An Improved Scheme for Combined Wavelet Domain and Temporal Filtering

Ljubomir Jovanov, Aleksandra Pižurica, Vladimir Zlokolica and Wilfried Philips

Abstract—**In this paper we propose an improved video denoising scheme that combines recursive temporal filtering and wavelet domain spatial denoising. In our previous work, we introduced a sequential scheme (SEQWT) where a wavelet domain spatial filter is followed by a motion detector and by a selective recursive temporal filter. This scheme is efficient but its limitation is the lack of motion compensation.**

In this paper, we introduce the idea of using motion estimation resources from wavelet video codec for video denoising. The benefit is twofold: Firstly, we improve the performance of the SEQWT video denoiser by using a motion estimator that is suitable for real-time processing. Secondly, we make a first step towards integrating wavelet domain video coder and denoiser by making them sharing the common resources such as motion estimation. Note that this is not straightforward though. The motion estimators aimed for video compression and coding, tolerate errors in the estimated motion field and hence are not directly applicable to video denoising. To solve this problem, we propose a novel motion field filtering step that refines the accuracy of the motion estimates to a degree that is required for denoising.

The main contributions of this paper are the following. (i) We introduce a novel motion field filtering step that improves significantly the mean squared error performance of the multiresolution motion estimator. (ii) We propose an original scheme where a video denoiser reuses motion estimation resources of a video coder with the added motion estimation refinement step. (iii) By using the refined motion estimator within the proposed scheme, we improve significantly the filtering performance over the SEQWT filter. The increase in peak signal to noise ratio (PSNR) ranges from 0.5dB to 0.8 dB depending on the test sequence.

Keywords—**Video denoising, wavelets, motion vector estimation, Bayesian estimation.**

I. Introduction

In the last few decades television became the most important media and its influence in modern society is still growing. Besides its role as a source of information, other important applications of video are surveillance, object tracking, medical and astronomical imaging.

Video recording systems and transmission means are not perfect, and as such introduce degradations of recorded video sequences. Among various types of noise, the most prevalent ones are the noise introduced by the camera (photon noise), shot noise (caused by the electronic hardware) and the channel (thermal) noise [1]. Adequate model for majority of noise sources is additive white Gaussian noise model, which is also treated in this paper. The level of noise in an image is most often specified in terms of signal-to-noise ratio (SNR). Besides degradation of visual quality, the noise patterns may mask important image detail and increase the entropy of the image, which can decrease the degree of effective compression. SNR of a video can be improved by spatio-temporal or 3-D filtering, which exploits image and noise modelling. Numerous existing approaches make compromises between computational complexity and performance. One advantage of video signal over still images is that video contains temporal redundancy, which can be successfully employed in denosing algorithms, using motion detection/estimation algorithms.

Denoising algorithm presented in this work, belongs to group of multiresolution video denoising algorithms. Our previous method [2] combines spatially adaptive wavelet thresholding of non-decimated wavelet coefficients and recursive pixel based temporal filtering based on motion detection algorithm. Main motivation for the sequential wavelet domain and temporal motion compensated filtering is the following. Spatially adaptive 2-D wavelet thresholding methods have achieved impressive results in still image denoising. Main drawback of these methods is their computational complexity, which makes them less attractive for real-time applications. Recently authors of [3] showed that it is possible to perform spatially adaptive wavelet thresholding with computational complexity and memory requirements acceptable for real time hardware implementation in FPGA with negligible loss in denoising performance. Also motion vector

Lj. Jovanov (Ljubomir.Jovanov@telin.UGent.be), A. Pižurica (Aleksandra.Pizurica@telin.UGent.be), V. Zlokolica (vzlokoli@telin.ugent.be) and W. Philips(philips@telin.UGent.be) are with the Department for Telecommunications and Information Processing (TELIN), Ghent University, Sint-Pietersnieuwstraat 41, B-9000 Gent, Belgium. Tel:+32 9 264 34 12, Fax: +32 9 264 42.

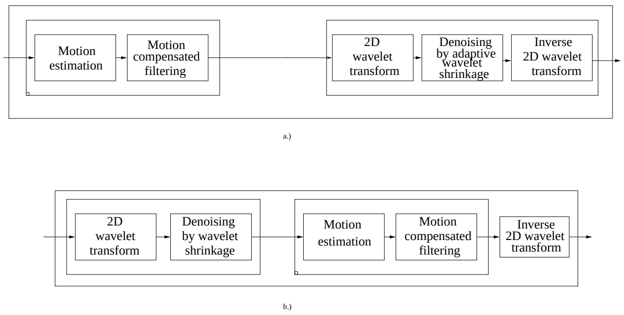

Fig. 1. Tested video denoising schemes.

field estimation approach presented in [4] is suitable for motion hardware implementation. VLSI implementations of this algorithm are described in the references [5] [6].

In this paper we describe multiple variants of denoising algorithms, concerning order and type of filtering. Two variants are tested: spatially adaptive filtering followed by temporal filtering, temporal filtering followed by spatially adaptive filtering using motion vectors estimated on noisy frames. We have used spatial wavelet filtering with noise variance σ_n fixed for all subbands. Tested video denoising schemes are depicted in Fig. 1. We test the method on four different video sequences corrupted by different amounts of additive white Gaussian noise. The results demonstrate that the proposed filter outperforms our previous method based on motion detection, while being compatible with video coding motion estimation algorithms. The paper is organized as follows. Section 2 explains two considered wavelet based denoising approaches. Section 3 addresses temporal filtering, performed before or after spatial denoising and the joint parameter optimisation. The results are presented and discussed in Section 4. Conclusions are given in Section 5.

II. 2-D Wavelet domain noise filtering

Wavelet coefficients of natural noise-free images can be precisely modeled by generalized Laplacian (or generalized Gaussian) probability density, since the histograms of wavelet coefficients in each subband are typically long-tailed and sharply peaked at zero. Analytical form of generalized Laplacian density can be written as:

$$p(y) = \frac{\lambda\nu}{2\Gamma(\frac{1}{\nu})} exp(-\lambda|y|^{\nu}), \quad \lambda,\ \nu > 0, \tag{1}$$

where $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$ is the Gamma function. Typical values for the shape parameter ν, for natural images are in the interval [0,1]. It can be shown [7] that the variance and the courtosis of a generalized Laplacian distribution are:

$$\sigma_y^2 = \frac{\Gamma(\frac{3}{\nu})}{\lambda^2\Gamma(\frac{1}{\nu})}, \quad \kappa_y = \frac{\Gamma(\frac{1}{\nu})\Gamma(\frac{5}{\nu})}{\Gamma^2(\frac{3}{\nu})} \tag{2}$$

When the image is contaminated with the additive white Gaussian noise, the model parameters ν and λ are estimated from the noisy coefficient histogram using the equations [7] [8]:

$$\begin{aligned} \frac{\Gamma(\frac{1}{\nu})\Gamma(\frac{5}{\nu})}{\Gamma^2(\frac{3}{\nu})} &= \frac{m_4 + 3\sigma^4 - 6\sigma^2\sigma_w^2}{(\sigma_w^2 - \sigma^2)^2}, \\ \lambda &= ((\sigma_w^2 - \sigma^2)\frac{\Gamma(\frac{1}{\nu})}{\Gamma(\frac{3}{\nu})})^{-\frac{1}{2}}, \end{aligned} \tag{3}$$

where σ_w^2 and $m_{4,w}$ are the variance and the fourth moment of the noisy histogram, respectively. For the value ν=1, generalized Laplacian density degenerates to Laplacian density, which is used quite often due its analytical tractability. For Laplacian density, scale parameter can be estimated as:

$$\lambda = [0.5(\sigma_w^2 - \sigma^2)]^{-\frac{1}{2}} \tag{4}$$

This simplified model usually does not produce a significant degradation in denoising performance. Wavelet coefficient estimator can be defined [2] as:

$$\hat{y}_l = \frac{\rho\xi_l\eta_l}{1 + \rho\xi_l\eta_l} w_l, \tag{5}$$

where

$$\rho = \frac{P(H_1)}{P(H_0)}, \quad \xi_l = \frac{p(w_l|H_1)}{p(w_l|H_0)}, \quad and \quad \eta_l = \frac{p(z_l|H_1)}{p(z_l|H_0)}. \tag{6}$$

H_1 denotes the hypothesis that the current wavelet coefficient contains a significant noise-free component, and H_0 denotes the opposite hypothesis and $p(w_l|H_0)$ and $p(w_l|H_1)$ denote the conditional probability density functions of the noisy coefficients given the absence and the presence of a signal of interest respectively. Probabilities $p(z_l|H_0)$ and $p(z_l|H_1)$ denote the conditional probabilities of the local spatial activity indicator. Parameters ρ, ξ_l, η_l are estimated from the observed image coefficients. For the prior (1), where ν=1, the conditional densities of *noise-free* coefficients are

$$p(y|H_0) = \begin{cases} A_0 exp(-\lambda|y|^{\nu}) & \text{if } y \le T \\ 0 & \text{if } y > T \end{cases} \tag{7}$$

$$p(y|H_1) = \begin{cases} 0 & \text{if } y \leq T \\ A_1 exp(-\lambda|y|^\nu & \text{if } y > T \end{cases} \quad (8)$$

where it is easy to show that $A_0 = (\lambda/2)e^{\lambda T}/(e^{\lambda T - 1})$ and $A_1 = (\lambda/2)e^{\lambda T}$. In this paper we adopt additive white noise model $w = y + \epsilon$, where $\epsilon\ N(0, \sigma^2)$. The densities of noisy cofficients $p(w|H_0)$ and $p(w|H_1)$ can be derived as convolution of normal density random variable $N(0, \sigma^2)$ with $p(y|H_0)$ and $p(y|H_1)$ respectively. This procedure is ussualy carried out numerically. Analytical expressions were derived in [2]. Statistical characterization of z_l can be simplified by assuming that all the coefficients within the small window are equally distributed and conditionally independent. Using these assumptions, and denoting the cofficient magnitude by $m_l = |w_l|$, we have that $f(Nz_l|H_1)$ is given by N convolutions of $f(m_l|H_1)$ with itself and $f(Nz_0|H_0)$ is given by N convolutions of $f(m_0|H_0)$ with itself, where $p(m_l|H_{0,1}) = 2p(w_l|H_{0,1}),\ m_l > 0$.

Prior ratio $\rho = P(H_1)/P(H_0)$ can be estimated as follows. From expression $P(H_1) = \int_{\infty}^{-T} p(y)dy + \int_{T}^{\infty} p(y)dy$, for the prior (1), we derive [2,8]

$$\rho = \frac{P(H_1)}{P(H_0)} = \frac{1 - \Gamma_{inc}((\lambda T)^\nu, \frac{1}{\nu})}{\Gamma_{inc}((\lambda T)^\nu, \frac{1}{\nu})}, \quad (9)$$

where $\Gamma_{inc}(x, a) = \frac{1}{\Gamma(a)} \int_0^x t^{a-1}e^{-t}dt$ is the *incomplete gamma function.* For $\nu = 1$ prior becomes Laplacian, and the previous expression reduces to

$$\rho = \frac{P(H_1)}{P(H_0)} = \frac{exp(-\lambda T)}{1 - exp(-\lambda T)}. \quad (10)$$

Optimal value of the threshold in the mean squared error sense is $T = \sigma$, where the proof is given in [2,8].

Method presented in this paper is implemented with the generalized Laplacian prior. Minor performance degradations are introduced when using Laplacian prior $\nu = 1$. Ussualy, the peak signal to noise ratio drops for 0.1 to 0.3 dB. Second denoising method tested in this paper [3] is more appropriate for hardware implementation.

III. Motion compensated temporal filter

It is well known that spatial denoising, if performed alone, produces annoying artifacts and unsatisfactory video quality [1]. These artefacts can be observed even when the sophisticated wavelet domain methods are used for spatial filtering. Annoying effects are caused by the fact that the residual noise and denoising artefacts differ from frame to frame, which can be observed as unpleasant "flickering" effect. In the proposed denoising scheme we have tested multiple different configurations. First of them performs spatial filtering followed by motion compensated temporal filtering, where the motion vectors are estimated based on spatially denoised frames. Second configuration performs motion compensated temporal filtering first. Motion vectors in this second method are estimated using either noisy or spatialy filtered frames.

As we mentioned earlier, in the configuration which yields best results, motion vectors are estimated based on spatially filtered frames. Motion vector field estimated on noisy sequence is not significantly different from motion field estimated on spatially filtered frames. This step is followed by motion compensated filtering. Final step is spatially adaptive wavelet domain filtering. Spatially adaptive wavelet domain filtering is performed by using wavelet domain shrinkage based on generalized Laplacian prior. Prior to Bayesian shrinkage, we perform noise variance estimation. Motion estimation algorithm employed in this denoising algorithm employs motion detector in order to remove spurious motion vectors.

Algorithm described in this paper performs filtering with different sets of coefficient, depending on detected motion. A formal description follows.

If we denote the k-th frame of a *noise-free* video sequence with f^k and noise field with n^k, noisy video frame can be expressed as $d^k = f^k + n^k$. 2-D denoised k-th frame is denoted as

$$\hat{f}^{2D,k} = [\hat{f}_1^{2D,k}\ , ..., \hat{f}_L^{2D,k}] \quad (11)$$

Temporal filtering is performed in the parts of the sequence, where motion field is defined (where significant motion exists). Differences in denoising performance when motion field is calculated using denoised and noisy frames are very small. Positions in the frame which belong to the area where no motion was detected, are filtered using time averaging between two neigbouring frames $\hat{f}^{tf,k} = \alpha f^k + (1 - \alpha) f^{k-1}$, where $0 \leq \alpha \leq 1$. At positions where motion is detected, we perform motion compensated temporal filtering $\hat{f}^{tf,k} = \beta f^{k,p} + (1 - \beta) f^{k-1,p-p_m}$, where p_m denotes motion vector in k-th frame. Finally, we get expression for the temporally filtered frame

$$\begin{aligned} \hat{f}_l^{temp,k} = (1 - m_l^k)[\alpha f^k + (1 - \alpha) f^{k-1}] + \\ m_l^k [\beta f^{k,p} + (1 - \beta) f^{k-1,p-p_m}] \end{aligned} \quad (12)$$

The last filtering step is spattialy adaptive filter-

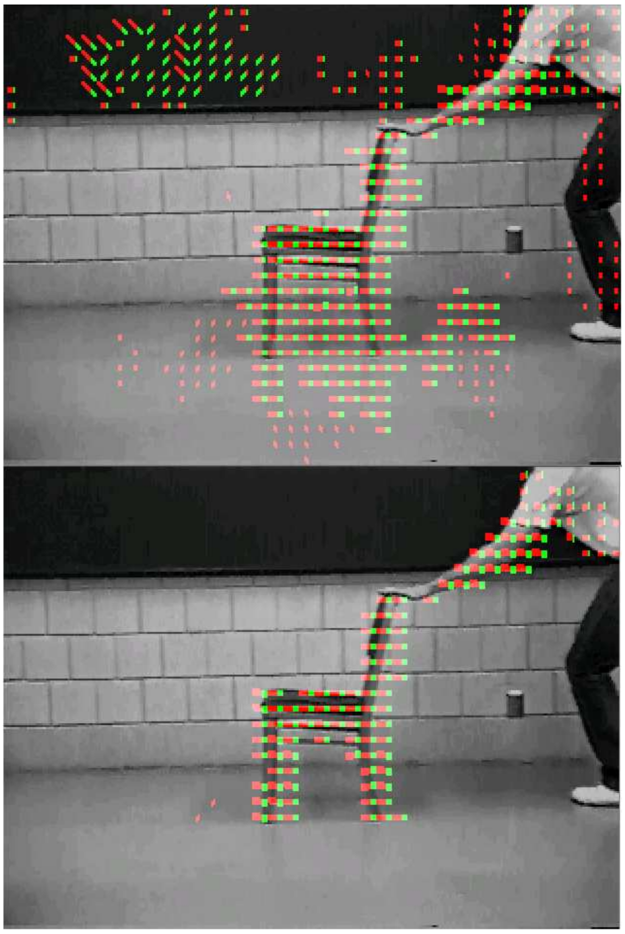

Fig. 2. Motion field for 15-th frame of 'chair' sequence, before and after filtering.

ing, where the filtering is performed with reduced, but fixed σ.

IV. Multiresolution motion estimation algorithm

Motion estimation algorithm used in this paper belongs to the group of the motion estimation algorithm which are used for video compression. This means that the algorithm tries to find best matching parts in the previous frame, no matter if the motion vectors obtained in this way do not have meaningful interpretation i.e. represent some objects. Because of that it is neccessary to perform additional filtering steps, in order to make it possible to use this algorithm for denoising purposes.

This algorithm belongs to group of multiresolution algorithms, where video frame is decomposed into different resolutions. Motion vector estimation is first performed on the coarsest resolution level, and then refined based on the information from finer resolution scales. Besides information from multiple resolution levels, algorithm described here exploits spatial and temporal correlations in motion field to obtain final result. Detailed description of the motion algorithm used in this paper is given in [4].

A. Motion field refinement step

We introduce additional filtering step in order to improve the quality of the motion field and video denoising performance. Motion field produced by algorithm [4] is aimed for applications in video coding, and therefore does not produce meaningful motion field, which follows the motion of objects in scene. Motion estimation algorithms used in video coding just try to find the best matching pixel value in previous frame no matter if it does not belong to the same object. Such motion field cannot be used for video denoising without further refinement, since they would introduce degradations of stationary parts of the scene. Motion field for 15-th frame of 'chair' sequence before and after filtering is shown in Fig. 2. The filtering step can be described as follows.

In the first step we calculate difference between blocks in neigbouring frames, denoted with $D^k_{i,j}$:

$$D^k_{i,j} = \frac{1}{N^2}\sum_{m=1}^{N}\sum_{n=1}^{N}|d^{k,i,j}_{m,n} - d^{k-1,i,j}_{m,n}|, \qquad (13)$$

where i, j are spatial coordinates of a block, m, n coordinates of a pixel inside block used for motion estimation and N is a block size.

We define threshold used for motion detection in k-th frame as follows:

$$THR = \gamma\frac{1}{N_{bx}N_{by}}\sum_{i=1}^{N_{bx}}\sum_{j=1}^{N_{by}} D^k_{i,j}, \qquad (14)$$

where γ is scalar whose value is chosen to get optimal denoising performance and N_{bx}, N_{by} are number of blocks along x and y axis.

In the filtering step, we make decision whether motion exists in each block based on comparison of the absolute block difference with the previously calculated threshold. If the absolute difference is less than threshold, both motion vector components are set to zero. Otherwise, motion vector keeps its original value.

V. Results

This section is divided into two subsections: first analyses performance of motion estimation algorithm and the second analyses performance of denoising algorithm.

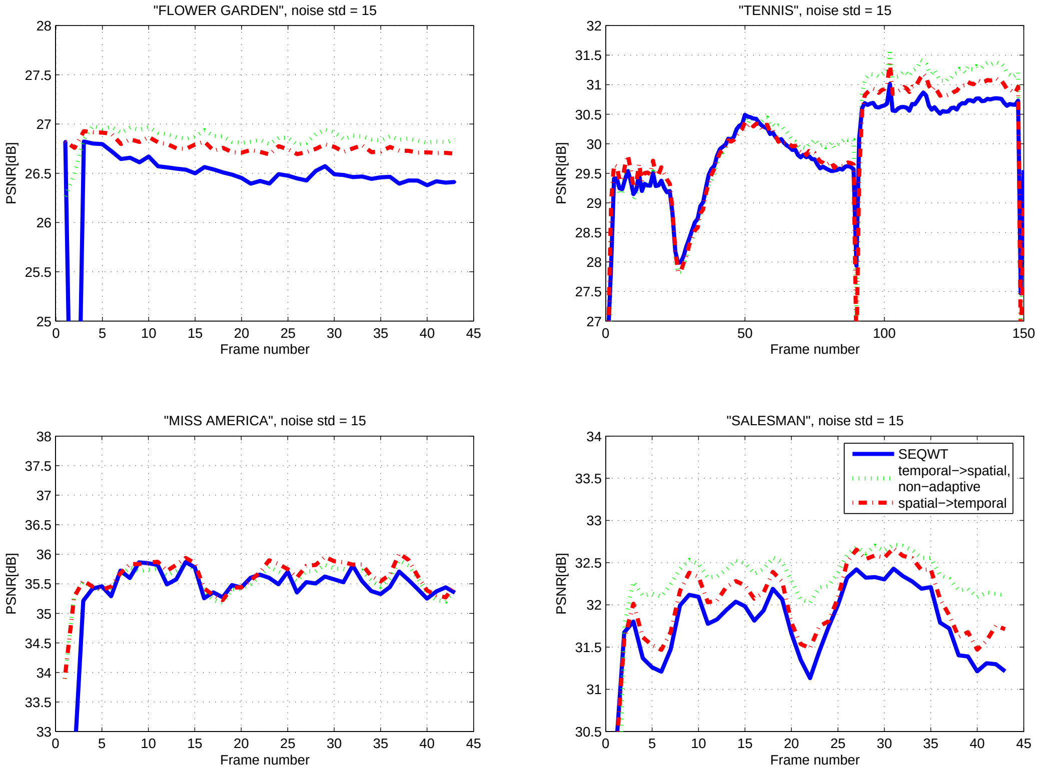

Fig. 3. Quantitative performance comparison of two variants of proposed method and SEQWT algorithm for the following sequences: (a) flower, σ =15 (b) tennis, σ=15 (c) miss America, σ = 15 (d) salesman, σ = 15

A. *Motion estimation performance*

Performance of the algorithm is evaluated by comparing mean square error of the motion compensated frame, obtained using estimated motion field with and without motion field filtering step.

TABLE I

MEAN SQUARE ERROR OF THE MOTION FIELD

Sequence	Without MV filtering	With MV filtering
"salesman"	0.188	0.118
"chair"	0.077	0.047
"miss America"	0.062	0.036
"tennis"	0.017	0.017
"flower garden"	0.49	0.345
"bus"	0.286	0.202

Mean square error of the motion field is defined as

$$MSE = \frac{1}{N_x N_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} (f_{i,j}^{k} - f_{i-m_x,j-m_y}^{k-1})^2, \quad (15)$$

N_x, N_y are the sizes of image frame along x and y axis, i, j coordinates of pixel inside image frame and m_x, m_y x and y components of the motion vector. Mean square errors of the described motion estimation methods are given in Table. I. The main criterion for additional filtering step optimization was improvement of denoising performance.

B. *Denoising results*

In this section we present the results of the proposed method with spatially adaptive wavelet shrinkage with fixed σ. Wavelet function used here is *symmlet* with eight vanishing moments. All results were obtained using non-decimated wavelet transform.

Four test sequences were used, namely *Miss America, salesman, tennis and flower*, each of which was corrupted by additive white Gaussian noise with standard deviations σ = 10, 15 and 20. Resulting PSNR

Fig. 4. (a) Noisy 24-th frame of the salesman sequence $\sigma = 15$. (b) Resulting frame after filtering with SEQWT filter. (c) Resulting frame after filtering with proposed filter. (d) Noise-free frame

of the proposed method is compared with the method presented in [2] which is illustrated in Fig. 3. Comparison with other methods can be done by consulting reference [2].

Denoising method which uses temporal filtering followed by spatial filtering, shows better results for all test sequences, with the PSNR improvements ranging from 0.5 - 0.8dB depending on the sequence. These improvements are possible mainly because method includes motion compensated temporal filtering. In the other sequences, such as "Miss America" and the parts of the "tennis" sequence which do not contain translatory movements algorithm performs similar as SEQWT algorithm. In the first experiment, sequences were first motion-compensated temporaly filtered and then spatially adaptive filtering was done, with fixed σ. In the second experiment we first perform spatially adaptive filtering with fixed σ, and then temporal filtering.

Proposed method also outperforms previous method in terms of visual quality, especially in the sequences which contain more dynamics. Denoised video frames are shown in Fig. 4. Differences between presented methods are even more percieveable after viewing and comparing denoised sequences.

VI. Conclusion

This paper has multiple contributions. First of them is novel filtering step, which makes possible usage of motion estimation algorithm, primary meant for video coding, for video denoising. This filtering step improves performance of the motion estimation algorithm in MSE sense of the motion compensated frame. Filtering step itself is very simple, and easy implementable in VLSI, since it contains only basic mathematical operations. Second major contribution is improvement of denoising performance of our previous denoising algorithm. Above mentioned motion estimation algorithm has its numerous VLSI implementations. Combined with low-complexity motion field filtering step and spatial denoising scheme from [3], this configuration can be easily implemented as a part of hardware video codec. Denoising scheme presented here outperforms our previous method in the terms of PSNR and visual quality.

Further improvements are expected from using more sophisticated motion estimation algorithm. Another source of improvements could be more sophisticated noise and image statistical modelling.

References

[1] J. Brailean, R. Kleinhorst, S. Efstratiadis, A. Katsaggelos, and A. Lagendijk, "Noise reduction filters for dynamic image sequences: A review," *Proceedings of the IEEE*, vol. 83, no. 9, pp. 1272–1292, 1995.

[2] A. Pizurica, V. Zlokolica and W. Philips, "Noise reduction in video sequences using wavelet-domain and temporal filtering," in *SPIE Symposium on Electronic Imaging*, San Jose, USA, 2003.

[3] M. Katona, A. Pižurica, N. Teslic, V. Kovacevic, and W. Philips, "Fpga design and implementation of a wavelet-domain video denoising system," *Lecture Notes in Computer Science*, vol. 3708, pp. 650–657, 2005.

[4] J. Chalidabhongse and C.-C. Jay Kuo, "Fast motion vector estimation using multiresolution-spatio-temporal correlations," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 7, no. 3, pp. 477–488, 1997.

[5] J. Hsieh and T. Meng, "Low-power parallel video compression architecture for a single-chip digital cmos camera," *Journal of VLSI Signal Processing Systems*, vol. 21, no. 11, pp. 195 –207, July 1999.

[6] B. Kim and C. J., "Vlsi architecture for low power motion motion estimation using high data access reuse," in *Proceedings of AP ASIC Conference*, Seoul, Korea, 1999.

[7] E. P. Simoncelli and E. H. Adelson, "Noise removal via Bayesian wavelet coring," in *Proc. IEEE Internat. Conf. Image Proc. ICIP*, Lausanne, Switzerland, 1996, 1996.

[8] A. Pižurica, W. Philips, I. Lemahieu, and M. Acheroy, "A joint inter- and intrascale statistical model for wavelet based Bayesian image denoising," *IEEE Trans. Image Proc*, vol. 11, no. 5, pp. 545–557, May 2002.