Video Denoising Using Spatio-temporal Filtering

Hongwei Zhu
Department of Fire Command
Chinese People's Armed Police
Force Academy
Langfang, Hebei 065000, China
zhw2006@126.com

Caijiao Xue
Department of Fire Command
Chinese People's Armed Police Force
Academy
Langfang, Hebei 065000, China
lf_xcj@126.com

Chunyan You
Image Communication Lab.
Chongqing Communication Institute
Chongqing 400035, China
bluesuny2002@126.com

ABSTRACT

Video denoising is very important to be used as a pre-processing stage, which should be considered both spatial and temporal dimensions. In this paper, video denoising scheme based on spatio-temporal filtering is researched, which contains spatially filtering in the two dimensional wavelet domain and temporal filtering in the time domain. Spatial filtering is firstly performed on individual video frames by wavelet shrinkage with an adaptive threshold selection. We use motion detection and recursive time weighted averaging to perform temporal filtering. It should be noted that both previous and following frames are considered to get the final filtered frame during the temporal filtering. Experimental results show that the proposed scheme yields better results in terms of both objective and subjective quality of the reconstructed video sequence.

Categories and Subject Descriptors

I.4.4 [**Image Processing and Computer Vision**]: Restoration

General Terms

Algorithms, Experimentation, Theory

Keywords

Video denoising, wavelet transform, temporal filtering

1. INTRODUCTION

Video image is often corrupted by noise due to the capturing or transmission process. Thus, it is desirable to remove noise without removing any important features of the original image. Video denoising is very important in numerous applications, such as restoration of old movies, television broadcasting systems and video surveillance. Noise reduction is also advantageous for the subsequent processing, especially for video coding, since it can considerably enhance coding efficiency and improve objective quality of video sequence. Most noise sources are considered to be additive white Gaussian noise model [1].

The wavelet transform for multiresolution analysis can remove

noise from images without excessive smoothing of actual discontinuities. Many denoising schemes by using wavelet transform for still image denoising have been developed [2-4], while video denoising based on wavelet transform is less studied so far. In the literature [5], Roosmalen proposed an algorithm for video denoising by thresholding the coefficients of a specific 3D multiresolution representation. Li and Selesnick studied wavelet thresholding in a non-separable 3D dual tree complex wavelet representation [6]. A scheme of motion adaptive filtering by using several successive video frames is proposed in [7], and literature [8] proposed a video denoising scheme, which combines 2D wavelet denoising and temporal filtering. However, there is a disadvantage for the scheme in [8] because only the previous frames are considered for temporal filtering.

Although wavelet shrinkage performs significantly better than most other commonly used denoising methods, visual quality of the denoised video frames can sometimes suffer from ringing type of artifacts, vallyes around the edges. In order to overcome this problem, the spatio-temporal filtering scheme for video denoising proposed in this paper contains spatial filtering in the 2D wavelet domain and temporal filtering in the time domain, where the temporal filtering is based on selective weighted averaging of the spatially filtered frames to form the final filtered frame. It should be noted that the temporal filter is improved by considering both previous and following frames during motion detection and temporal filtering which is different from the scheme in literature [8]. It is very similar to the concept of bi-directional frames used in many video coding standards. Experimental results demonstrate that the proposed method has an excellent denoising performance.

This paper is organized as follows. In Section 2, we describe spatial filtering by wavelet shrinkage and in Section 3, we discuss the temporal filtering scheme. Finally, experimental results are given in Section 4, followed by a conclusion in Section 5.

2. SPATIAL FILTERING

The problem of image denoising can be stated as follows: given the zero-mean observation data $d(i, j)$ as a noisy realization of the signal $f(i, j)$, the noisy version of the image is modeled as

$$d(i, j) = f(i, j) + n(i, j) \ \left(i = 1, \mathrm{K}, M, j = 1, \mathrm{K}, N \right) \quad (1)$$

where $M \times N$ denotes the image size. The goal is to estimate the original image data $f(i, j)$ from the noise observations $d(i, j)$ so that mean squared error is minimum. In this work, we assume that $n(i, j)$ is white Gaussian noise of zero mean and variance

σ^2. Due to linearity of the wavelet transform, the noise remains additive in the wavelet domain as well.

Wavelet-based methods for the removal of the noise without blurring the important features in images have received a great deal of attention. Generally, important features are characterized by large wavelet coefficients across the scales, while most of the noise power is confined to several fine scales, thus facilitating the separation of the feature-related coefficients and noise-related coefficients. The general procedure for wavelet-based denoising algorithms consists of the following steps: forward wavelet transform, modification of detail coefficients and inverse wavelet transform. The selections of threshold and thresholding function are two of the critical components in this process [9].

Thresholding is a simple non-linear method, which operates on one wavelet coefficient at each time. In the most basic form, each coefficient is thresholded by comparing against the threshold, if the coefficient is smaller than the threshold, set to zero, otherwise it is kept or modified. Replacing the small noisy coefficients by zero and inverse wavelet transform on the results may lead to reconstruction with the essential signal characteristics and less noise. The most common choices for the thresholding function are hard and soft thresholding function. Soft thresholding has been used over hard thresholding because it gives more visually pleasant images as compared with hard thresholding. In this paper, spatial filtering in the wavelet domain is performed by soft thresholding function with the data driven subband dependent threshold, which is adaptive to different subband characteristics. Threshold T for subband b of length L in R level wavelet decomposition is given by [10]

$$T_b = \sqrt{\ln(L/R)}\,\sigma^2 \Big/ \sqrt{\max\left(\sigma_b^2 - \sigma^2, 0\right)} \qquad (2)$$

where σ_b^2 is the variance of the subband b. Generally, the noise variance $\hat{\sigma}^2$ is not known, one usually estimates it as the median absolute deviation of the diagonal detail coefficients ω_{ij} on level 1 (i.e. subband HH$_1$) divided by 0.6745 [2].

$$\hat{\sigma}^2 = \left[\operatorname{median}\left(\left|\omega_{ij}\right|\right)\Big/0.6745\right]^2 \quad \omega_{ij} \in subband\,HH_1 \quad (3)$$

3. TEMPORAL FILTERING

As we have mentioned, temporal filtering can further enhance the performance of video denoising by using the temporal correlation of adjacent frames. This is achieved by performing motion detection and temporal filtering over every three adjacent spatially filtered frames. After denoising in the 2D wavelet domain, the interframe differences caused by remaining noise are small as compared with the actual interframe differences caused by changes of the video content along time axes [8]. Thus we can use a pixel-based motion detector and we make the recursive filtering work at those positions where motion is detected. Unlike

[8], where only the current frame and previous frames are used for temporal filtering, following frames are also considered in our approach.

Let $\mathbf{f_k}$ denotes the k-th frame of an original video sequence and $\mathbf{d_k}$ the corresponding noisy frame, equation (1) becomes

$$\mathbf{d_k} = \mathbf{f_k} + \mathbf{n_k} \qquad (4)$$

where $\mathbf{n_k}$ is the noise of the k-th frame. Further on, let

$$\hat{\mathbf{f}}_\mathbf{k}^{\mathbf{2D}} = \left[\hat{f}_k^{2D}(i,j)\right]_{M \times N} \qquad (5)$$

denotes the k-th denoised frame after spatial filtering, where $i = 0,1,\mathrm{K}, M-1$ and $j = 0,1,\mathrm{K}, N-1$. As mentioned above, the current frame, previous frames and following frames are all considered for motion detection and temporal filtering. For the purpose of computational complexity, the number of previous frames and following frames are all set to one. Assuming that the current frame is $\hat{\mathbf{f}}_\mathbf{k}^{\mathbf{2D}}$, besides this frame, we will consider both the previous frame $\hat{\mathbf{f}}_{\mathbf{k-1}}^{\mathbf{2D}}$ and the following frame $\hat{\mathbf{f}}_{\mathbf{k+1}}^{\mathbf{2D}}$. Our intention is to further improve the current frame $\hat{\mathbf{f}}_\mathbf{k}^{\mathbf{2D}}$ on the basis of motion information along time axis. Towards this end, we define the motion field $\mathbf{S_k}$ as

$$\mathbf{S_k} = \left[s_k(i,j)\right]_{M \times N} \qquad (6)$$

We can estimate $s_k(i,j)$ by using two difference intensities $d_{k-1,k}(i,j)$ and $d_{k,k+1}(i,j)$, which are defined as follows

$$\begin{cases} d_{k-1,k}(i,j) = \left|\hat{f}_{k-1}^{2D}(i,j) - \hat{f}_k^{2D}(i,j)\right| \\ d_{k,k+1}(i,j) = \left|\hat{f}_k^{2D}(i,j) - \hat{f}_{k+1}^{2D}(i,j)\right| \end{cases} \qquad (7)$$

For equation (7), the first expression represents the changes between $\hat{\mathbf{f}}_\mathbf{k}^{\mathbf{2D}}$ and $\hat{\mathbf{f}}_{\mathbf{k-1}}^{\mathbf{2D}}$ while the second one represents the changes between $\hat{\mathbf{f}}_\mathbf{k}^{\mathbf{2D}}$ and $\hat{\mathbf{f}}_{\mathbf{k+1}}^{\mathbf{2D}}$ at the spatial position (i,j). The motion field $\mathbf{S_k}$ can be estimated by

$$\hat{s}_k(i,j) = \begin{cases} 0 & d_{k-1,k}(i,j) < \lambda_1 \,\&\,\&\, d_{k,k+1}(i,j) < \lambda_2 \\ 1 & else \end{cases} \qquad (8)$$

with λ_1 and λ_2 being the motion threshold. The case of $\hat{s}_k(i,j) = 0$ illustrates that there is no obvious changes over $\hat{\mathbf{f}}_{\mathbf{k-1}}^{\mathbf{2D}}$, $\hat{\mathbf{f}}_\mathbf{k}^{\mathbf{2D}}$ and $\hat{\mathbf{f}}_{\mathbf{k+1}}^{\mathbf{2D}}$ at the spatial position (i,j), meaning that

$$\hat{f}_{k-1}^{2D}(i,j) \approx \hat{f}_k^{2D}(i,j) \approx \hat{f}_{k+1}^{2D}(i,j) \qquad (9)$$

let $\hat{s}_k(i,j)$ be 1 if there is obvious changes over $\hat{\mathbf{f}}_{\mathbf{k-1}}^{\mathbf{2D}}$, $\hat{\mathbf{f}}_\mathbf{k}^{\mathbf{2D}}$ and $\hat{\mathbf{f}}_{\mathbf{k+1}}^{\mathbf{2D}}$ at the position (i,j). At those spatial positions where

$\hat{s}_k(i,j) = 0$, we apply weighted averaging of those three adjacent denoised intensities to generate the final filtered intensity $\hat{f}_k^{3D}(i,j)$ as follows

$$\hat{f}_k^{3D}(i,j) = \alpha_1 \hat{f}_{k-1}^{2D}(i,j) + \alpha_2 \hat{f}_k^{2D}(i,j) + \alpha_3 \hat{f}_{k+1}^{2D}(i,j) \quad (10)$$

where the weighting factors α_1 , α_2 and α_3 satisfy

$$\begin{cases} \alpha_1 + \alpha_2 + \alpha_3 = 1 \\ \alpha_2 > \alpha_1, \alpha_2 > \alpha_3 \end{cases} \quad (11)$$

the size relationship between α_1 and α_3 is determined by

$$\begin{cases} \alpha_1 < \alpha_3 & d_{k-1,k}(i,j) > d_{k,k+1}(i,j) \\ \alpha_1 > \alpha_3 & d_{k-1,k}(i,j) < d_{k,k+1}(i,j) \end{cases} \quad (12)$$

At those positions where motion is detected (i.e. $\hat{s}_k(i,j) = 1$), the temporal filtering is closed and the spatial filtered value is remained: $\hat{f}_k^{3D}(i,j) = \hat{f}_k^{2D}(i,j)$. Thus the final filtered frame $\hat{\mathbf{f}}_\mathbf{k}^{\mathbf{3D}}$ after spatio-temporal filtering is given briefly as follows

$$\hat{\mathbf{f}}_\mathbf{k}^{\mathbf{3D}} = \begin{cases} \alpha_1 \hat{\mathbf{f}}_{\mathbf{k-1}}^{\mathbf{3D}} + \alpha_2 \hat{\mathbf{f}}_\mathbf{k}^{\mathbf{2D}} + \alpha_3 \hat{\mathbf{f}}_{\mathbf{k+1}}^{\mathbf{2D}} & \hat{s}_k(i,j) = 0 \\ \hat{\mathbf{f}}_\mathbf{k}^{\mathbf{2D}} & \hat{s}_k(i,j) = 1 \end{cases} \quad (13)$$

4. EXPERIMENTAL RESULTS

Similations have been performed on the three standard CIF test sequences named Miss American, Mother&daughter and Salesman, each of which is corrupted by additive white Gaussian noise with standard deviations σ=10, 15, 20 and 30. Table 1 shows the performance of three test sequences at different noise levels and the corresponding average peak signal to noise ratios (PSNR/dB) for 150 frames of 352 by 288. In our experiments, we got the following values: $\lambda_1 = \lambda_2 = 20$, $\alpha_2 = 0.5$ and $\alpha_1 = 0.2$ if $\alpha_1 < \alpha_3$, otherwise $\alpha_1 = 0.3$.

Table 1. Performance table for three test sequences

	Miss American	Mother&daughter	Salesman
σ=10	37.8864	36.7889	34.7385
σ=15	36.4585	35.2665	32.2412
σ=20	35.0375	33.9664	29.8476
σ=30	32.4678	31.2458	27.2112

In terms of visual quality, figure 1 shows the results of Miss American sequence by using the proposed scheme and spatial

(a)   (b)

filtering in the wavelet domain without temporal filtering. The advantage of the proposed filter is even more evidence when viewing video, furthermore, the increase of computational complexity with respect to the just spatial filtering in the 2D wavelet domain is not significant.

5. CONCLUSION

Video denoising has become an important task, and has found its usage in many fields. In this paper, an improved spatio-temporal filtering for video denoising is proposed, which combines the two dimensional wavelet denoising and temporal filtering. The results show that spatial filtering in the wavelet domain can cause an unpleasant effect due to the lack of temporal filtering along time axis.Temporal filtering can further improve the results of the spatial filtering, which is performed on adjacent denoised frames. The scheme of temporal filtering is very simple to implement and computationally more efficient. Experimental results demonstrate that the spatio-temporal filtering outperforms the other techniques. However, there are also some problems for future work, such as the design of advanced spatial filtering techniques and temporal filter. In the future work, we intend to employ the motion compensation to design the temporal filter. However, how to decrease the computational complexity also needs further research.

6. ACKNOWLEDGMENTS

REFERENCES

[1] Brailean, J., Kleihorst, R., Efstratiadis, S., et al. Noise reduction filters for dynamic image sequences. *Proceedings of the IEEE, 83,* 9 (1995), 1272-1292.

[2] Chang, S. G., Yu, B., and Vetterli, M. Spatially adaptive wavelet thresholding with context modeling for image denoising. *IEEE Trans. Image Processing, 9,* 9 (2000), 1522-1531.

[3] Crouse, M. S., Nowak, R. D., and Baranuik, R. G. Wavelet-based statistical signal processing using hidden markov models. *IEEE Trans. Signal Processing, 46,* 4 (1998), 886-902.

[4] Crouse, M. S., Nowak, R. D., and Baranuik, R. G. Wavelet-based statistical signal processing using hidden markov models. *IEEE Trans. Signal Processing, 6,* 12 (1999), 300-303.

[5] Simoncelli, E. P., Freeman, W., Adelson, E. et al. Shiftable multiscale transforms. *IEEE Trans. Information Theory, 38,* 2(1992), 587-607.

[6] Selesnick, I., and Li, K. Video denoising using 2D and 3D dual-tree complex wavelet transforms. In *Proceedings of SPIE in Wavelet Applications in Signal and Image Processing* (Aug. 4-8, 2003). SPIE Press, San Diego, 2003, 607-618.

[7] Zlokolica, V., Pizurica, A., and Philips, W. Video denoising using multiple class averaging with multiresolution. *Lectures Notes in Computer Science of Springer Verlag. 2849*(2003), 172-179.

[8] Pizurica, A., Zlokolica, and V., Philips, W. Combined wavelet domain and temporal video denoising. In *Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS'03)* (July 21-22, 2003). IEEE Press, Miami, 2003, 334-341.

[9] Johnstone, I. M., and Donoho, D. L. Adapting to smoothness via wavelet shrinkage. *Journal of the Statistical Association, 90,* 432(1995), 1200-1224.

[10] Rajpoot, N., Yao, Z., Wilson, R. Adaptive wavelet restoration of noisy video sequences. In *Proceedings of the International Conference on Image Processing (ICIP'04)* (Oct. 24-27, 2004), IEEE Press, Singapore, 2004, 957-960.

.