VIDEO DENOISING USING A SPATIOTEMPORAL STATISTICAL MODEL OF WAVELET COEFFICIENTS

Gijesh Varghese[1] *and* *Zhou Wang*[2]

[1]Mobilygen Corporation, Santa Clara, CA, USA
[2]Dept. of Electrical & Computer Engineering, University of Waterloo, Waterloo, ON, Canada
Email: gijesh.varghese@ieee.org, zhouwang@ieee.org

ABSTRACT

We propose a video denoising algorithm based on a spatiotemporal Gaussian scale mixture (ST-GSM) model in the wavelet transform domain. This model simultaneously captures local correlations between the wavelet coefficients of natural video sequences across both space and time. Bayes least square estimation is used to recover the original signal from the noisy observation. To further improve the performance, motion compensation is employed before ST-GSM denoising, where a Fourier domain noise-robust cross correlation approach is proposed for motion estimation. Experiments show that the performance of the proposed method is highly competitive when compared with state-of-the-art video denoising algorithms.

Index Terms— video signal processing, video denoising, statistical image modeling, image restoration, motion estimation

1. INTRODUCTION

Video signals are often contaminated by noise during acquisition and transmission. Denoising of video is highly desirable, which can enhance perceptual quality, increase compression effectiveness, facilitate transmission bandwidth reduction, and improve the accuracy of the subsequent feature extraction and pattern recognition processes.

Most video denoising algorithms proposed in the literature consider additive white Gaussian noise and can be roughly categorized into pixel domain and transform domain methods. Pixel domain denoising is usually done with weighted averaging within local 3D windows, where the weights can either be fixed or adapted based on the local image content. Improved results are obtained by applying weighted averaging after motion compensation, so that temporal correlation is taken better into account. A review of pixel domain method can be found in [1].

Transform domain methods first decorrelate the video signal using a linear transform, and then denoise the signal in the transform domain (e.g., by soft/hard thresholding, maximum likelihood estimation, or Bayesian estimation), followed by an inverse transform that brings the signal back to the pixel domain. Motion information or temporal correlations may be incorporated into the algorithms by employing an advanced or adapted transform [2, 3] or by using an advanced statistical model that reflects the joint distributions of wavelet coefficients over space and time [4–6]. Recursive filtering method has also been proposed to filter the wavelet coefficients along the estimated motion trajectories [7].

The interest in statistical image modeling has been growing steadily in recent years. Here we are specifically interested in the Gaussian scale mixture (GSM) model, which was originally proposed in the statistics literature and has recently been applied to the modeling of static natural images [8]. Indeed, it was found to be both effective and convenient in describing the marginal and joint statistics of wavelet coefficients. Bayesian estimation method based on the GSM model has also been developed for the denoising of still images, and has achieved superior performance [8].

In this paper we extends the general GSM idea for the modeling and denoising of video. In particular, we propose to use a spatiotemporal GSM (ST-GSM) model that simultaneously captures local correlations between the wavelet coefficients of natural video sequences across both space and time. Moreover, we find that applying a motion compensation process beforehand can further improve the performance of ST-GSM based denoising. We have also developed and incorporated a new noise-robust motion estimator for motion compensation.

2. METHOD

2.1. Denoising Based on GSM Models

A random vector $\mathbf{x}$ is a GSM if it can be expressed as the product of two independent components:

$$\mathbf{x} = \sqrt{z}\mathbf{u}, \tag{1}$$

where $\mathbf{u}$ is a zero-mean Gaussian vector, and z is called a mixing multiplier. The density of $\mathbf{x}$ is then given by:

$$p_x(\mathbf{x}) = \int \frac{1}{[2\pi]^{N/2}|z\mathbf{C_u}|^{1/2}} \exp(-\frac{\mathbf{x}^T(z\mathbf{C_u})^{-1}\mathbf{x}}{2})p_z(z)dz \tag{2}$$

where $\mathbf{C_u}$ is the covariance matrix of $\mathbf{u}$, $p_z(z)$ is the mixing density, and N is the size of the vector $\mathbf{x}$. It has been found that a GSM model can well account for 1) the marginal distributions of the wavelet coefficients computed from natural images, and 2) the strong correlations between the amplitudes of neighboring wavelet coefficients [8].

Now assume that $\mathbf{x}$ is a vector of neighboring wavelet coefficients of the original image, then a noise corrupted coefficient vector $\mathbf{y}$ can be written as

$$\mathbf{y} = \mathbf{x} + \mathbf{w} = \sqrt{z}\mathbf{u} + \mathbf{w}\,, \tag{3}$$

where $\mathbf{w}$ is an additive Gaussian noise coefficient vector with a covariance matrix $\mathbf{C_w}$ (here we have assumed that the original noise contamination in the image domain is additive independent Gaussian). The group of neighboring coefficients constitute a sliding window that moves across the wavelet subband. At each step, only the center coefficient of the window is estimated (i.e., denoised). Therefore, the objective here is converted to estimating the center coefficient x_c of $\mathbf{x}$, given the noisy observation $\mathbf{y}$.

The Bayes least square (BLS) estimator is simply the conditional mean, which can be computed by [8]

$$E\{x_c|\mathbf{y}\} = \int_0^\infty p(z|\mathbf{y})E\{x_c|\mathbf{y}, z\}dz\,. \tag{4}$$

On the right hand side, $E\{x_c|\mathbf{y}, z\}$ is linear based on the facts that $\mathbf{w}$ is Gaussian and $\mathbf{x}$ is also Gaussian when conditioned on z. The posterior density of $p(z|\mathbf{y})$ can also be estimated by Bayes' rule, provided that the prior $p_z(z)$ of the multiplier is given. The integral in Eq. (4) is evaluated with a 1D numerical integration. More implementation details as well as further discussions can be found in [8].

2.2. Spatiotemporal GSM denoising

The general approach of GSM denoising described above can have many variations, depending on how the neighboring wavelet coefficient vector is formed. Since natural video signals have strong correlations between adjacent frames, it is useful to include temporal neighborhood coefficients in the vector. In addition, when the adjacent frames are properly aligned, the temporal correlation becomes even stronger. Therefore, we apply motion compensation before the formation of the neighboring wavelet coefficient vector and the subsequent ST-GSM denoising process.

The diagram of the proposed denoising algorithm is illustrated in Fig. 1. The denoising of the current frame involves not only the frame itself, but also a set of adjacent past and future frames. Motion estimation is performed between the current frame and the past/future frames. The estimation results are then used for motion compensation (simply by spatial translation). Wavelet transform is then applied to the current frame as well as the motion compensated past and future frames. Next, wavelet coefficient vectors are formed from a spatiotemporal neighborhood, and an ST-GSM denoising method similar to the BLS estimator discussed earlier is employed. Finally, an inverse wavelet transform is applied to the denoised wavelet coefficients and a denoised current frame is created.

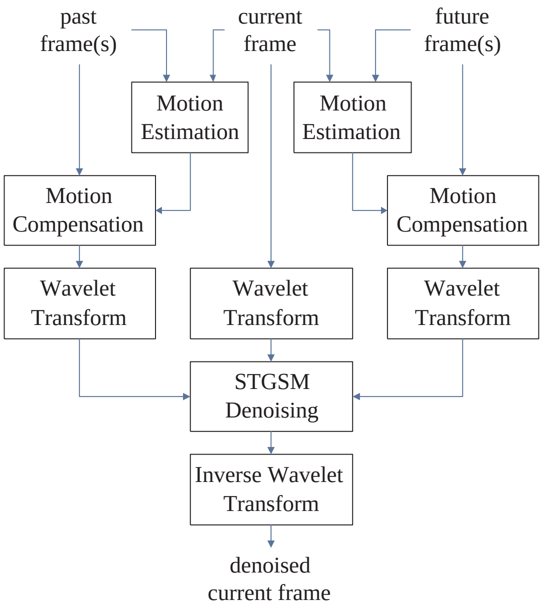


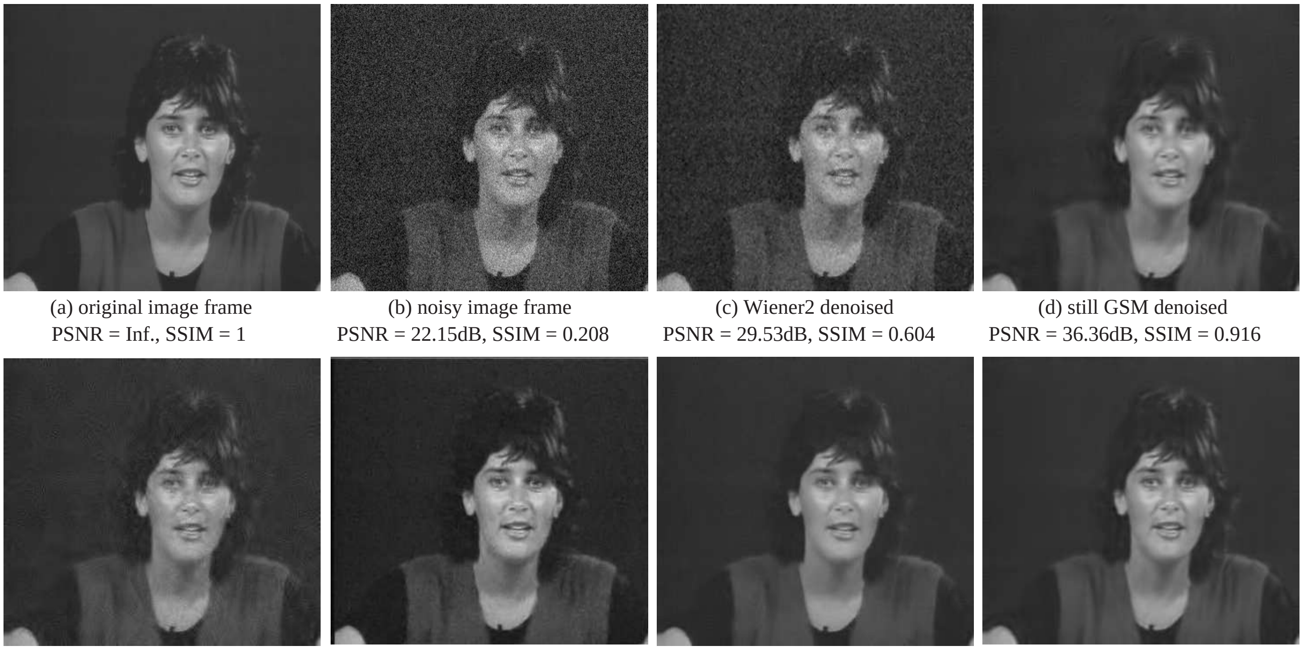
Fig. 1. Diagram of the proposed algorithm.

One of the challenges in the implementation of the above algorithm is to estimate motion in the presence of noise. Here we propose a simple but reliable noise-robust cross correlation method for global motion estimation at integer pixel precision (i.e., a single translation by an integer number of pixels for an entire frame). The use of global and integer-pixel motion estimation avoids interpolation operations in the motion compensation process, which may change signal/noise statistics and severely affect the adequacy of the signal/noise models as well as the denoising estimator.

Let $f_1(x, y)$ and $f_2(x, y)$ represent two image frames, between which a motion vector is to be estimated. The 2D Fourier transform of an image frame is expressed as $F(\omega_1, \omega_2) = \mathscr{F}\{f(x, y)\}$, where $\mathscr{F}$ denotes the Fourier transform operation. In the standard cross correlation based approach, the motion vector can be estimated by finding the peak of the cross correlation function [9], which can be computed efficiently using a Fourier transform approach:

$$h_{CC}(u, v) = \mathscr{F}^{-1}\{Y(\omega_1, \omega_2)\} \tag{5}$$

where $Y(\omega_1, \omega_2) = F_1(\omega_1, \omega_2)F_2^*(\omega_1, \omega_2)$. This method is optimal when the image frames are noise-free. Nevertheless, in our application, only the noisy frames are available. Since natural image signals have significant energy concentration at low frequencies and the noise spectrum is flat, the signal-to-noise ratio at high frequencies is much lower. Therefore, we propose a noise-adapted approach that matches



(a) original image frame
PSNR = Inf., SSIM = 1

(b) noisy image frame
PSNR = 22.15dB, SSIM = 0.208

(c) Wiener2 denoised
PSNR = 29.53dB, SSIM = 0.604

(d) still GSM denoised
PSNR = 36.36dB, SSIM = 0.916

(e) IFSM denoised
PSNR = 33.57dB, SSIM = 0.808

(f) WRSTF denoised
PSNR = 34.78dB, SSIM = 0.849

(g) 3DSWDCT denoised
PSNR = 36.38dB, SSIM = 0.906

(h) Proposed ST-GSM denoised
PSNR = 38.63dB, SSIM = 0.937

Fig. 2. Denoising results of Frame 80 in "Miss America" sequence corrupted with a noise standard deviation $\sigma = 20$.

the non-uniform distribution of signal-to-noise ratios across the Fourier spectrum. In particular, the motion vector is estimated as the peak position (u_0, v_0) in the following noise-robust cross correlation function:

$$h_{NRCC}(u,v) = \mathscr{F}^{-1}\left\{ Y(\omega_1,\omega_2)\left(1 - \frac{|W(\omega_1,\omega_2)|^2}{|Y(\omega_1,\omega_2)|}\right)\right\}, \tag{6}$$

where $W(\omega_1, \omega_2)$ is the noise spectrum (flat for white noise). Note that this function converges to the standard cross correlation function when the images are noise-free.

After motion compensation, each frame is decomposed using the steerable pyramid [10], a redundant wavelet transform that avoids aliasing in subbands. The noisy wavelet coefficient vector $\mathbf{y}$ at a particular position in a subband is then formed from a window of size $N = N_1 \times N_2 \times N_f$, where N_1 and N_2 are the spatial dimensions, and N_f is the number of frames involved. In our implementation, $N_1 = N_2 = 3$ and $N_f = 9$ (i.e, 4 past, 1 current and 4 future frames). Assuming that the noise standard deviation is given, the covariance matrices $\mathbf{C_W}$ and $\mathbf{C_u}$ (both of size N^2) for each subband are estimated from an noise image and the data, respectively. Similar BLS estimation process as in Section 2.1 can then be used to denoise the center coefficient of the current frame. Note that since the noise added to each frame is independent, the cross-terms in $\mathbf{C_W}$ between any pair of coefficients reside in different subbands are zero. By contrast, the same terms in $\mathbf{C_u}$ can be significant because of the strong temporal correlations in the video signal. We regard this as the key reason that distinguishes signals from noise (in a statistical sense) and leads to the major improvement of ST-GSM over intra-frame denoising approaches.

3. RESULT

The proposed ST-GSM denoising algorithm is tested with five standard video sequences ("Tennis", "Garden", "Salesman", "Miss America" and "Foreman") contaminated with additive white Gaussian noise. For easy 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. We have compared the proposed algorithm with state-of-the-art video denoising algorithms, including WRSTF [7], SEQWT [6], 3DWTF [2], IFSM [5], and 3DSWDCT [3]. To better place the performance evaluation in the context, we have also included three baseline algorithms, which are 1) Wiener2D: MATLAB's *Wiener2* function applied on a frame-by-frame basis; 2) Wiener3D: a simple extension of *Wiener2* to 3D, with a window size of $3 \times 3 \times 3$; and 3) still GSM: static image GSM denoising [8] applied on a frame-by-frame basis. The denoising results for WRSTF [7] and 3DWTF [2] were obtained from the processed sequences available at [12], and the rest of the results were computed by ourselves.

Peak signal-to-noise ratio (PSNR) and the structural similarity (SSIM) index [11] are used to provide quantitative evaluations of the algorithms. The latter has shown to be a better indicator of perceived image quality [11]. In Fig. 2, a denoised frame using different denoising algorithms is extracted

Table 1. PSNR and SSIM [11] comparisons of video denoising algorithms.

video sequence	Foreman			Salesman			Miss America			Tennis			Garden		
noise std (σ)	10	15	20	10	15	20	10	15	20	10	15	20	10	15	20
PSNR Results															
Wiener2D	33.14	30.46	28.55	31.97	29.51	27.80	34.51	31.64	29.56	31.07	28.55	26.78	29.73	26.77	24.80
Wiener3D	29.54	29.26	28.87	29.59	29.30	28.88	36.95	35.60	34.06	22.88	22.81	22.71	18.36	18.33	18.29
WRSTF [7]	35.48	33.37	31.82	35.54	33.56	32.00	37.82	36.17	34.79	33.68	31.35	29.71	30.59	27.95	26.13
SEQWT [6]	NA	NA	NA	32.86	30.59	29.02	NA	NA	NA	31.19	29.14	27.59	29.30	26.43	24.38
3DWTF [2]	NA	NA	NA	34.96	33.33	32.03	NA	NA	NA	31.96	29.91	28.56	30.25	27.70	25.95
IFSM [5]	34.13	31.98	30.50	34.22	31.85	30.22	37.52	35.41	33.86	32.41	30.10	28.56	30.05	27.25	25.40
still GSM [8]	35.05	33.10	31.70	33.80	31.73	30.28	38.52	37.14	36.14	31.82	29.87	28.65	30.40	27.65	25.76
3DSWDCT [3]	36.17	34.46	33.07	36.98	35.12	33.75	38.87	37.72	36.74	33.83	31.79	30.50	**31.80**	**29.40**	**27.70**
Proposed	**36.74**	**34.98**	**33.72**	**38.04**	**36.03**	**34.61**	**40.57**	**39.40**	**38.50**	**34.05**	**31.97**	**30.59**	31.48	29.08	27.49
SSIM Results															
Wiener2D	0.860	0.774	0.694	0.859	0.778	0.704	0.818	0.704	0.602	0.813	0.712	0.625	0.916	0.853	0.792
Wiener3D	0.865	0.839	0.808	0.839	0.818	0.786	0.907	0.868	0.808	0.577	0.560	0.539	0.510	0.500	0.488
WRSTF [7]	0.914	0.877	0.841	0.932	0.901	0.868	0.908	0.877	0.846	0.897	0.839	0.790	0.953	0.922	0.889
SEQWT [6]	NA	NA	NA	0.900	0.846	0.796	NA	NA	NA	0.842	0.772	0.716	0.941	0.893	0.842
3DWTF [2]	NA	NA	NA	0.923	0.903	0.882	NA	NA	NA	0.856	0.793	0.740	0.909	0.872	0.840
IFSM [5]	0.886	0.836	0.793	0.904	0.851	0.801	0.904	0.857	0.812	0.855	0.776	0.709	0.927	0.882	0.837
still GSM [8]	0.916	0.889	0.867	0.909	0.865	0.825	0.936	0.922	0.913	0.831	0.758	0.711	0.939	0.899	0.857
3DSWDCT [3]	0.932	0.907	0.884	0.955	0.930	0.905	0.946	0.928	0.909	**0.894**	0.834	0.790	**0.959**	**0.931**	**0.900**
Proposed	**0.937**	**0.917**	**0.901**	**0.960**	**0.941**	**0.923**	**0.952**	**0.943**	**0.936**	**0.894**	**0.841**	**0.797**	0.950	0.925	**0.900**

from the "Miss America" sequence. It can be observed that the proposed ST-GSM algorithm is quite effective at removing the noise while maintaining the edge and texture details of the image. More comparisons are shown in Table 1, where the proposed method gives the best performance in most cases.

4. CONCLUSION

We propose an ST-GSM model for natural video signals and use it for video denoising. We find that applying motion compensation before ST-GSM denoising is effective in further improving its performance. A Fourier domain noise-robust method is proposed to provide reliable motion estimation in the presence of noise. Experiments with a set of standard video sequences contaminated with additive white Gaussian noise show that the proposed denoising algorithm is highly competitive in terms of both PSNR and SSIM evaluations when compared to state-of-the-art methods.

5. REFERENCES

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

[2] W. I. Selesnick and K. Y. Li, "Video denoising using 2D and 3D dualtree complex wavelet transforms," in *Proc. SPIE, Wavelets:Applications in Signal and Image Processing X*, vol. 5207, (San Diego), pp. 607–618, Nov. 2003.

[3] D. Rusanovskyy and K. Egiazarian, "Video denoising algorithm in sliding 3D DCT domain," in *Advanced Concepts for Intelligent Vision Systems, ACIVS*, (Antwerp, Belgium), Sept. 2005.

[4] N. Lian, V. Zagorodnov, and Y. Tan, "Video denoising using vector estimation of wavelet coefficients," in *Proc. IEEE Int. Sym. Circuits and Systems*, pp. 2673–2676, May 2006.

[5] S. M. M. Rahman, M. O. Ahmad, and M. N. S. Swamy, "Video denoising based on inter-frame statistical modeling of wavelet coefficients," *IEEE Trans. Circuits and Systems for Video Tech.*, vol. 17, pp. 187–198, Feb. 2007.

[6] A. Pizurica, V. Zlokolica, and W. Philips, "Combined wavelet domain and temporal video denoising," in *IEEE Conf. on Advanced Video and Signal Based Surveillance*, pp. 334–341, July 2003.

[7] V. Zlokolica, A. Pizurica, and W. Philips, "Wavelet-domain video denoising based on reliability measures," *IEEE Trans. Circuits and Systems for Video Tech.*, vol. 16, pp. 993–1007, Aug. 2006.

[8] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using Gaussian scale mixtures in the wavelet domain," *IEEE Trans. Image Processing*, vol. 12, pp. 1338–1351, Nov. 2003.

[9] R. Manduchi and G. A. Mian, "Accuracy analysis for correlation-based image registration algorithms," in *Proc. IEEE Int. Sym. Circuits and Systems*, pp. 834–837, 1993.

[10] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable multi-scale transforms," *IEEE Trans. Information Theory*, vol. 38, pp. 587–607, 1992.

[11] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Processing*, vol. 13, pp. 600–612, Apr. 2004.

[12] V. Zlokolica, "Results of 'Wavelet-domain video denoising based on reliability measures'," `http://telin.ugent.be/~vzlokoli/Results_J`.