VIDEO DENOISING BASED ON ADAPTIVE TEMPORAL AVERAGING

David BARTOVČAK – Miroslav VRANKIĆ

Abstract: *This paper proposes a video denoising algorithm based on adaptive, pixel-wise, temporal averaging. The algorithm decomposes videos into a set of 1-D time dependent signals and then removes the noise by establishing temporal averaging intervals throughout each signal from the set. Temporal averaging intervals are established by simple, yet effective comparison processes which include two-way thresholding. The proposed algorithm is tested on several types of 1-D signals and benchmark videos. Experiments suggest that the proposed algorithm, despite its simplicity, produces high-quality denoising results and even outperforms some state-of-the-art competitors.*

Keywords: – signal processing
– video denoising
– temporal averaging
– averaging interval
– pixel-domain method
– additive white Gaussian noise

1. INTRODUCTION

Image and video quality enhancement is a long-standing area of the research. Image and video signals are often contaminated by noise during acquisition and transmission and noise is a dominant factor that degrades image/video quality. Low-end camera market is growing rapidly (digital cameras, web-cams, cell phones etc.) and there is a need now more than ever for fast, effective and reliable image and video enhancement technologies to improve their output. Even high-end and professional equipment (surveillance cameras, medical devices) have to cope with image degradation and noise corruption (especially in extreme conditions). Nowadays, practically every image-capturing device incorporates some sort of noise removal technology. Video denoising is highly desirable, not only for improving perceptual quality, but also for increasing compression speed and effectiveness and facilitating transmission bandwidth reduction. Most video denoising algorithms proposed in the literature consider additive white Gaussian noise (AWGN) and can be categorized into pixel domain and transform domain methods. Pixel domain methods can be further subcategorized into spatial, temporal and spatiotemporal domain methods and mostly reduce noise by weighted averaging. Majority of the recently proposed pixel domain algorithms argue that spatiotemporal filtering performs better than temporal filtering, e.g. ST-GSM algorithm [1]. However, spatiotemporal filtering, as well as spatial filtering, may significantly reduce the effective resolution of the video due to spatial blurring [2]. Most videos are temporally consistent and every new frame can be predicted from previous frames. If two videos are given with similar *PSNR* values, one filtered with spatial and the other with temporal algorithm, the latter may be preferred just because of the temporal coherence. Motion information and temporal coherence information can be incorporated in video denoising algorithms through applying advanced transforms [3, 4]. However, most transform domain video denoising methods tend to be complicated, slow, and not applicable on consumer electronics. Pixel domain temporal denoising algorithms also offer the ability of motion estimation in order to preserve full resolution of the input image sequence and temporal coherence. But, even pixel domain methods, which use block-matching or similar techniques, often require lots of computer resources for effective noise removal. This paper proposes a pixel-wise temporal video denoising method which is simple, intuitive, yet effective and competitive. The method observes videos as a group of 1-D signals – every video is decomposed in m-by-n one-dimensional signals (m represents width, and n height of image frame in a number of pixels) and then processed. This approach allows the algorithm to remain simple, fast and highly adaptable. Noise removal is achieved by an adaptive temporal averaging applied on averaging intervals.

This article intends to show that high-quality video denoising doesn't require the use of wavelet transforms, NLM methods of searching of similar patches and other methods (which are unavoidably and inherently complex). This paper presents a simple and "lightweight" pixel-wise method of video denoising that can produce high-quality noise removal results and even outperform some of its more sophisticated competitors. Experiments were conducted to test developed algorithms. The proposed algorithm is first tested on a set of test 1-D signals corrupted by additive white Gaussian noise and the results are used for tweaking the algorithm. The algorithm is then evaluated using gray-scale benchmark videos. The proposed method is presented in Section 2. Section 3 contains experimental results and the conclusion is presented in Section 4.

2. THE PROPOSED ALGORITHM

2.1. Observation model

Before examining the algorithm, the observation model and notation used throughout the paper must be introduced. A noisy signal is considered:

$$f[k] = s[k] + n[k] \tag{1}$$

where $s[k]$ is a noise-free signal and $n[k]$ is additive white Gaussian noise with zero mean and variance σ^2 (standard deviation σ). If signal $f[k]$ is a representation of time-dependent pixel intensity taken from a monochromatic video then $f[k]$ takes values between 0 and 255 and has a finite number of samples.

2.2. ATA method

In this paper ATA (Adaptive Temporal Averaging) method is proposed. The method is based on a simple idea. A noisy input signal $f[k]$ is considered where k = 1, 2, 3, …, m. Noise is removed by establishing estimation intervals and applying averaging. Estimation intervals are established for every sample of the noisy input and every sample of the resulting (noise-free) signal is obtained by using the corresponding estimation/averaging interval of the noisy input.

The key to high-quality denoising is reliable estimation of averaging intervals. This paper proposes the use of temporal coherence of the signal in order to establish averaging intervals meaning that every sample of a noisy input has a certain amount of similar samples occurring immediately before and/or after it. If that is presented graphically (with time placed on the x-axis and sample values placed on the y-axis), then every sample has a certain amount of similar samples residing immediately to the left and/or to the right. Groups of similar samples are used to form averaging intervals and an averaging interval is formed for every sample of the noisy input.

Forming of averaging intervals is simple – ATA algorithm is comparing currently processed sample with consecutive samples to its left and right side (comparison of samples to the left and to the right are mutually independent). When the comparison process reaches a sample to the left that is significantly different from the one currently processed, the algorithm stops the left-hand side consecutive comparison. This is how the left-hand side border of the averaging interval is obtained. The same procedure is used for obtaining the right-hand side border of the averaging interval. When the averaging interval is determined, then the estimate of the noise-free signal is obtained using mean value of the averaging interval as follows:

$$e[k] = \frac{f[l_k] + \ldots + f[k] + \ldots + f[r_k]}{r_k - l_k + 1}, \qquad r_k \geq k \geq l_k \tag{2}$$

where $e[k]$ is estimated noise-free signal, l_k is left-hand-side and r_k is right-hand side border of the averaging interval. A two-way threshold criterion is used to determine borders of the averaging intervals. While determining borders of the averaging intervals, the algorithm examines absolute differences between the sample that is being currently processed and samples residing to its left/right. When examined, absolute difference is greater than predetermined value (which was named Threshold A), the border of the averaging interval is

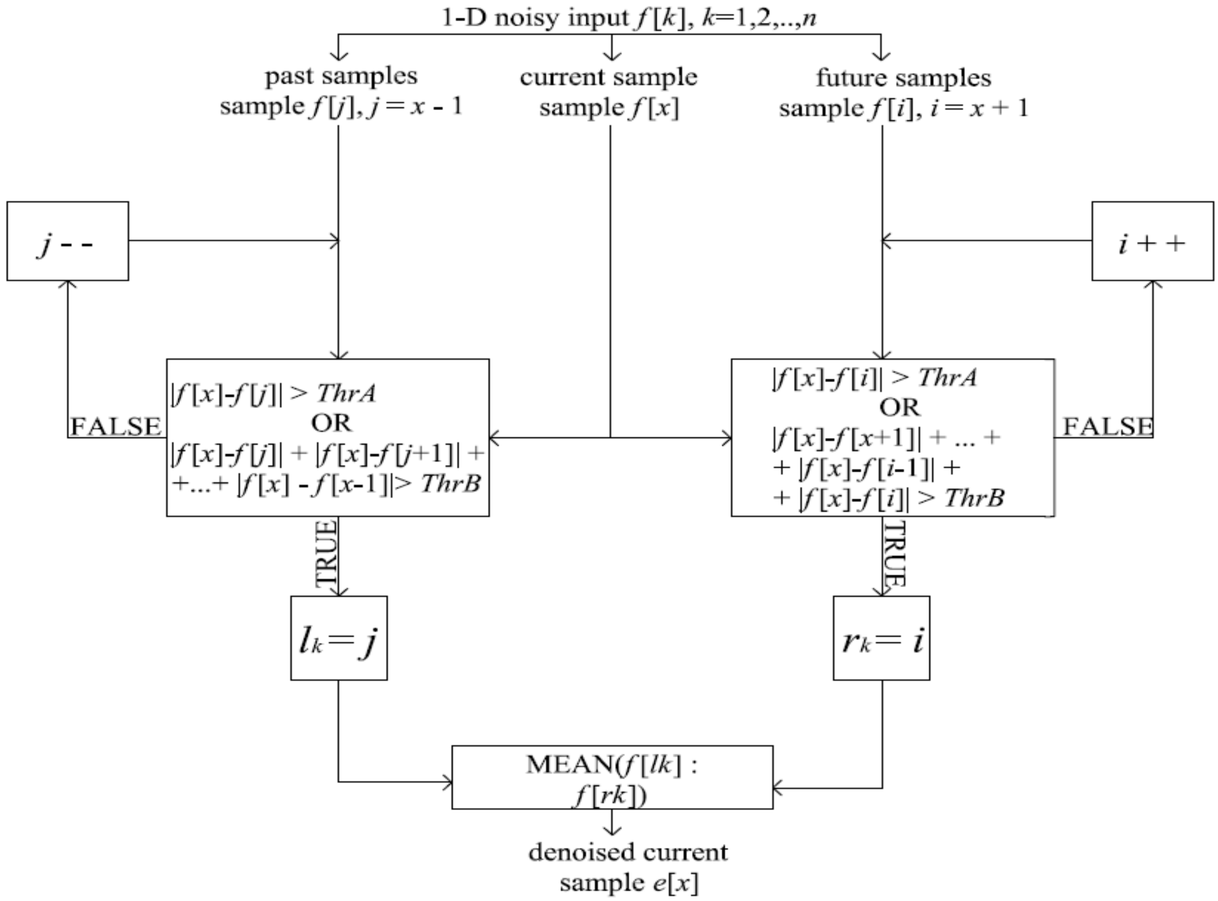

Figure 1. Flowchart of the proposed algorithm

found. Simultaneously, the ATA method is cumulating the above mentioned differences. When that sum is greater than predetermined value (which was named Threshold B), the border of the averaging interval is found (refer to Figure 1, which shows the algorithm flowchart). Hence, two threshold criteria are responsible for determining the borders of the averaging intervals (bear in mind that the left-hand side border is completely independent from the right-hand one). Both criteria are used simultaneously, meaning that the averaging interval border is determined by the one criterion that provides an averaging interval border first. The Threshold A is designed to react to abrupt changes in the input signal and Threshold B is designed to react to continuous changes in the input signal. An iterative method was used for determining optimal values of Threshold A and B – various signals and videos were processed iteratively until the best values of *MSE*, *PSNR* and *SSIM* were achieved. By using this approach, empirical optimal values for Threshold A and Threshold B were determined and used in further experiments. They are as follows: *ThresholdA* = $5 \cdot \sigma$, *ThresholdB* = $10 \cdot \sigma$.

3. EXPERIMENTAL RESULTS

Two 1-D signals were used in experiments to determine the effectiveness of the ATA method. Also, several different gray-scale sequences were used for testing the algorithm: “Salesman”, “Miss America”, “Tennis”. They were corrupted with additive white Gaussian noise of various values of the standard deviation σ = 10, 15, 20 and processed with the proposed filter.

3.1. ATA method applied on 1-D signals

Two different test signals were used:

- **Signal A** – quasi-rectangular signal; features areas of constant values and major leaps in signal values.
- **Signal B** – quasi-sine wave; features continuous changes in signal values.

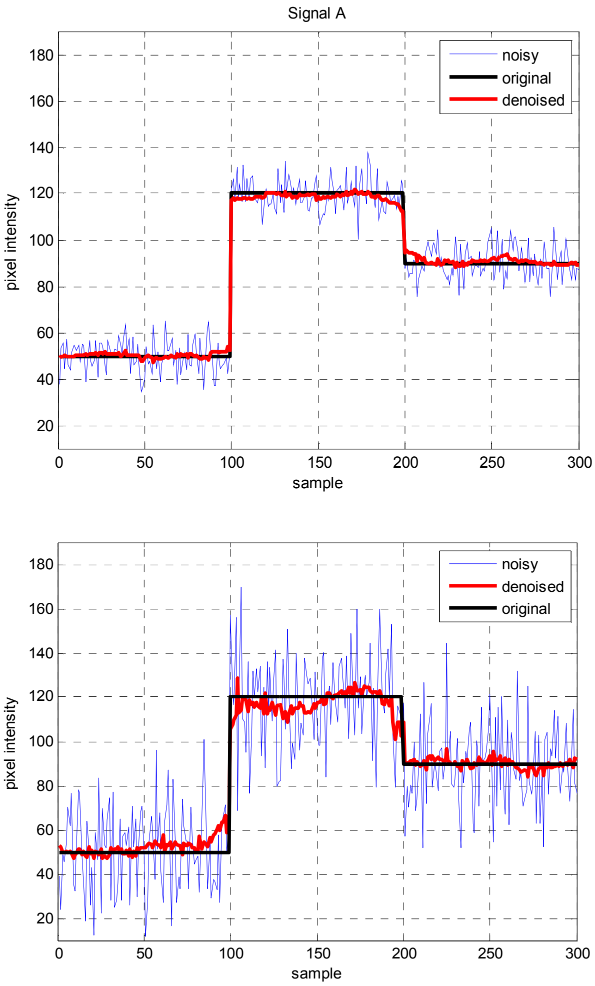

Figure 2. *Denoising of Signal A using ATA method; up: noise σ = 10, down: noise σ = 20*

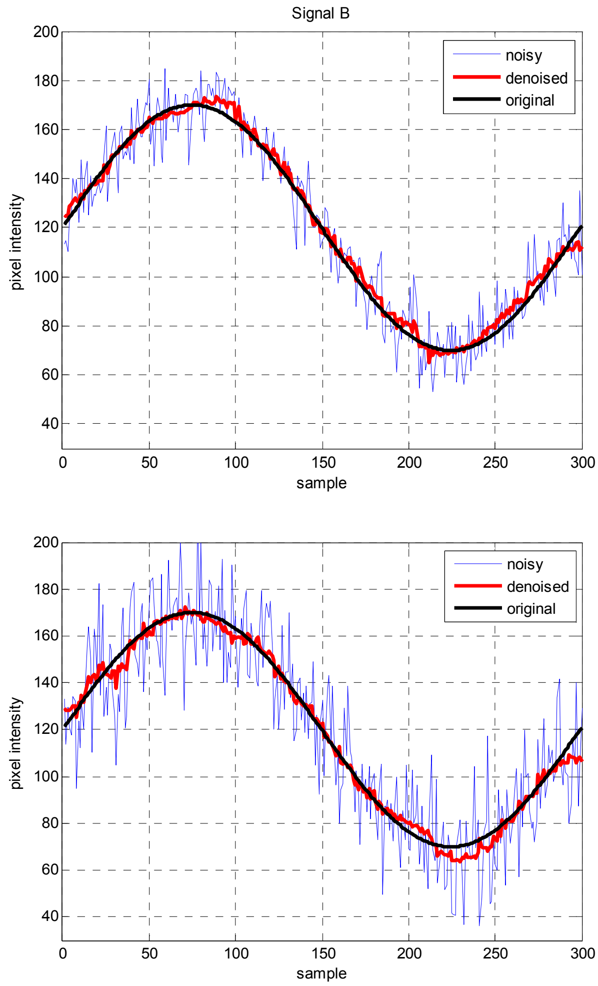

Figure 3. *Denoising of Signal B using ATA method; up: noise σ = 10, down: noise σ = 20*

Both signals were corrupted with Gaussian noise of standard deviation $\sigma = 10, 15, 20$. The ATA method was used for noise removal and results are expressed by determining *MSE* (Mean Squared Error) of the denoised signals (Table 1). Figures 2 and 3 illustrate denoising of Signals A and B using the ATA method. By examining the results shown in Table 1 (along with Figures 2 and 3), the conclusion is reached that substantial noise removal is achieved.

3.2. ATA method applied on image sequences

Several different gray-scale sequences were used to determine the method performance. For straight-forward comparison with existing algorithms, the noise standard deviations are chosen to be 10, 15 and 20, respectively, and only the results of the luminance (Y) channel are reported. Peak signal-to-noise ratio (*PSNR*) and the structural similarity (*SSIM*) index [5] are used to provide quantitative evaluations of the algorithm. The latter has shown to be a better indicator of perceived image quality [5]. The proposed algorithm has been compared with state-of-the-art video denoising algorithms, including SEQWT [6], IFSM [7] and ST-GSM [1]. To better place the performance evaluation in the context, one baseline algorithm has also been included in comparison; still GSM [8]: static image GSM denoising applied on a frame-by-frame basis. Values of *PSNR* and *SSIM* obtained by the ATA method and comparison to other state-of-the-art

Table 1. MSE values of 1-D signals denoised using ATA method

Noise σ	10	15	20
MSE values			
Signal A	2.917	3.038	21.488
Signal B	9.150	10.361	13.180

Figure 4. Denoising of video "Salesman" corrupted with noise σ = 20 using ATA method, frame 100; from left to right: original, noisy, denoised

Figure 5. Denoising of video "Tennis" corrupted with noise σ = 20 using ATA method, frame 300; from left to right: original, noisy, denoised

Table 2. PSNR and SSIM values of video sequences denoised using ATA method and comparison to other methods

Video	"Salesman"			"Miss America"			"Tennis"		
Noise σ	10	15	20	10	15	20	10	15	20
PSNR [dB]									
Noisy	28.16	24.72	22.32	28.15	24.62	22.29	28.22	24.72	22.25
SEQWT [6]	32.86	30.59	29.09	NA	NA	NA	31.19	29.14	27.59
IFSM [7]	34.22	31.85	30.22	37.52	35.41	33.86	32.41	30.10	28.56
still GSM [8]	33.80	31.73	30.28	38.52	37.14	36.14	31.82	29.87	28.65
ST-GSM [1]	38.04	36.03	34.61	40.57	39.40	38.50	34.05	31.97	30.59
Proposed	35.64	33.78	32.51	37.16	35.77	34.30	32.95	30.55	28.72
SSIM									
Noisy	0.718	0.574	0.467	0.493	0.321	0.226	0.719	0.573	0.466
SEQWT [6]	0.900	0.846	0.796	NA	NA	NA	0.842	0.772	0.716
IFSM [7]	0.904	0.851	0.801	0.904	0.857	0.812	0.855	0.776	0.709
still GSM [8]	0.909	0.865	0.825	0.936	0.922	0.913	0.831	0.758	0.711
ST-GSM [1]	0.960	0.941	0.923	0.952	0.943	0.936	0.894	0.841	0.797
Proposed	0.941	0.920	0.899	0.910	0.884	0.861	0.870	0.813	0.756

methods are contained in Table 2. In Figures 4 and 5 denoised frames of videos "Salesman" and "Tennis" are shown along with original and noisy frames for comparison. The evaluation shows that the amount of noise removed is significant and the ATA method is highly competitive when compared to state-of-the-art noise removal algorithms.

4. CONCLUSION

This paper proposed a simple and intuitive, yet highly competitive video denoising method based on determining averaging intervals. The method exploits video's temporal coherence to obtain averaging intervals by simple two-way thresholding designed to produce the averaging interval's left-hand side and right-hand side borders. Noise is then removed by applying averaging on sample-wise temporally adaptive intervals. Several 1-D signals and benchmark videos were used to examine performance of the proposed method. Denoising 1-D signals with the proposed method has produced significant improvement in *MSE* values. Denoising videos with the proposed method has also produced

significant increase in *PSNR* and *SSIM* values. Compared to other state-of-the-art methods, the proposed method outperforms many of its competitors, e.g. SEQWT [6], IFSM [7] and still GSM [8]. The proposed algorithm is simple, easy to use and doesn't require more complex video denoising procedures that are numerically more demanding. Further improvements of the proposed method are possible and desirable, especially taken into account that ST-GSM [1] method is still able to produce somewhat better denoising results.
Improvements will be explored not only considering effectiveness of the proposed method but also its speed. The future research will focus on the development of more sophisticated and more reliable threshold criteria for determining borders of the averaging intervals, which will enable better and faster performance of the method.

5. LIST OF SYMBOLS

Noisy signal	$f[k]$,	-
Original (noise-free) signal	$s[k]$,	-
Noise	$n[k]$,	-
Estimated noise-free signal	$e[k]$,	-
Standard deviation	σ,	-
Right-hand side border of the averaging interval	r_k	-
Left-hand side border of the averaging interval	l_k,	-
Mean squared error	*MSE*,	-
Peak signal-to-noise ratio	*PSNR*,	dB
Structural similarity index	*SSIM*,	-

REFERENCES

[1] Varghese, G., Wang, Z.: *Video denoising using a spatiotemporal statistical modeling of wavelet coefficients*, IEEE Acoustics, Speech And Signal Processing International Conference, Las Vegas, April 2008.
[2] Zlokolica, V., Pizurica, A., Philips, W.: *Recursive temporal denoising and motion estimation of video*, International Conference on Image Processing, Singapore, October 2004.
[3] Rusanovskyy, D., Egiazarian, K.: *Video denosing algorithm in sliding 3D DCT domain*, Advanced Concepts for Intelligent Vision Systems, ACIVS, Antwerp, September 2005.
[4] Selesnick, W. I., Li, K. Y.: *Video denoising using 2D and 3D dualtree complex wavelet transforms*, in Proc. SPIE, Wavelets: Applications in Signal and Image Processing X, vol. 5207, pp. 607-618, San Diego, November 2003.
[5] Wang, Z., Bovik, A. C., Sheikh, H. R., Simoncelli, E. P.: *Image quality assessment: from error visibility to structural similarity*, IEEE Trans. Image Processing, vol. 13, pp. 600-612, April 2004.
[6] Zlokolica, V., Pizurica, A., Philips, W.: *Combined wavelet and temporal video denoising*, IEEE Conf. on Advanced Video and Signal Based Surveillance, pp. 334-341, Miami, July 2003.
[7] Rahman, S. M. M., Ahmad, M. O., Swamy, M. N. S.: *Video denoising based on inter-frame statistical modeling of wavelet coefficients*, IEEE Trans. Circuits and Systems for Video Tech., vol. 17, pp. 187-198, February 2007.
[8] Portilla, J., Strela, V., Wainwright, M. J., Simoncelli, E. P.: *Image denoising using Gaussian scale mixtures in the wavelet domain*, IEEE Trans. Image Processing, vol. 12, pp. 1338-1351, November 2003.

Received: 01.09.2011. Accepted: 09.04.2012.

Author´s address:

David Bartovčak
Miroslav Vrankić
Faculty of Engineering, University of Rijeka
Vukovarska 58, 51000 Rijeka
david.bartovcak@gmail.com
miroslav.vrankic@riteh.hr