Video Denoising Algorithm in Sliding 3D DCT domain

Dmytro Rusanovskyy and Karen Egiazarian

Institute of Signal Processing
Tampere University of Technology, Finland
E-mails: {FirstName.LastName}@tut.fi

Abstract. The problem of denoising of video signals corrupted by additive Gaussian noise is considered in this paper. A novel 3D DCT-based video-denoising algorithm is proposed. Video data are locally filtered in sliding/running 3D windows (arrays) consisting of highly correlated spatial layers taken from consecutive frames of video. Their selection is done by the use of a block matching or similar techniques. Denoising in local windows is performed by a hard thresholding of 3D DCT coefficients of each 3D array. Final estimates of reconstructed pixels are obtained by a weighted average of the local estimates from all overlapping windows. Experimental results show that the proposed algorithm provides a competitive performance with state-of-the-art video denoising methods both in terms of PSNR and visual quality.

1. Introduction

Digital images and video nowadays are essential part of everyday life. Often imperfect instruments of data acquisition process, natural phenomena, transmission errors and compression can degrade a quality of collected data. Presence of noise may sufficiently affect the further data processing such as analysis, segmentation, classification and indexing. Denoising is typically applied before any aforementioned image/video data processing. Herein, the problem of denoising of video corrupted by additive independent white Gaussian noise is considered.

Historically, first algorithms for video denoising operated in spatial or spatio-temporal domains [1]. Recent research on denoising has demonstrated a trend towards transform-based processing techniques. Processing in a transform domain (e.g. in DCT, DFT or wavelet domains) provides a superior performance comparing to the spatio-temporal methods due to a good decorrelation and compaction properties of transforms.

Wavelet-based video denoising was inspired by the results of the intensive work on the wavelet-based image denoising [3-5] initiated by Donoho's *wavelet shrinkage* approach [2]. Several multiresolution (*wavelet-based*) approaches were recently proposed to the problem of video denoising, see, e.g. [6] and [7].

Local adaptive *sliding window DCT* (SWDCT) image denoising method [8], [9] is a strong alternative to the wavelet-based methods. This paper gives an extension of it to SWDCT denoising of video. This extension is not a straightforward one. Video data in the temporal direction are not stationary due to a motion present in videos. Thus, two pixels located at the same spatial location of consecutive frames could be uncorrelated. On the other hand, DCT is a good approximation of the statistically optimal Karhunen-Loeve transform in the case of highly correlated data [10]. Thus, a wise selection of local 3D data through the different frames should be performed before any application of a 3D DCT. One approach to this is to use a block matching technique to correlate 2D image blocks in sequential frames via minimization of some cost function (MSE or MAE). Full search or any of fast block matching schemes could be utilized here.

This paper is organized as following. In Section 2, the SWDCT image denoising approach is briefly described. A 3D-SWDCT video denoising algorithm is proposed in Section 3. In Section 4, denoising performance of the proposed algorithm is analyzed in comparison with the recent wavelet-based video denoising algorithms [6], [7], [13]. Conclusions are given in Section 5.

2. Sliding Window DCT Denoising of Images

Sliding window DCT denoising approach is well developed tool for image denoising, (see, e.g. [8] and [9]). In this section we will briefly describe its basic principles. SWDCT denoising scheme is graphically depicted in **Fig. 1**.

Suppose, we wish to recover unknown image $x(t)$ from noisy observations $y(t) = x(t) + n(t)$, where $t = (t_1, t_2)$ are coordinates in 2D space, $n(t)$ is an additive Gaussian noise $N(0, \sigma^2)$ with variance σ^2.

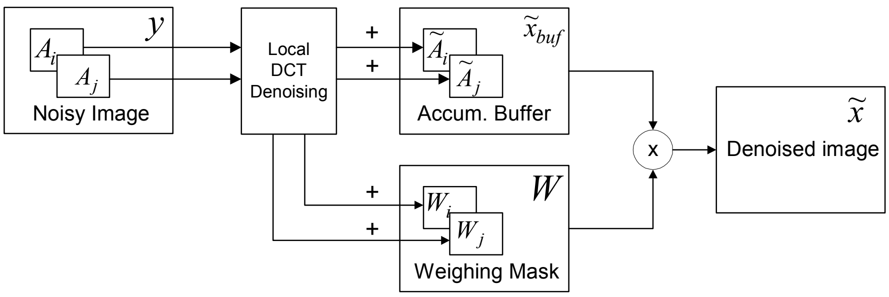


Fig. 1. A general SW-DCT image denoising scheme [9]

Noisy image $y(t)$ is locally processed in the overlapped blocks (windows) $\{A_i\}$. Running over the image each A_i is separately filtered in the DCT domain (computing 2D DCT of $\{A_i\}$, thresholding obtained coefficients and applying an inverse 2D DCT to the result) to obtain a local estimate $\tilde{A}_i$. For every $\tilde{A}_i$ its "relevance" is reflected by a weight W_i evaluated from the local DCT spectrum properties (selected to be a reciprocal of the number of remaining (nonzero) after a threshold DCT coefficients in the block). These estimates $\tilde{A}_i$ and weights W_i are further accumulated in the buffer

$\tilde{x}_{buf}$ and in the weighting mask W, respectively. Finally, every denoised image pixel $\tilde{x}(t)$ is obtained by a weighted average of denoised local estimates of the same pixel from all overlapped estimates $\tilde{A}_i$.
The SWDCT denoising algorithm can be expressed by equations (1)-(4)

$$Y(w) = F\{A_i\}, \tag{1}$$

$$\hat{X}(w) = T\{Y(w)\}, \tag{2}$$

$$\tilde{A}_i = F^{-1}\{\hat{X}(w)\}, \tag{3}$$

$$\tilde{x}(t) = W(t)\tilde{x}_{buf}. \tag{4}$$

where $F\{\ \}$ is a separable 2D forward DCT and $F^{-1}\{\ \}$ is its inverse, $w = (w_1, w_2)$ are coordinates of 2D DCT coefficients and $T\{\ \}$ is a hard thresholding function

$$\hat{X}(w) = \begin{cases} Y(w), & |Y(w)| \geq Thr \\ 0, & \text{else} \end{cases}. \tag{5}$$

The SW-DCT denoising assumes several tunable parameters, such as the local window size and sliding steps along the image directions. They can be user-specified [9] either adaptive to a local signal statistic [12] in order to achieve a better performance/complexity tradeoff.

3. Video Denoising Based on a 3D DCT

The SW-DCT denoising method is well developed for images. In the case of video, SW-DCT should be performed in the 3D space, and the use of a temporal redundancy of video can improve the filtering performance. Let us assume that SW-DCT operates in the spatial domain of each video frame as it is described above. In the temporal direction 1D sliding DCT can be similarly applied along the temporal axis. On the other hand, SW-DCT performance can be significantly improved, if the transform will operate over a highly correlated signal. However, pixels along the temporal axis may be uncorrelated due to dynamical nature of a video signal.
Due to this, we propose to perform a local 3D DCT denoising on an array B_i (of size $L_h \times L_w \times L_t$) that is built from correlated 2D blocks $A_{i,k}$ (k – is an index of the current frame) taken from the L_t consecutive frames. These $A_{i,k}$ blocks are selected using a block matching or similar technique [11]. Here, the full search or a fast block matching scheme via minimization of some cost function (MSE or MAD) can be employed.

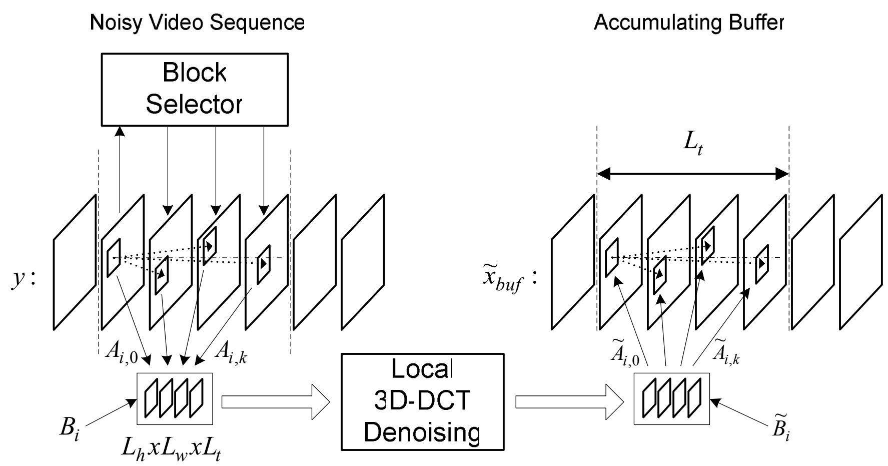


Fig. 2. A general block-diagram of proposed video denoising algorithm

A general scheme of the proposed algorithm for video processing is depicted in **Fig. 2.** A noisy sequence y is processed locally in the 3D windows of size $L_h \times L_w \times L_t$. An accumulating buffer $\tilde{x}_{buf}$ keeps L_t consecutive frames. *Block selector* uses a sliding window $A_{j,0}$ in the 0[th] frame of the buffer as a reference and searches in every sequential frame for the best match $\{A_{i,1},...,A_{i,k}; k = L_t - 1\}$ in terms of their correlation to $A_{i,0}$. These 2D blocks $A_{i,0}$ and $\{A_{i,1},...,A_{i,k} : k = L_t - 1\}$, are filled to the buffer B_i. Note here that it may appear that a block selector will fail to find, in some frame, a subblock $A_{i,k}$ that correlates with $A_{i,0}$. This could appear either due to a dynamic nature of a video or due to a global scene change. In order to prevent an error propagation further, block selection for the current $A_{i,0}$ should be terminated. In such a case a local 3DDCT denoising should be performed on a shorter in the temporal direction array B_i, in other words we implement an adaptive window size selection in the temporal domain. As a result, we have produced a 3D array B_i filled with L_t highly correlated 2D blocks $\{A_{i,0},...,A_{i,k}; k = L_t - 1\}$ which is now a subject of denoising. We retrieve a locally denoised estimate of $\tilde{B}_i$ as a result of hard-thresholding of the 3D-DCT coefficients of B_i and accumulate it in the buffer $\tilde{x}_{buf}$. Amount of the retrieved estimates for a particular $\tilde{x}_{buf}(t)$ and statistical properties of the local DCT spectra define $W(t)$, as it was specified in Section 2.

After the current L_t frames are processed, the sliding window shifts in the temporal direction and a new L_{t+1} frame becomes to be involved in the denoising procedure. Described operations (namely, block matching, local denoising of B_i, accumulation of $\tilde{B}_i$ in $\tilde{x}_{buf}$) are recursively performed on a group of frames until the last frame of the sequence is processed. Finally, every pixel $\tilde{x}(t)$ is reconstructed from a coordinate-wise weighting of $\tilde{x}_{buf}(t)$ with a mask $W(t)$.

The computational complexity of the proposed algorithm is mostly depends on the computation of the 3D DCTs and a block matching procedure performed for every spatial block. If we assume that sizes of a local 3D DCT are L in all directions, and the sliding parameters for both spatial coordinates are P, then the number of arithmetic operations per output sample for the transform part of the algorithm is equal to $2\cdot\mu\cdot\left(\frac{2L}{P^2}+1\right)$, where μ is a complexity of the 1D DCT. If $L=8$ and $P=4$, this number is $4\cdot\mu$. Few operations per output sample should be added to this in the case of application of a fast block matching procedure.

4. Experimental results

To evaluate the performance of the proposed method, several standard test CIF and QCIF video sequences were used, see Tables 1 and 2. Original sequences were corrupted by an additive Gaussian noise with a standard deviations equal to 10, 15 and 20, and then processed with the denoising algorithm proposed in Section 3.

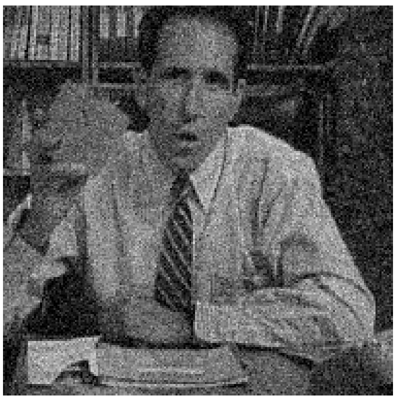
In our simulations we chose processing buffer B_i to be of size of 8x8x8 due to existed and well developed software and hardware solutions for 8-point DCT [10]. Buffer B_i was chosen to be sliding over a video data with the steps equal to 2 in both spatial directions and 1 in the temporal direction. The hard thresholding procedure was applied to all 3D-DCT coefficients of the buffer B_i to get a locally denoised estimate $\tilde{B}_i$. Threshold value *Thr* in the Equation 5 was chosen to be equal to 2σ [8]. To select highly correlated $A_{i,k}$ blocks from 7 consecutive frames, we have used a fast block matching algorithm in the pixel domain (so called "logarithmic search" [11]) with a minimal absolute error (MAD) as a cost function. To prevent error propagation, an adaptive window size selection in temporal domain was performed. The block selection procedure was terminated for a particular $A_{i,0}$ if a correlated $A_{i,k}$ can not be found in the current frame. This improves a filtering performance especially in the presence of high motion or scene change. Furthermore, a selection algorithm that operates in the pixel domain and is based on MAD or MSE criteria may provide $A_{i,k}$ correlated rather with a noise pattern of the reference $A_{i,0}$ than with the original video signal. This could become a problem in video fragments with a very low signal to noise ratios, for example, dark flat regions corrupted with heavy noise. To suppress such false motion prediction we rejected (set to zero) motion vectors if MAD of a prediction is lower than a predefined threshold within the distance of 3σ.

In **Table 1,** a performance of our algorithm is compared with the results of wavelet-based denoising schemes of [6] and [7]. The average PSNR values presented in **Table 1** are computed over 40 ("Salesman") and 52 ("Tennis" and "Flower Garden") frames. The first four frames of processed sequences are excluded from PSNR calculation due to the recursive nature of the WRSTF algorithm [7] in order to make the comparison more objective.

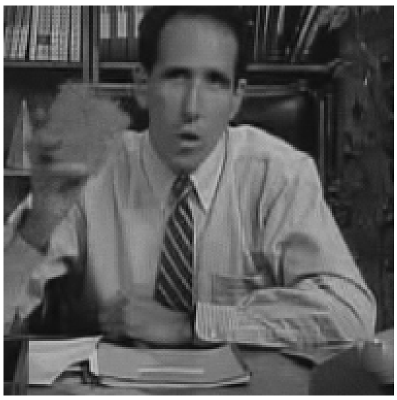
Table 1. Video denoising, comparative results

Video	Noise, σ	Average PSNR, dBs			
		Noisy	Soft3D [6]	WRSTF [7]	3D SWDCT
"Tennis"	10	28.16	31.86	32.41	**33.34**
	15	24.63	29.86	30.12	**30.80**
	20	22.15	28.58	28.68	**29.52**
"Salesman"	10	28.15	34.85	35.82	**37.01**
	15	24.72	33.29	33.91	**34.83**
	20	22.35	32.00	32.40	**33.29**
"Flower"	10	28.34	30.23	30.80	**31.25**
	15	24.88	27.71	28.19	**28.62**
	20	22.44	26.01	26.39	**26.80**

To compare performance of the proposed algorithm with the results reported in [13], we have applied our algorithm to "Miss America" and "Hall" video sequences corrupted with additive Gaussian noise (average PSNR of noisy video are 20 dBs). Results are shown in **Table 2**. Analysis of **Tables 1** and **2** demonstrates that our algorithm outperforms those from [6], [7] and [13]. **Fig 3.** and **4** give some examples of denoised frames to subjective judgment of visual quality of denoised video sequences.



(a)



(b)

Fig. 3. A fragment of the 30th frame of the "Salesman" video sequence. (a) Noisy (PSNR of fragment 22.29 dBs). (b) Denoised with the proposed algorithm (PSNR of fragment 33.06 dBs)

(a)

(b)

Fig. 4. A fragment of the 30th frame of the "Flower" video sequence. (a) Noisy (PSNR of fragment 22.41dBs). (b) Denoised with the proposed algorithm (PSNR of fragment 26.39 dBs)

Table 2. Video denoising, comparative results

Video	Average PSNR, dBs		
	Noisy	Proposed in [13]	3D SWDCT
"Miss America"	20	34.1	**34.9**
"Hall"	20	29.1	**31.8**

Detailed information on the developed algorithm and video sequences processed by 3D-SWDCT are available from: http://www.cs.tut.fi/~rusanovs/.

5. Conclusions

A problem of denoising of video signals corrupted by an additive Gaussian noise is considered in this paper. A novel 3D DCT based video denoising algorithm is proposed. High filtering performance of the local 3D DCT based thresholding is achieved by a proper selection of video volume data to be locally denoised. A 3D DCT thresholding is performed on a group of highly correlated sliding in spatial directions 2D windows that are selected from the set of sequential frames. Weighted average of overlapped denoised estimates provides a final denoised video. We have tested the proposed algorithm on a group of standard video test sequences corrupted by an additive Gaussian noise with a variety of standard deviations. Results have demonstrated that the proposed algorithm provides competitive results with wavelet-based video denoising methods both in terms of PSNR and subjectively quality.

Acknowledgements

This work has been financially supported by the Academy of Finland, Finnish Center of Excellence Programme 2000-2005, project no. 5202853, as well as EU project NoE FP6-PLT 511568-3DTV.
The authors would like to thank V. Zlokolica and I. Selesnick for providing processed video sequences resulting from applications of their denoising algorithms (presented in Table 1).

References

1. J.C. Brailean, R.P. Kleihorst, S.Efstratiadis, A.K.Katsaggelos, R.L.Lagendijk, "Noise Reduction Filters for Dynamic Image Sequences: A Review", IEEE Proc, Vol. 83, no. 9, Sep, 1995
2. D.L.Donoho, "De-noising by soft-thresholding", *IEEE Trans. on Information Theory*, vol.41, no.3, pp.613-627, May 1995
3. L.Sendur, I.W.Selesnick, "Bivariage Shrinkage Functions for Wavelet-Based Denoising Exploiting Interscale Dependency", *IEEE Trans. on Signal Proc.*, vol. 50, n. 11, pp. 2745-2756, November 2002
4. R.Coifman,D.Donoho, "Translation Invariant de-noising", in Lecture Notes in Statistics: Wavelets and Statistics, vol. New York: Springer-Verlag, pp.125-150, 1995
5. N.Kingsbury, "Complex Wavelets and Shift Invariance", available by the URL: http://ece-www.colorado.edu/~fmeyer/Classes/ECE-5022/Projects/kingsbury1.pdf
6. W.I. Selesnick and K.Y. Li, "Video denoising using 2d and 3d dualtree complex wavelet transforms", in *Wavelet Applications in Signal and Image Processing (Proc. SPIE 5207), San Diego*, Aug. 2003.
7. V. Zlokolica, A. Pizurica, W. Philips "Wavelet Domain Noise-Robust Motion Estimation and Noise Estimation for Video Denoising " *First International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, Scotssdale, Arizona, 23-25 January, 2005, USA
8. L. Yaroslavsky and K. Egiazarian and J.Astola, "Transform domain image restoration methods: review, comparison and interpretation", *TICSP Series #9*, TUT, Tampere, Finland, December 2000, ISBN 952-15-0471-4.
9. R. Öktem, L. Yaroslavsky and K. Egiazarian, "Signal and Image Denoising in Transform Domain and Wavelet Shrinkage: A Comparative Study", *in Proc. of EUSIPCO'98*, Sept. 1998, Rhodes, Greece.
10. K. Rao, P.Yip, *Discrete Cosine Transform: Algorithm, Advantages, Applications*, Academic Press, 1990.
11. R. J. Clarke, *Digital Compression of Still Images and Video*, Academic Press, 1995.
12. K. Egiazarian, V. Katkovnik, H. Öktem and J. Astola, "Transform-based denoising with window size adaptive to unknown smoothness of the signal", *In Proc. of First International Workshop on Spectral Techniques and Logic Design for Future Digital Systems (SPECLOG)*, June 2000, Tampere, Finland, pp. 409-430.
13. N. Gupta, E. Plotkin, M. Swamy, "Bayesian Algorithm for Video Noise Reduction in the Wavelet Domain", IEEE International Symposium on Circuits and Systems, ISCAS 2005, May 23-26, Kobe, Japan