 4d fluoresence microscopic imaging. In *Proc. of 8th International Conference on Medical Image Computing and Computer Assisted Intervention*, Palm Spring, California, October 2005.

[4] J.C. Brailean, R.P. Kleihorst, S. Efstratiadis, A.K. Katsaggelos, and R.L. Lagendijk. Noise reduction filters for dynamic image sequences: A review. *Proc. of the IEEE*, 83(9):1272–1291, 1995.

[5] A. Buades, B. Coll, and J.M. Morel. Denoising image sequences does not require motion estimation. Pre-Print CMLA, 2005.

[6] A. Buades, B. Coll, and J.M. Morel. A review of image denoising algorithms, with a new one. *Multiscale Modeling and Simulation: A SIAM Interdisciplinary Journal*, 4(2):490–530, 2005.

[7] H-Y Cheong, A.M. Tourapis, J. Llach, and J. Boyce. Adaptive spatio-temporal filtering for video-denoising. In *Proc. of IEEE Int. Conf. on Image Processing, ICIP'2004*, volume 2, pages 965–968, Singapore, October 2004.

[8] A. Criminisi, P. Pérez, and K. Toyama. Region filling and object removal by exemplar-based inpainting. *IEEE Trans. on Image Processing*, 13(9):1200–1212, 2004.

[9] F. Dekeyser, P. Bouthemy, and P. Pérez. Spatio-temporal Wiener filtering of image sequences using a parametric motion model. In *Proc. of IEEE Int. Conf. on Image Processing, ICIP'2000*, pages 208–211, Vancouver, September 2000.

[10] R. Dugad and N. Ahuja. Video denoising by combining kalman and wiener estimates. In *Proc. of IEEE Int. Conf. on Image Processing, ICIP'1999*, volume 4, pages 152–156, Kobe, October 1999.

[11] A. Efros and T. Leung. Texture synthesis by non-parametric sampling. In *Proc. of 7th IEEE Int. Conf. on Computer Vision, ICCV'99*, pages 1033–1038, Kerkyra, Greece, 1999.

[12] C. Ercole, A. Foi, V. Katkovnik, and K. Egiazarian. Spatio-temporal pointwise adaptive denoising in video: 3d non parametric approach. In *Proc. of the 1st International Workshop on Video Processing and Quality Metrics for Consumer Electronics, VPQM'2005*, Scottsdale, 2005.

[13] I.W. Fei Shi, Selesnick. Video denoising using oriented complex wavelet transforms. In *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'04)*, volume 2, pages 949–952, Montreal, 2004.

[14] T. Gasser, L. Sroka, and C. Jennen Steinmetz. Residual variance and residual pattern in nonlinear regression. *Biometrika*, 73:625–633, 1986.

[15] V. Katkovnik, K. Egiazarian, , and J. Astola. Adaptive window size image denoising based on intersection of confidence intervals (ICI) rule. *Journal of Mathematical Imaging and Vision*, 16(3):223–235, 2002.

[16] C. Kervrann. An adaptive window approach for image smoothing and structures preserving. In *Proc. of 8th Eur. Conf. on Computer Vision, ECCV'2004*, pages 132–144, Prague, 2004.

[17] C. Kervrann and J. Boulanger. Unsupervised patch-based image regularization and representation. In *Proc. of 9th Eur. Conf. on Computer Vision, ECCV'2006*, Graz, Austria, May 2006.

[18] Suk Ho Lee and Moon Gi Kang. Spatio-temporal video filtering algorithm based on 3-D anisotropic diffusion equation. In *Proc. of IEEE Int. Conf. on Image Processing, ICIP'1998*, volume 3(2), pages 447–450, Chicago, 1998.

[19] O. Lepski. Asymptotically minimax adaptive estimation 1: upper bounds. *SIAM Journal Theory of Probability and Application*, 36(4):654–659, 1991.

[20] M. Mahmoudi and G. Sapiro. Fast image and video denoising via nonlocal means of similar neighborhoods. *SPLetters*, 12(12):839–842, December 2005.

[21] G. Motta, E. Ordentlich, I. Ramirez, G. Seroussi, and M.J. Weinberger. The dude framework for continuous tone image denoising. In *Proc. of IEEE Int. Conf. on Image Processing, ICIP'2005*, volume 3, pages 345–348, 2005.

[22] J.M. Odobez and P. Bouthemy. Robust multiresolution estimation of parametric motion models. *Journal of Visual Communication and Image Representation*, 6(4):348–365, december 1995.

[23] J. Polzehl and V.G. Spokoiny. Adaptive weights smoothing with applications to image restoration. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(2):335–354, 2000.

[24] N. Rajpoot, Zhen Yao, and R. Wilson. Adaptive wavelet restoration of noisy video sequences. In *Proc. of IEEE Int. Conf. on Image Processing, ICIP'2004*, pages 957–960, 2004.

[25] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60(1-4):259–268, 1992.

[26] Y. Wexler, E. Shechtman, and M. Irani. Space-time video completion. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'04)*, volume 1, pages 120–127, Washington, 2004.

[27] D. Zang and Z. Wang. Image information restoration based on long-range correlation. *IEEE Trans. on Circuits and Systems for Video Technology*, 12(5):331–341, May 2002.

[28] V. Zlokolica and W. Philips. Motion and detail adaptive denoising in video. *Proc. of SPIE – Image Processing: Algorithms and Systems III*, 5298:403–412, 2004.