LETTER

Movement Awareness-Adaptive Spatio Temporal Noise Reduction in Video

Sangwoo AHN[†a)], Jongjoo PARK[††], Linbo LUO[†††], *Nonmembers*, *and* Jongwha CHONG[††], *Member*

SUMMARY In this letter, we present an efficient video matching-based denoising method. Two main issues are addressed in this paper: the matched points and the denoising algorithm based on an adaptive spatial temporal filter. Unlike previous algorithms, our method adaptively selects reference pixels within spatially and temporally neighboring frames. Our method uses more information about matched pixels on neighboring frames than other methods. Therefore, the proposal enhanced the accuracy of video denoising. Simulation results show that the proposed method produces cleaner and sharper images.

key words: *denoising, video matching, ASTA, bilateral filter*

1. Introduction

Video databases are contaminated by noise during acquisition and transmission. Video denoising is highly valuable because it can enhance perceived video quality, facilitate the reduction of transmission bandwidth, and improve the accuracy of possible subsequent processes such as feature extraction and pattern analysis.

Video denoising methods can be classified into two comprehensive criteria: one is the implementation domains, the spatial domain or the transform domain; the other is the utilization of motion information [1]. Spatial domain methods are usually processed with weighted averaging within frames, where the weights can be adaptively fixed based on the local information [2]. Transform domain methods are processed by de-correlating the noisy signal using a linear transform and recovering the original signal, followed by an inverse transformation [3], [4].

Motion information or temporal correlations can be incorporated by employing an advanced or adapted transform [5] or by using an advanced statistical model reflecting the joint distributions of wavelet coefficients over space and time [6]. Nonlocal patch-based methods are being researched [7]–[10], in which motion information is incorporated implicitly by adaptively clustering similar 2-D or 3-D patches.

In this paper, we propose a new video denoising method based on the spatial domain and motion-recognized video matching in the temporal domain. The proposed method is based on the adaptive spatio-temporal accumulation (ASTA) filter [2]. ASTA is an effective video denoising method using a 2-D bilateral filter in the spatial domain and a 1-D bilateral filter in the temporal domain. However, a limitation exists in the temporal domain, because ASTA uses pixels in the same position in consecutive frames. If the camera or objects are moved, the temporal filter of ASTA is useless. Therefore, we enhance the information of consecutive frames by video matching, which improves the reliability of denoising performance.

2. Related Work

2.1 Adaptive Spatio-Temporal Accumulation (ASTA) Filter

This section describes the conventional ASTA before inclusion of the proposed algorithm [2]. Two issues regarding the ASTA filter are discussed: the number of pixels to be combined and the motion of corresponding pixels. The ASTA combines temporal-only and spatial-only bilateral-inspired filtering with fixed parameters based on local illumination.

There are three major parts of the ASTA: the spatial filter, the temporal filter, and the similarity distance assessment. The spatial filter is the well-known, edge-preserving, bilateral filter shown below.

$$B(s,\sigma_h,\sigma_i) = \frac{n_*}{d_*} = \frac{\sum\limits_{p\in N_s} g\left(\|p-s\|,\sigma_h\right) g(D(p,s),\sigma_i) I_p}{\sum\limits_{p\in N_s} g\left(\|p-s\|,\sigma_h\right) g(D(p,s),\sigma_i)} \tag{1}$$

where s is the base pixel, p is the neighboring pixel, N_s is the set of neighboring pixels of s, σ_h is the control value of the spatial Gaussian, and σ_i is the control value of the. D is the function of similarity distance described below. n_* and d_* represent numerator, denominator respectively.

$$D(p_{xyt},s_{xyt}) = \frac{\sum\limits_{x=s_x-n}^{s_x+n}\sum\limits_{y=s_y-n}^{s_y+n} g\left(\left\|x-p_x, y-p_y\right\|,\sigma_e\right)\left|I_{x,y,pt}-I_{x,y,st}\right|}{\sum\limits_{x=s_x-n}^{s_x+n}\sum\limits_{y=s_y-n}^{s_y+n} g\left(\left\|x-p_x, y-p_y\right\|,\sigma_e\right)} \tag{2}$$

Manuscript received June 27, 2013.
Manuscript revised September 11, 2013.

[†]The author is with the Department of Nanoscale Semiconductor Engineering, Hanyang University, 17 Haengdang-Dong, Seongdong-Gu, Seoul, 180–8585, Rep. of Korea.

[††]The authors are with the Department of Electronics and Computer Engineering, Hanyang University, 17 Haengdang-Dong, Seongdong-Gu, Seoul, 180–8585, Rep. of Korea.

[†††]The author is with the Department of Science and Technology, China University of Geosciences, Wuhan, P.R. China, 430074.

a) E-mail: ahnsangwoo@hanyang.ac.kr (Corresponding author)

DOI: 10.1587/transinf.E97.D.380

Copyright © 2014 The Institute of Electronics, Information and Communication Engineers

where σ_e is a temporal edge tolerance control value.

The temporal filter is similar to the spatial bilateral filter; however, the temporal filter is a 1-D bilateral filter based on temporally consecutive pixels at the same location.

ASTA is a voting scheme between the spatial and temporal filters; each vote is a measure of support for the filter. Voting scheme means that processing algorithms are elected by voting, in here, voting is determined by the result of temporal and spatial bilateral filter.

Before ASTA is run on a pixel, we determine how many pixels are required (defined as λ). n_* and d_* representing numerator, denominator of bilateral filters are needed before. ASTA is implemented as shown below.

$$\begin{aligned} \frac{n_T}{d_T} &= temporalBilateral(x, y, t, \sigma_h, \sigma_i) \\ \frac{n_S}{d_S} &= spatialBilateral(x, y, t, \sigma_h, \sigma_i) \\ w &= \lambda \times g(0, \sigma_h) \times g(0, \sigma_i) \end{aligned} \tag{3}$$

$$ASTA(x, y, t, \lambda) = \begin{cases} \frac{n_T}{d_T}, & d_T \geqslant w \\ \frac{n_T + n_S}{d_T + d_S}, & d_T < w \& d_T + d_S < w \\ \frac{n_T + n_S\,({}^{w-d_T}/_{d_S})}{w}, & d_T < w \& d_T + d_S \geqslant w \end{cases} \tag{4}$$

According to the voting scheme, information from the temporal filter and spatial filter produce denoising in each pixel. The ASTA is a well-known video denoising method that has been widely researched in various applications. However, there is room for improvements. This research focuses on the second voting scheme, which shows that both similarity distances of the temporal and spatial filters are lower than the expected values. In this scheme, information about temporal and spatial is unreliable. Therefore, we suggest a method for supplying more information in that case.

2.2 Speeded Up Robust Features (SURF)

In this research, we use SURF to match consecutive points. SURF is a robust translation, rotation, and scale invariant representation algorithm. It extracts the key points and points of interest from target images. Each of the extracted key points contains the coordinate location of the point, laplacian value, size of the feature, direction, value of hessian, and its descriptor [11]. SURF has been successfully applied to object recognition with the renowned algorithm Scale Invariant Feature Transform (SIFT).

As mentioned above, we researched a method to supply more information for the second voting scheme in ASTA. Object detectors like SURF can improve the information. SURF has the characteristics which are scale invariant and rotation invariant. Therefore, in consequence frames, SURF can compensate the difference caused by camera rotation or shaking. The matched points on consecutive frames provide reliable information that does not exist in the original ASTA. Therefore, we can improve the accuracy of ASTA using matched points with SURF.

3. Proposed Method

The two methods mentioned above construct the main flow of the proposed method. However, we must consider the combination of the two methods and the utilization of matched points; these are crucial research topics in video denoising. Therefore, we focus on a solution to utilize matched points.

Before solving the problem, we must recognize the characteristics of matched points. Their matching sets involve each point on the frame being the same on the other frames in the matching set.

In this letter, the matched points are re-evaluated by the proposed method for introducing to de-noising filter. After that, the movement awareness-temporal bilateral filter which is core difference with original ASTA is approximated with the re-evaluated matched points. Furthermore, against to original ASTA, voting schemes of the proposed de-noising method are categorized into four sections for more reliable results.

The reason why we introduce movement awareness method can be briefly explained here. In a temporal-bilateral filter on ASTA, points on the same coordinate of consecutive frames are used in a one-dimensional temporal filter. This means that, if the object moves from one point to another, then the temporal filter is useless, and a spatial-bilateral filter must be used, which may involve a blurring effect. Moreover, natural scenes captured by video cameras tend to involve moving objects, and due to hand shake, there is a slight difference between neighboring frames. As a result, matching points make advance from the temporal-bilateral filter to the movement awareness temporal-bilateral filter as shown in Fig. 1. In Fig. 1, we can see that the MA-temporal filter can be used as temporal filter with moving areas of consequence frames, it makes the usages of spatial filter reduced while temporal filter cannot.

3.1 Movement Awareness Temporal-Bilateral Filter

We focus on the movement awareness temporal-bilateral filter, which is the main method in which the proposed algorithm differs from the original ASTA. Explaining the movement awareness (MA) temporal-bilateral filter, we suppose that the matched points on consecutive frames are deter-

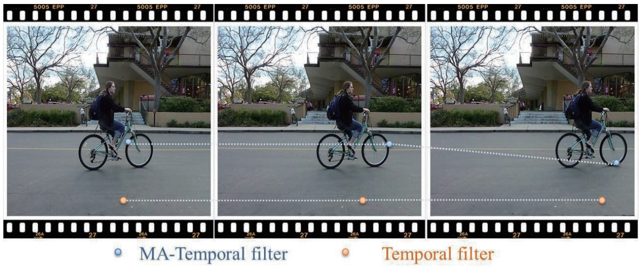


Fig. 1 Example of difference reference pixels.

mined by SURF in advance.

The MA temporal-bilateral filter differs from the regular temporal-bilateral filter: distance estimation between neighboring frames and reference pixels for denoising.

First, a distance estimation method for weights of MA-ASTA between neighboring frames is proposed to improve accuracy. In the *MA* temporal-bilateral filter, matched points from SURF are highly reliable; however, weight-based distance (rather than direct introduction) is more reliable. Weights are calculated by the similarity distance between matched points and their neighbor pixels as follows.

$$D_T(p,q) = \frac{\sum_{q\in S} g_\sigma\left(\left\|p_x - q_x, p_y - q_y\right\|\right)\left|I_p - I_q\right|}{\sum_{q\in S} g_\sigma\left(\left\|p_x - q_x, p_y - q_y\right\|\right)} \tag{5}$$

where D_T is the similarity distance on a *MA* temporal bilateral filter, p and q are matched points in distinctive frames, and S is the set of neighboring pixels centered at q. This becomes the weight applied to the spatial distance of a *MA* temporal-bilateral filter. Therefore, spatial distances in the *MA* temporal filter are on a three-dimensional plane with x-y coordinates and D_T. In the result, the *MA* temporal-bilateral filter can be shown as follows.

$$\begin{aligned} & MA - TBF(p, \sigma_h, \sigma_i) \\ & = \frac{n_*}{d_*} = \frac{\sum_{q\in S} g(z_s(p,q), \sigma_{h'}) g\left(\left|I_p - I_q\right|, \sigma_{i'}\right) I_q}{\sum_{q\in S} g(z_s(p,q), \sigma_{h'}) g\left(\left|I_p - I_q\right|, \sigma_{i'}\right)} \end{aligned} \tag{6}$$

where *MA*-T*BF* represents the *MA* temporal-bilateral filter, $\sigma_{x'}$ is the control value, n_* and d_* represent numerator, denominator of bilateral filter respectively, and z_s is the virtual three-dimensional spatial distance with the axis of x-coordinate, y-coordinate and the evaluated value in advance as follows.

$$z_s = sqrt((p_x - q_x)^2 + (p_y - q_y)^2 + D_T^2) \tag{7}$$

In conclusion, the proposed *MA* temporal-bilateral filter is a virtual three-dimensional bilateral filter on a modified temporal scale.

3.2 Movement Awareness Adaptive-Spatio Temporal Filter

The proposed Movement Awareness Adaptive Spatio Temporal Filter (MA-ASTA) also has a voting scheme among the spatial filter, temporal filter, MA-temporal filter and the other, where each vote is a measure of support for the filter. The differences between ASTA and MA-ASTA are the number of voting stages because of the MA-temporal filter and their related filter formulations. In ASTA, there is an insufficient information stage; to overcome this disadvantage, MA-ASTA has two stages (with and without a matched point set). MA-ASTA is implemented as follows.

$$\begin{aligned} \frac{n_T}{d_T} &= temporalBilateral(x, y, t, \sigma_h, \sigma_i) \\ \frac{n_S}{d_S} &= spatialBilateral(x, y, t, \sigma_h, \sigma_i) \\ \frac{n_{MT}}{d_{MT}} &= MA - temporalBilateral(x, y, t, \sigma'_h, \sigma'_i) \end{aligned} \tag{8}$$

$$\begin{aligned} & w = \lambda \times g(0, \sigma_h) \times g(0, \sigma_i) \\ & MA - ASTA(x, y, t) \\ & = \begin{cases} \dfrac{n_T}{d_T}, & d_T \geqslant w \\ \dfrac{n_S + n_{MT}(e - d_S / d_{MT})}{w}, & d_T + d_S < w \& (x, y, t) \notin M \\ \dfrac{n_{MT}}{d_{MT}}, & d_T + d_S < w \& (x, y, t) \in M \\ \dfrac{n_T + n_S(e - d_T / d_S)}{w}, & d_T < w \& d_T + d_S \geqslant e \end{cases} \end{aligned} \tag{9}$$

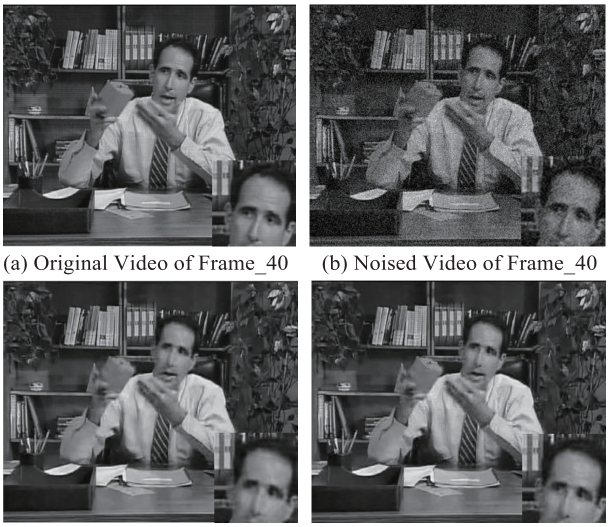
Filters are selectively applied to denoising according to the voting scheme. Voting is based on the original ASTA to maintain the verified accuracy and set of matched points. In temporal filter and MA-filter, the result of each filters are directly used in the final, however, in spatial filter and the other, the final results are re-arranged for reducing unintentional artifacts. In MA-ASTA, the number of pixels in the insufficient information stage is less than in ASTA due to adoption of matched points.

4. Simulation Results

Simulations are based on a video denoising database published by the Department of Signal Processing on the homepage of Tampere University of Technology. The database provides the famous video sequences such as coastguard, foreman, tennis, bicycle, bus, flower, missa, and salesman. The proposed algorithm is implemented using MATLAB R2012a.

In the simulation, we demonstrate that the proposed MA-ASTA outperforms the original ASTA. The first aim of the simulation is to show that MA-ASTA has better movement-awareness performance than the original ASTA. Therefore, the video sequence with noise and camera movement is simulated on both ASTA and MA-ASTA. The sample video sequences are flower, which has moving frames and artificial Gaussian noise, and coastguard, which has a moving object with a stable background and artificial Gaussian noise. The second aim of the simulation is to show that MA-ASTA can be applied generally to various video sequences.

Figure 2 shows the 40th frame of the original video, artificially noised video, video de-noised by ASTA, and video de-noised by the proposed MA-ASTA. The result of the proposed MA-ASTA is closer to the original video than is the result of ASTA. The prominent difference is noticed around the left ear and the left eye of the man enlarged in the video. In ASTA, the left ear is unclear; however, in MA-ASTA, the left ear is clearly observable. This is because MA-ASTA provides the information of the left ear from the previous se-



(a) Original Video of Frame_40 (b) Noised Video of Frame_40

(c) ASTA VIDEO of Frame_40 (d) MA-ASTA Video of Frame_40

Fig. 2 Comparison of ASTA with the proposed algorithm.

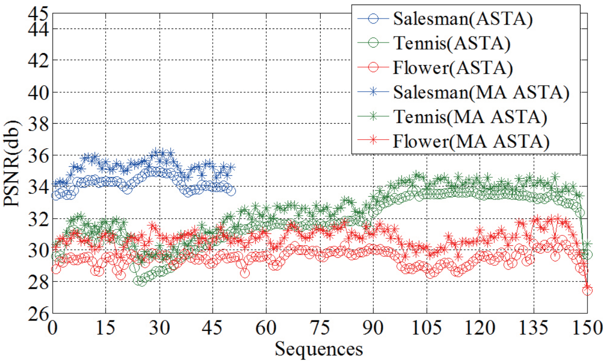


Fig. 3 PSNR evaluation (ASTA vs. MA-ASTA).

quence despite the movement. The simulation demonstrates that the MA-ASTA performs better with slightly moving objects.

Figure 3 shows the PSNR scores of ASTA and MA-ASTA on various test sequences. The PSNR scores are improved by 1.04 dB, 0.83 dB, and 1.28 dB on *Salesman*, *Tennis*, and *Flower* video sequences, respectively. Most of the MA-ASTA scores in the figure are higher than the ASTA scores, though a few are similar or slightly lower.

5. Conclusion

In this paper, we proposed a movement-awareness adaptive spatio-temporal filter for a video denoising algorithm. The main contributions of this paper are the similarity distance evaluation methods and filter scheme. A simulation verified that the proposed MA-ASTA is more accurate than ASTA through movement awareness steps for detail information.

Acknowledgments

This research was supported by Hynix Semiconductor Inc. This work was sponsored by ETRI SW-SoC R&BD Center, Human Resource Development Project.

References

[1] G. Varghese and Z. Wang, “Video Denoising Based on a Spatiotemporal Gaussian Scale Mixture Model,” IEEE Trans. Circuits Syst. Video Technol., vol.20, no.7, pp.1032–1040, 2010.

[2] E.P. Bennett and L. McMillan, “Video enhancement using per-pixel virtual exposures,” ACM Trans. Graphics, vol.24, no.3, pp.845–852, 2005.

[3] D.L. Donoho, “De-noising by soft-thresholding,” IEEE Trans. Inf. Theory, vol.41, no.3, pp.613–627, 2005.

[4] E.P. Simoncelli and E.H. Aden, “Noise removal via Bayesian wavelet coring,” IEEE International Conference on Image Processing, vol.1, pp.379–387, 1996.

[5] W.I. Selesnick and K.Y. Li, “Video denoising using 2-D and 3-D dual-tree complex wavelet transforms,” Wavelet Applications in Signal and Image Processing X (SPIE), vol.5207, pp.607–618, 2003.

[6] S.M.M. Rahman, M.O. Ahmad, and M.N.S. Swamy, “Video denosing based on inter-frame statistical modeling of wavelet coefficients,” IEEE Trans. Circuits Syst. Video Technol., vol.17, no.2, pp.187–198, 2007.

[7] K. Dabov, A. Foi, and K. Egiazarian, “Video denoising by sparse 3-D transform-domain collaborative filtering,” IEEE Trans. Image Process., vol.16, no.8, pp.2080–2095, 2007.

[8] J. Boulanger, C. Kervrann, and P. Bouthemy, “Space-time adaptation for patch-based image sequence restoration,” IEEE Trans. Pattern Anal. Mach. Intell., vol.29, no.6, pp.1096–1102, 2007.

[9] M. Protter and M. Elad, “Image sequence denoising via sparse and redundant representations,” IEEE Trans. Image Process., vol.18, no.1, pp.27–36, 2009.

[10] X. Li and Y. Zheng, “Patch-based video processing: A variational Bayesian approach,” IEEE Trans. Circuit Syst. Video Technol., vol.19, no.1, pp.27–40, 2009.

[11] H. Bay, A/ Ess, T. Tuytelaars, and L.V. Gool, “Speededup robust features (SURF),” J. Comput. Vis. Image Understand., vol.110, no.3, pp.346–359, 2008.