A Kalman Filter based Video Denoising Method Using Intensity and Structure Tensor

Yu Liu, Chenlin Zuo, Xin Tan, Huaxin Xiao, Maojun Zhang
College of Information System and Management, National University of Defense Technology,
Changsha, PR China
[e-mail: jasonyuliu@nudt.edu.cn]
*Corresponding author: Yu Liu

Abstract

We propose a video denoising method based on Kalman filter to reduce the noise in video sequences. Firstly, with the strong spatiotemporal correlations of neighboring frames, motion estimation is performed on video frames consisting of previous denoised frames and current noisy frame based on intensity and structure tensor. The current noisy frame is processed in temporal domain by using motion estimation result as the parameter in the Kalman filter, while it is also processed in spatial domain using the Wiener filter. Finally, by weighting the denoised frames from the Kalman and the Wiener filtering, a satisfactory result can be obtained. Experimental results show that the performance of our proposed method is competitive when compared with state-of-the-art video denoising algorithms based on both peak signal-to-noise-ratio and structural similarity evaluations.

---

1. Introduction

Digital video surveillance is prevalent in our daily life. Large numbers of monitoring cameras are installed in public and private places, such as government buildings, military bases, and car parks. To obtain high quality surveillance, video denoising techniques have been well studied in the field of image processing. Apart from denoising itself, these techniques can be used to increase compression efficiency, reduce transmission bandwidth, and improve the effectiveness of further processes, such as feature extraction, object detection, and pattern classification.

Even though video and image denoising can be considered different research topics, some basic image denoising ideas and algorithms are borrowed for video denoising, such as Gaussian filter, bilateral filter [1-2], domain transformation [3-5], similar blocks matching [4-6, 28-29], sparse representations [30-32] etc. Compared to a single image, video can provide sufficient additional information from nearby frames, which can bring better denoising results. Moreover, with the emergence of new multi-resolution tools, such as the wavelet transform [7-8], video denoising methods performed in the transform domain have been proposed continually [9-13]. Zlokolica et al. [9] introduced new wavelet-based motion reliability measures, and performed motion estimation and adaptive recursive temporal filtering in a closed loop, which is followed by an intra-frame spatially adaptive filter. Rahman et al. [10] proposed a joint probability density function to model the video wavelet coefficients of any two neighboring frames, and then applied this statistical model for denoising. Jovanov et al. [11] reused motion estimation resources from a video-coding module for video denoising. They proposed a novel motion field-filtering step and a novel recursive temporal filter with the reliability of the estimated motion field appropriately defined. Yu et al. [12] integrated both spatial filtering and recursive temporal filtering into the 3-D wavelet domain and effectively exploited spatial and temporal redundancies. Maggioni et al. [13] exploited the temporal and nonlocal correlation of the video and constructed 3-D spatiotemporal volumes separately by tracking blocks along trajectories defined by motion vectors. Jin et al. [33] proposed a multi-resolution motion analysis method in the wavelet domain. In [34], the change was estimated in the 3D SCT domain. Lian et al. [35] used vector estimation of wavelet coefficients. In addition, other proposed video denoising methods, such as one that uses low-rank matrix completion [14], achieved relatively better results.

Video denoising technology has made great progress over the previous decades. However, most existing methods cannot obtain ideal results when dealing with large noisy video sequences captured under low light environment. This requirement is urgently demanded in many fields, especially for security monitoring, where a camera is mounted at a stable position with a fixed angle in which the captured video sequences have relatively unchanged backgrounds. In practical applications, the characteristics of both still and moving objects must be clearly seen in the video sequences. This requirement can easily be satisfied during the day. However, at night, statistical noise due to low light illumination seriously affects the video sequences.

In this paper, a novel video denoising method based on Kalman filter is proposed. Taking advantage of the strong spatiotemporal correlations of neighboring frames, motion estimation based on intensity and structure tensor [15-17] is performed by comparing current noisy frame with previous denoised frames. Then, based on motion estimation results, current noisy frame is processed in temporal domain using the Kalman filter [18]. During the filtering process,

different positions of the noisy frame have different filtering strengths according to the motion estimation results. Motion positions have weak filtering strength and keeping their motion characteristic is difficult, whereas still positions have strong filtering strength for reducing noise. Simultaneously, the noisy frame is also processed in the spatial domain using the Wiener filter [19]. Finally, by weighting the two denoised frames using Kalman and Wiener filtering methods, a satisfactory result can be obtained. The still region is obtained largely from Kalman filtering, while the motion region is the result of Wiener filtering. Experimental results show that the performance of our proposed method is effective over current competing video denoising methods.

The remainder of the paper is organized as follows. Section 2 describes our proposed video denoising method. Section 3 provides quantitative quality evaluations of the denoising results. Section 4 discusses the experiments as well as the results. Finally, Section 5 concludes this article.

2. Proposed Denoising Method

Fig. 1 illustrates the diagram of our proposed video denoising method. The denoising of current noisy frame involves not only the frame itself, but also a series of previously denoised frames. Motion estimation is performed based on intensity and structure tensor between the current noisy frame and the previous denoised frames. Then, the estimation results guide the Kalman filtering on the current noisy frame. In this operation, the final denoised frame from Kalman filtering is needed. Simultaneously, Wiener spatial filtering is also performed on the current noisy frame. Thus, after processing, two denoised frames are obtained. One is obtained using Kalman filtering, and another is obtained using Wiener filtering. Finally, by weighting the two denoised frames, a satisfactory result can be obtained.
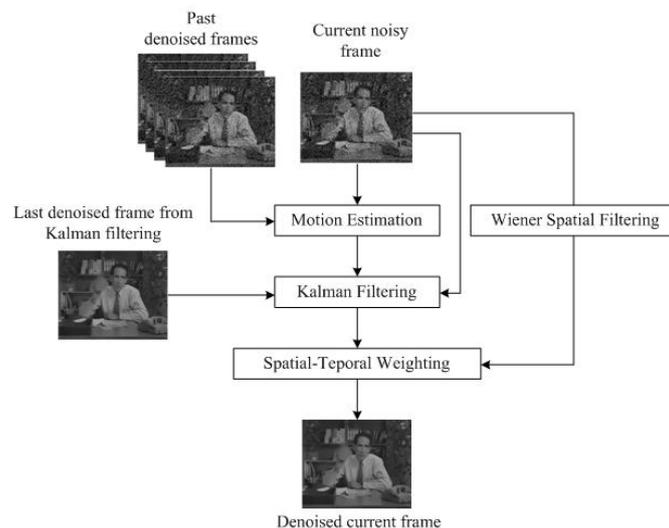

Fig. 1. Diagram of proposed video denoising method

2.1 Motion Estimation based on Intensity and Structure Tensor

To take advantage of the strong correlations between adjacent frames, intensity and structure tensor based motion estimation is performed by comparing the current noisy frame with previous denoised frames.

2.1.1 Intensity based Motion Estimation

In order to suppress the noise influence, a strong filter is firstly used to pre-process the noise images. Prefilter is frequently used in many denoising algorithms, such as VBM3D [4]. Considering the algorithm complexity and the noise suppressing ability, we employ the Gaussian filter with large kernel size. Then, the intensity distance could be calculated as follows.

$$d_I(k,i) = \begin{cases} |K_{\rho_1}(p_k) - K_{\rho_1}(p_i)|, k \neq i \\ 0, k = i \end{cases} \tag{1}$$

In above equation, k is the temporal index of the frame. In particular, i is the current frame's index, namely, $k = ..., i-2, i-1, i, i+1, i+2, ...$. p_k is the pixel value in some position of the frame. In particular, p_i is the pixel value of the current frame. K_{ρ_1} is the Gaussian filter kernel with the standard variance ρ_1. $d_I(k,i)$ is the intensity distance between frame k and frame i.

Fig. 2(a1) and **(a2)** are the past and current frames with additive Gaussian white noise, whose $\sigma = 50$. Before calculating the intensity distance, the two frames are prefiltered with a 10×10 Gaussian filter whose $\rho_1 = 5$, and the results are shown in **Fig. 2(b1)** and **(b2)**. The choice of the filter kernel follows the noise level. The larger the noise is, the larger the kernel size is. Then, the intensity distance is calculated based on this two prefiltered frames and the result is shown in **Fig. 2(b3)**.
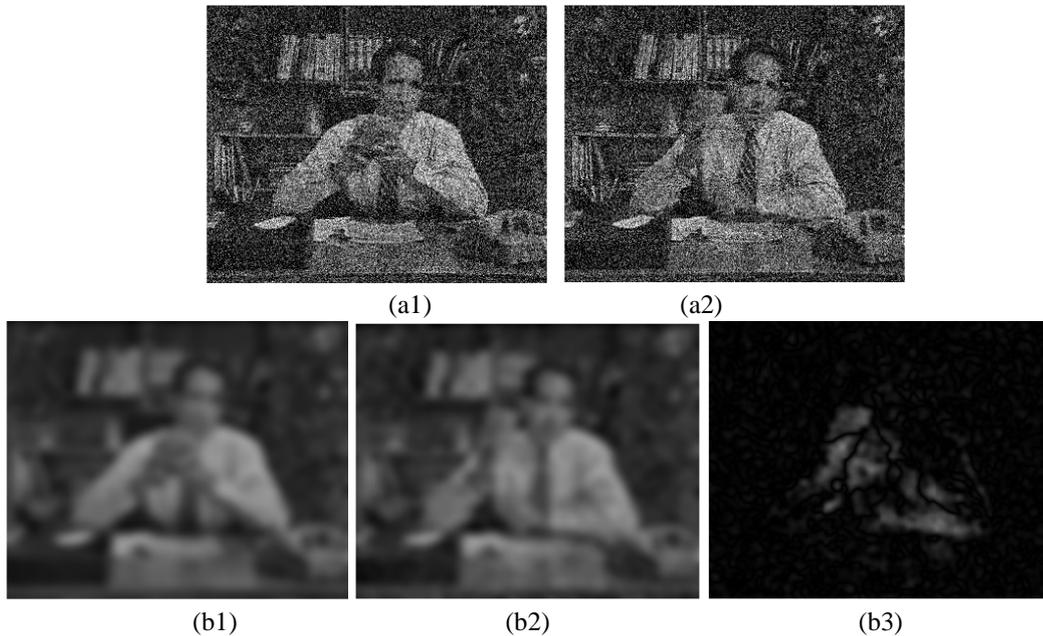

(a1)   (a2)

(b1)   (b2)   (b3)

Fig. 2. Intensity based motion estimation. (a1) and (a2) are the past and current frame with additive Gaussian white noise ($\sigma = 50$). (b1) and (b2) are the prefiltered results of (a1) and (a2) with a 10×10 Gaussian filter whose $\rho_1 = 5$. (b3) is the intensity distance of (b1) and (b2).

2.1.2 Structure tensor based Motion Estimation

Although the strong prefilter effectively suppresses the large scale noise, it destroys the edges of the motion area too. Some detail variations are also damaged and even lost. Weickert et al. [15-17] first proposed the structure tensor, which is used as a tool for analyzing image structure, extracting the geometric feature, etc. In this paper, the simple linear structure tensor is used to analyze the image. This simple linear structure tensor is defined as

$$\mathbf{J}_{\rho_2}(p) = K_{\rho 2} * (\nabla p_{\sigma'} \otimes \nabla p_{\sigma'}^T) = K_{\rho 2} * \begin{pmatrix} (I_x(p_{\sigma'}))^2 & I_x(p_{\sigma'})I_y(p_{\sigma'}) \\ I_x(p_{\sigma'})I_y(p_{\sigma'}) & (I_y(p_{\sigma'}))^2 \end{pmatrix} \quad (2)$$

In the above equation, ∇ is the image gradient operator, and $p_{\sigma'}$ is the Gaussian filtered image of input p with the Gaussian standard variance σ'. In addition, $\otimes$ is the structure tensor product. The image gradients $I_x(p_{\sigma'})$ and $I_y(p_{\sigma'})$ can be used in x and y directions. Moreover, $*$ is the convolution of Gaussian filter $K_{\rho 2}$ with standard variance ρ_2 and the structure tensor product. Generally, $\rho_2 > \sigma'$. The Gaussian filter σ' before gradient operation and the filter $K_{\rho 2}$ play the role of the strong pre-filter. The Gaussian filter $K_{\rho 2}$ isotropically synthesizes the local neighborhood structure tensor information, and is thus, called "linear structure tensor."

$\mathbf{J}_{\rho_2}$ contains the image geometric structure information. By orthogonally decomposing $\mathbf{J}_{\rho_2}$, we obtain eigenvalues, λ_1 and λ_2, and eigenvectors, $\vec{e}_1$ and $\vec{e}_2$. The eigenvalues describe the strength of the direction of the eigenvectors, which reflect the direction of the image structures. The corresponding eigenvector $\vec{e}_1$ of the maximum eigenvalue λ_1 indicates the direction of the maximum gradient contrast, i.e., the normal direction. The corresponding eigenvector $\vec{e}_2$ of eigenvalue λ_2 indicates the tangential direction.

Different image structures can be described using different eigenvalues. Usually, $\lambda_1 + \lambda_2$ is used to reflect the strength of the structure. **Fig. 3(1)** and **(2)** show the maps of the structure strength extracted from the noise frames in **Fig. 2(a1)** and **(a2)**, respectively.

When motion occurs, variation in the structure tensor is unavoidable. The structure tensor could be used to detect the motion. Thus, the structure tensor distance should be measured. Given that the structure tensor resides in non-Euclidean space, we use a Riemannian metric called Log-Euclidean metric [20] with simple and fast computations. The metric is computed as

$$d_{ST}(p_{current}, p_{past,i}) = \sqrt{Trace\left(\left(\log\left(\mathbf{J}_{\rho_2}\left(p_{current}\right)\right) - \log\left(\mathbf{J}_{\rho_2}\left(p_{past,i}\right)\right)\right)^2\right)}. \quad (3)$$

In the above equation, $Trace(\cdot)$ is the trace of the matrix, and $\log(\cdot)$ is the structure tensor logarithmic operator defined in [20]. In addition, $\mathbf{J}_{\rho_2}\left(p_{current}\right)$ represents the structure tensor of the current noisy frame, and $\mathbf{J}_{\rho_2}\left(p_{past,i}\right)$ represents the structure tensor of the *i-th* previous denoised frame. **Fig. 3(3)** shows the Log-Euclidean metric distance of **Figs. 3(1)** and **(2)**.

(1)   (2)   (3)
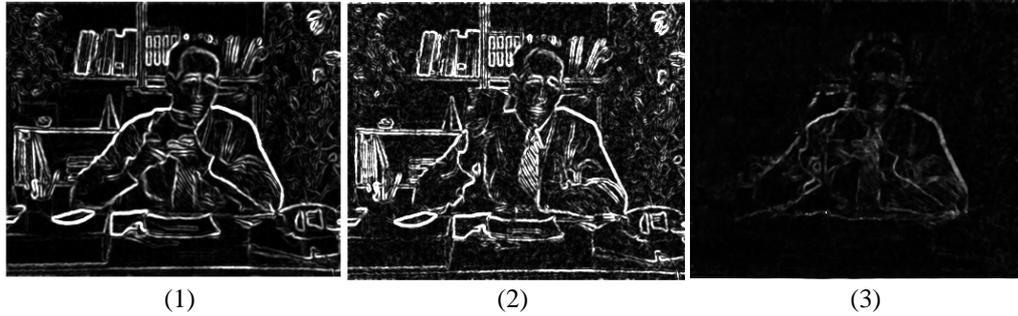
Fig. 3. Structure tensor based motion estimation. (1) and (2) are the maps of the structure strength $\lambda_1 + \lambda_2$ extracted from the noise frames in **Fig. 2(a1)** and **(a2)**. (3) is the Log-Euclidean metric distance of (1) and (2).

Structure tensor based motion estimation is a good supplement for intensity based motion estimation. The intensity and structure tensor combined motion estimation is shown in **Fig. 4**. The combination follows:

$$d_{IST}(k,i) = \alpha \cdot d_{ST}(k,i) + \beta \cdot d_I(k,i) \tag{4}$$

where α and β are weighted parameters. In **Fig. 4**, $\alpha = 0.1$ and $\beta = 1$.

Fig. 4. Intensity and structure tensor combined change segmentation

2.2 Motion Estimation based Kalman Filtering in Temporal Domain

The discrete Kalman filter [18] can provide an efficient solution to the least squares method. Generally, the step is made up of two consecutive stages, namely, prediction and updating.

The prediction equations are defined as

$$x_k^- = A_k x_{k-1}^+ + B_k u_k \tag{5}$$

and

$$p_k^- = A_k p_{k-1}^+ A_k^T + Q_k \tag{6}$$

where the superscripts "-" and "+" denote "before" and "after" each measurement, respectively. Moreover, x_{k-1}^+ represents the estimated state matrix and p_{k-1}^+ represents the state covariance matrix of last state; x_k^- and p_k^- represent the a priori estimates of state matrix and state covariance matrix for the current state, respectively; and A_k represents the state transition matrix that determines the relationship between the present state and the previous

one. Matrix B_k relates the control input u_k to the current state, and Q_{k-1} represents the covariance matrix of process noise.

In our proposed method, we attempt to estimate the current frame based on the last one. Thus, the state matrix in the equations can be expressed by using the frame matrix. Otherwise, no control input is available, hence, $u_k = 0$. The priori estimates for current state is assumed to be the same as that of the previous state, so the initial A_k is an identity matrix. Then, the following equations can be obtained.

$$x_k^- = x_{k-1}^+ \tag{7}$$

$$p_k^- = p_{k-1}^+ + Q_k \tag{8}$$

The motion in the video sequences brings the process noise. Thus, for any pixel (x, y) of the current noisy frame,

$$Q_{k-1}(x, y) = d_{IST}(x, y), \tag{9}$$

which keeps the covariance of motion region larger than that of the still region.

The updating equations are defined as

$$Kg_k = p_k^- H_k^T (H_k p_k^- H_k^T + R_k)^{-1} \tag{10}$$

$$x_k^+ = x_k^- + Kg_k(z_k - H_k x_k^-) \tag{11}$$

$$p_k^+ = (I - Kg_k H_k) p_k^- \tag{12}$$

where Kg_k is known as the blending factor for minimizing the posteriori error covariance, called the Kalman gain. Variables x_k^- and p_k^- are the priori estimates calculated in the prediction stage. Matrix H_k describes the relationship between the measurement vector, z_k, and the posteriori state vector, x_k^+. R_k is the covariance matrix of measurement noise, and p_k^+ is the posteriori estimate of state covariance matrix for the current state.

In our proposed method, the current noisy and denoised frames are described as z_k and x_k^+. H_k is the unit matrix. The measurement noise just represents the noise in the video sequences. Thus, the following equations can be obtained.

$$Kg_k = p_k^- (p_k^- + R_k)^{-1} \tag{13}$$

$$x_k^+ = x_k^- + Kg_k(z_k - x_k^-) \tag{14}$$

$$p_k^+ = (I - Kg_k) p_k^- \tag{15}$$

After Kalman filtering, a denoised frame can be obtained. In this frame, the still region is denoised well. However, the moving region still has much noise because the Kalman filter keeps the information of this region intact. Therefore, the noise in the moving region must still be reduced. Reducing the noise in the moving region of denoised frame from Kalman filtering is complicated. Thus, the Wiener filter [19] is applied on the entire current noisy frame. In this case, both the still and moving regions are denoised. Then, by weighting the two denoised frames using Kalman and Wiener filtering, an integrated denoised frame can be obtained. In the denoised frame, the still region is obtained by using Kalman filtering, and the moving region is obtained by using Wiener filtering.

2.3 Spatial-Temporal Weighting

After Kalman and Wiener filtering, two denoised frames are obtained. The image from Kalman filtering showed the still regions are well denoised, but the motion regions retained the noisy information. The result of the Wiener filtering indicated that the motion regions were denoised to some extent. Thus, we integrated the two denoised frames by weighting them based on motion estimation results. The weight is based on Gaussian distribution, and, for any pixel，whose position is (x, y), its weight value, $w_c(x, y)$, can be calculated as follows.

$$w_c(x, y) = e^{-\frac{(d_{IST,x,y})^2}{\sigma_c^2}} \tag{16}$$

In the above equation, $d_{IST,x,y}$ is the corresponding motion estimation value in the position (x, y), and σ_c is used to control the degree of attenuation. The larger the value of motion estimation is, the smaller the weight will be. Thus, the motion and still regions can be further distinguished effectively.

The weighted denoised frame can be calculated as follows.

$$X_c = W_c X_{Kalman} + [I - W_c] X_{Wiener} \tag{17}$$

Here, W_c represents the weight matrix calculated using Equation (16). X_{Kalman} and X_{Wiener} represent the denoised frame matrices through Kalman filtering and Wiener filtering, respectively. X_c is simply the desired weighted frame matrix. After obtaining the weighted average, both the motion and still regions of the weighted frame have been denoised.

2.4 Complexity Analysis

We assume that the size of each frame (total pixel number) is N. The proposed method includes three steps: motion estimation, Kalman filtering and Wiener filtering. Firstly, in motion estimation, intensity based and structure tensor based motion estimation are implemented, respectively. In intensity based motion estimation, the size of Gaussian convolution kernel is assumed to be $r \times r$. If we divide the convolution to the vertical and horizontal one, the time complexity will be $O(Nr)$. However, in our method, the size of Gaussian convolution kernel is usually invariable, such as 5×5, 10×10 or 15×15, and it will not increase along with the increase of frames' size. So, the time complexity of Gaussian filtering will be $O(N)$. After that, calculating the intensity distance is implemented, in which the time complexity is $O(N)$. So, the total time complexity of intensity based motion estimation still is $O(N)$. Then, in structure tensor based motion estimation, because the size of Gaussian convolution kernel and gradient convolution kernel are also not increase along with the increase of frames' size, the time complexity of Gaussian filtering and gradient operator are $O(N)$, respectively. Then, the time complexity of calculating the structure tensor distance is $O(N)$. So, the total time complexity of structure tensor based motion estimation still is $O(N)$. Therefore, the total time complexity of the motion estimation is $O(N)$. After motion estimation, Kalman filtering and Wiener filtering are implemented respectively, in which the time complexity are both $O(N)$. Finally, the time complexity of the proposed method is $O(N)$, which is linear.

3.   Denoising Validation Criteria

To provide quantitative quality evaluations of the denoising results, we employed two objective criteria, namely, PSNR and SSIM [21-23]. PSNR is defined as

$$PSNR = 10 \cdot \log_{10}(\frac{L^2}{MSE}),$$

(18)

where L is the dynamic range of the image (for 8 bits/pixel images, $L = 255$). MSE is the mean squared error between the original and distorted images. SSIM is first calculated within local windows using

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)},$$

(19)

where x and y are the image patches extracted from the local window from the original and noisy images, respectively. μ_x, σ_x^2, and σ_{xy} are the mean, variance, and cross-correlation computed within the local window, respectively. The overall SSIM score of a video frame is computed as the average local SSIM scores. PSNR is the mostly widely used quality measure in existing literature, but has been criticized for not correlating well with human visual perception [24]. SSIM is believed to be a better indicator for perceived image quality [24] as it also supplies a quality map that indicates the variations of images quality over space. The final PSNR and SSIM results for a denoised video sequence are computed as the frame average of the full sequence.

4.   Experiments and Results

To evaluate the performance of the proposed method, we compared some state-of-the-art video denoising algorithms, such as ST-GSM [3] and VBM3D [4]. The original codes of these two algorithms can be downloaded online [25-26]. Besides, we also gave the experimental results of using Kalman filter and Wiener filter separately.

The standard test videos can be downloaded at video sequence base [27]. Two types of videos are available in the base, namely, stationary and moving backgrounds. Given that our method is for videos with a stationary background, we chose four former types of videos in our experiment, which are Salesman, Paris, Akiyo, and Hall. The size of the video is 288×352, and the duration is 300 frames. The experiment was conducted on the luminance channel of the video. The noisy video sequences are simulated by adding independent white Gaussian noises at a given variance σ² on each frame.

Table 1 shows the PSNR and SSIM results of ST-GSM, VBM3D, Kalman-only, Wiener-only, and our proposed method for the four video sequences at five noise levels. As seen from the table, both Kalman-only and Wiener-only methods could not obtain good denoising results. When the noise level was relatively low, the proposed method worked well, but a gap still existed in ST-GSM and VBM3D. However, when the noise level was high, the proposed method performed better than ST-GSM and VBM3D for most test sequences. In particular, the SSIM of our proposed method was better than the other two algorithms.

Table 1. PSNR and SSIM Comparisons of Video Denoising Algorithms for Four Video Sequences at Five Noise Levels

Video sequence	Salesman					Paris				
Noise std (σ)	10	15	20	50	100	10	15	20	50	100
PSNR Results (dB)										
ST-GSM [3]	37.93	35.56	33.89	26.43	20.72	36.42	34.17	32.59	26.15	18.85
VBM3D [4]	**39.11**	**36.65**	**34.72**	27.93	22.18	**38.15**	**35.86**	**34.14**	**27.34**	20.57
Kalman-only	33.71	32.82	32.19	26.16	21.91	27.54	27.05	26.85	23.77	20.89
Wiener-only	31.90	29.83	27.95	24.23	20.67	25.76	24.90	22.38	19.18	16.83
Proposed method	35.33	33.62	33.27	**29.28**	**22.48**	30.57	28.04	28.01	25.06	**21.44**
SSIM Results										
ST-GSM [3]	0.970	0.950	0.928	0.699	0.452	0.967	0.951	0.936	0.840	0.510
VBM3D [4]	**0.976**	**0.958**	**0.932**	0.742	0.489	**0.977**	**0.964**	**0.949**	**0.847**	0.554
Kalman-only	0.920	0.902	0.899	0.641	0.596	0.901	0.882	0.862	0.751	0.627
Wiener-only	0.874	0.811	0.751	0.563	0.417	0.851	0.812	0.713	0.519	0.351
Proposed method	0.936	0.921	0.914	**0.857**	**0.738**	0.943	0.913	0.909	0.838	**0.731**
Video sequence	Akiyo					Hall				
Noise std (σ)	10	15	20	50	100	10	15	20	50	100
PSNR Results (dB)										
ST-GSM [3]	40.67	38.34	36.53	28.44	21.89	38.28	35.99	34.12	27.16	19.99
VBM3D [4]	**42.00**	**39.72**	**37.85**	**30.69**	23.36	**39.96**	**37.93**	**36.31**	28.14	21.97
Kalman-only	33.60	32.26	30.85	28.48	22.76	32.69	31.99	31.34	26.83	22.52
Wiener-only	33.71	31.11	29.40	24.60	20.85	30.68	29.03	26.36	21.77	18.62
Proposed method	34.67	33.82	32.26	30.43	**23.49**	36.26	34.52	32.14	**28.30**	**23.05**
SSIM Results										
ST-GSM [3]	0.980	0.969	0.958	0.852	0.673	0.975	0.965	0.955	0.882	0.620
VBM3D [4]	**0.984**	**0.976**	**0.964**	**0.871**	0.614	**0.980**	**0.973**	**0.966**	0.887	0.601
Kalman-only	0.948	0.931	0.907	0.790	0.576	0.954	0.941	0.914	0.795	0.612
Wiener-only	0.901	0.825	0.828	0.633	0.477	0.893	0.831	0.806	0.622	0.453
Proposed method	0.958	0.947	0.931	0.865	**0.741**	0.971	0.967	0.956	**0.900**	**0.778**

Fig. 5 demonstrates the visual effects of above five video denoising algorithms. Specifically, Frame 100 was extracted from the Akiyo sequence together with a noisy version of the same frame. The denoised frames were obtained by using the five video denoising algorithms. The Kalman-only and our proposed method are obviously effective at suppressing background noise, but Kalman-only method is failed to remove the noise of motion region, such as the woman's head in the frame, while our method could suppress the noise of motion region to some extent. This finding is further verified by examining the SSIM quality maps of the corresponding frames. The results show that our proposed method is effective for the large noisy video sequences and can achieve state-of-the-art denoising performance.
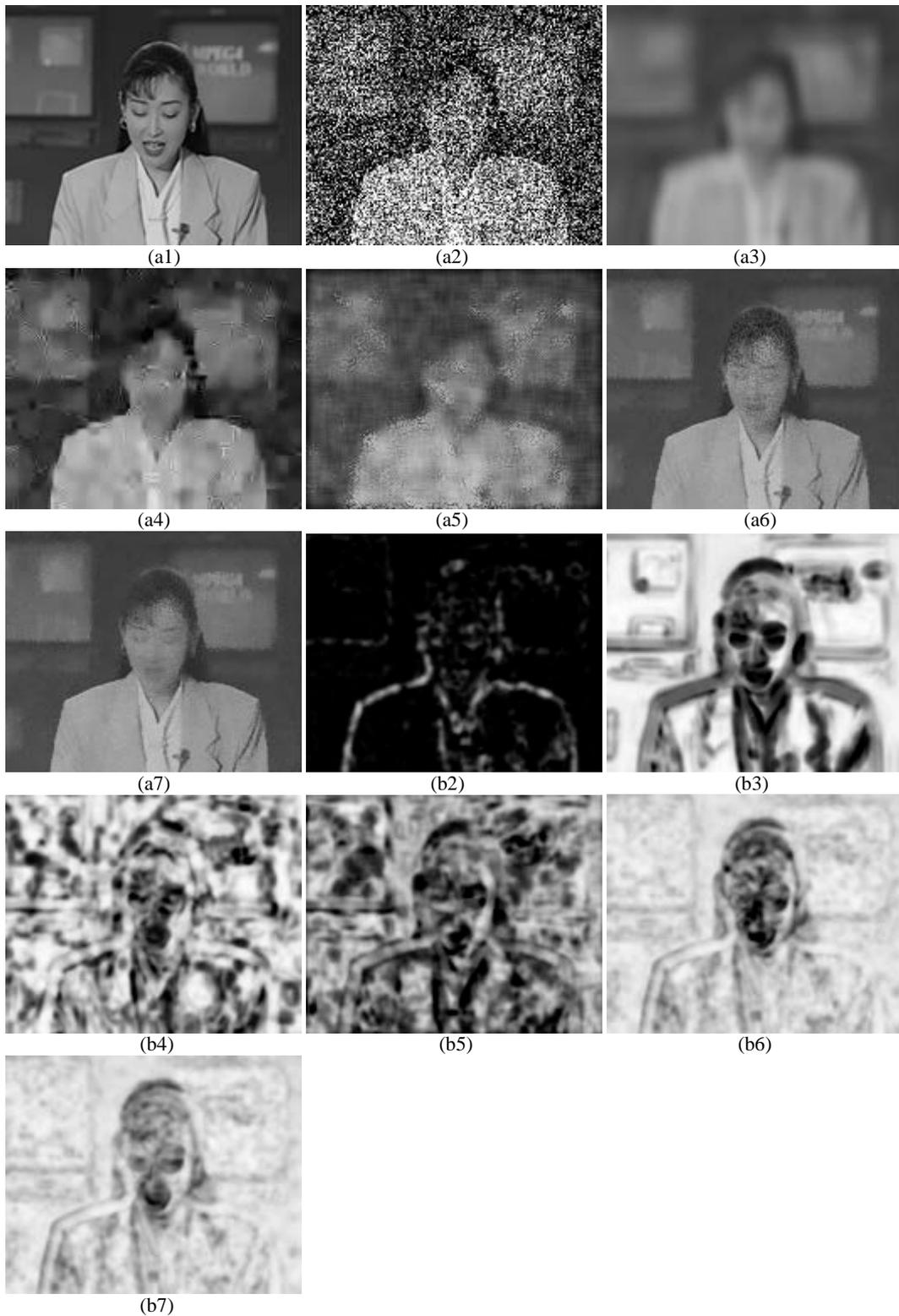
Fig. 5. Denoising results of frame 100 in the Akiyo sequence corrupted with noise with a standard deviation σ = 100. (a1) to (a7): Frames in the original, noisy, ST-GSM [3], VBM3D [4], Kalman-only, Wiener-only, and our proposed method denoised sequences. (b2) to (b7): Corresponding SSIM quality maps (brighter areas indicate larger SSIM values).

5. Conclusion

This paper presented a video denoising method based on Kalman filter for large noisy video signals. This method was applied to the restoration of noisy video sequences with added white Gaussian noise. Motion estimation was performed by employing intensity and structure tensor comparing the current noisy frame with previous denoised frames. Then, the Kalman and the Wiener filters were applied on the current noisy frame. Finally, by weighting the denoised frames from the filtering methods, a satisfactory result was obtained. The experimental comparisons with state-of-the-art algorithms show that the proposed method achieved competitive results for large noisy video sequences with a fixed background in terms of both subjective and objective evaluations.

References

[1] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. of IEEE Int. Conf. Computer Vision*, pp. 839-846, Bombay, India, 1998. Article (CrossRef Link)

[2] E. P. Bennett and L. McMillan, "Video enhancement using per-pixel virtual exposures," in *Proc. of ACM SIGGraph 05 Conference*, pp. 845–852, Jul. 2005. Article (CrossRef Link)

[3] G. Varghese and Z. Wang, "Video denoising based on a spatiotemporal Gaussian scale mixture model", *IEEE Trans. Circuits and Systems for Video Technology*, vol. 20, no. 7, pp. 1032–1040, Jul. 2010. Article (CrossRef Link)

[4] K. Dabov, A. Foi, and K. Egiazarian, "Video denoising by sparse 3-D transform-domain collaborative filtering," *IEEE trans. Eur. Signal Process. Conf.*, Poznan, Poland, pp. 1257-1260, Sep. 2007. Article (CrossRef Link)

[5] F. Luiser, T. Blu, and M. Unser, "SURE-LET for Orthonormal Wavelet-Domain Video Denoising," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 20, no. 6, pp. 913–919, Jun. 2010. Article (CrossRef Link)

[6] Y. Han, and R.Chen, "Efficient video denoising based on dynamic nonlocal means," *Image and Vision Computing*, vol. 30, no. 2, pp. 78-85, Feb. 2012. Article (CrossRef Link)

[7] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989. Article (CrossRef Link)

[8] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Comm. Pure Appl. Math.*, vol. 41, no. 7, pp. 909–996, Oct. 1988. Article (CrossRef Link)

[9] V. Zlokolica, A. Pizurica and W. Philips, "Wavelet-domain video denoising based on reliability measures," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 8, pp. 993–1007, Aug. 2006. Article (CrossRef Link)

[10] S. M. M. Rahman, F. M. Omair Ahmad, and M. N. S. Swamy, "Video denoising based on inter-frame statistical modeling of wavelet coefficients," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 2, pp. 187–198, Feb. 2007. Article (CrossRef Link)

[11] L. Jovanov, A. Pizurica, S. Schulte, P. Schelkens, A. Munteanu, E. Kerre, and W. Philips, "Combined wavelet-domain and motion-compensated video denoising based on video codec motion estimation methods," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 3, pp. 417–421, Mar. 2009. Article (CrossRef Link)

[12] S. Yu, M. O. Ahmad and M. N. S. Swamy, "Video denoising using motion compensated

3-D wavelet transform with integrated recursive temporal filtering," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 6, pp. 780–791, Jun. 2010. Article (CrossRef Link)

[13] M. Maggioni, G. Boracchi. A. Foi and K. Egiazarian, "Video denoising, deblocking, and enhancement through separable 4-D nonlocal spatiotemporal transforms," *IEEE Transactions on Image Processing*, vol. 21, no. 9, pp. 3952-3966, Sep. 2012. Article (CrossRef Link)

[14] H. Ji, C. Liu, Z. Shen and Y. Xu, "Robust video denoising using Low rank matrix completion," *in Proc. of CVPR*, pp. 13-18, Jun. 2010. Article (CrossRef Link)

[15] J.Weickert, H. Scharr, "A scheme for coherence-enhancing diffusion filtering with optimized rotation invariance," *J. Visual Comm. Imag. Repres*, vol. 13, pp.103-118, 2002. Article (CrossRef Link)

[16] J.Weickert, "Anisotropic Diffusion in Image Processing," *Teubner-Verlag*, Stuttgart, Germany, 1998.

[17] J. Weickert, "Coherence-enhancing diffusion filltering," *Int. J. Computer Vision*, vol. 31, pp. 111-127, Apl. 1999. Article (CrossRef Link)

[18] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME, Journal of Basic Engineering*, vol. 82, pp. 35-45, 1960. Article (CrossRef Link)

[19] J. S. Lim, "Two-Dimensional Signal and Image Processing," *Englewood Cliffs*, NJ, Prentice Hall, pp. 548, 1990. Article (CrossRef Link)

[20] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache, "Log-Euclidean metrics for fast and simple calculus on diffusion tensors," Magnetic Resonance in Medicine, vol. 56, no. 2, pp. 411–421, Jun. 2006. Article (CrossRef Link)

[21] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002. Article (CrossRef Link)

[22] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004. Article (CrossRef Link)

[23] Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment based on structural distortion measurement," *Signal Process.: Image Commun.*, vol. 19, no.2, pp. 121–132, Feb. 2004. Article (CrossRef Link)

[24] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at signal fidelity measures," *IEEE Signal Process.* Mag., vol. 26, no. 1, pp. 98–117, Jan. 2009. Article (CrossRef Link)

[25] Original codes of ST-GSM: https://ece.uwaterloo.ca/~z70wang/research/stgsm/

[26] Original codes of VBM3D: http://www.cs.tut.fi/~foi/GCF-BM3D/

[27] Video Sequence Database: http://media.xiph.org/video/derf/

[28] X. Li and Y. Zheng, "Patch-based video processing: A variational Bayesian approach," *IEEE Trans. on Circuits Syst. Video Tech.*, vol. 19, no. 1, pp. 27–40, Jan. 2009. Article (CrossRef Link)

[29] A. Buades, B. Coll, and J. Morel, "Nonlocal image and movie denoising," *Int. J. Comput. Vision*, vol. 76, no. 2, pp. 123–139, 2008. Article (CrossRef Link)

[30] M. Protter, and M. Elad, "Image sequence denoising via sparse and redundant representations," *IEEE Trans. on Image Process.*, vol. 18, no. 1, pp. 27–35, Jan. 2009. Article (CrossRef Link)

[31] M. Elad, and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. on Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006. Article (CrossRef Link)

[32] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration,"

IEEE Trans. on Image Process., vol. 17, no. 1, pp. 53–69, Jan. 2008.
Article (CrossRef Link)

[33] F. Jin, P. Fieguth, and L. Winger, "Wavelet video denoising with regularized multiresolution motion estimation," *Eup. Assoc. Speech,Signal, Image Process. J. Appl. Singal Process.*, vol. 2006, no. 72705, pp. 1–11, 2006.

[34] D. Rusanovskyy and K. Egiazarian, "Video denoising algorithm in sliding 3-D DCT domain," in *Proc of ACIVS*, Sep. 2005, pp. 618–625. Article (CrossRef Link)

[35] N. Lian, V. Zagorodnov, and Y. Tan, "Video denoising using vector estimation of wavelet coefficients," in *Proc.of IEEE Int. Sym. Circuits Syst.*, May 2006, pp. 2673–2676.
Article (CrossRef Link)

Yu Liu received his BS from Northwestern Polytechnical University, Xi'an, China in 2005. He then received his MSc on image processing and PhD on computer graphics from University of East Anglia, Norwich, UK, in 2007 and 2011, respectively. He is currently a lecturer in the department of system engineering, National University of Defense Technology. His research interests include image/video processing, computer graphics, and visual-haptic technology.

Chenlin Zuo received the BS and MS degree in system engineering from National University of Defense Technology, Changsha, China, in 2008 and 2010 respectively. He is currently pursuing his PhD degree in control science and engineering from NUDT. His research interests include image/video denoise and computer vision.

Xin Tan received his BS from Sichuan University, Chengdu, China in 2007. He then received his MS on image processing from National University of Defense Technology, Changsha, China, in 2009. He is currently pursuing his PhD degree in control science and engineering from NUDT. His research interests include image/video denoise and image processing.

Huaxin Xiao received his BS degree in automation from University of Electronic Science and Technology of China. He is currently pursuing his MS degree in control science and engineering from National University of Defense Technology, Changsha, China. His research interests include sparse representation and computer vision.

Maojun Zhang received the BS and PhD degree in system engineering from National University of Defense Technology, Changsha, China, in 1992 and 1997 respectively. He is currently a professor in the department of system engineering, National University of Defense Technology. His research interests include computer vision, information system engineering, system simulation and virtual reality technology.