Video Denoising Algorithm via Multi-scale Joint Luma–Chroma Bilateral Filter

Yuanyuan Gao[#], Hai-Miao Hu[#1], Jiawei Wu[#]

[#]Beijing Key Laboratory of Digital Media, School of Computer Science and Engineering, Beihang University, Beijing 100191
[1]Corresponding Author: frank0139@163.com

***Abstract*— Video denoising is important for display and subsequent analysis, but remains to be a challenging problem. Key insights that limit the performance of algorithms include two main aspects. First, low-frequency scene information and the coarse-grained noise in the chroma are mixed with each other, which is different from that in the luma. Thus, denoising the chroma by using only its own information is difficult. Second, it is impossible to directly use the denoised luma edge information to guide the chroma denoising for high quality results. Therefore, in this paper, we propose a multi-scale joint luma-chroma bilateral filter to improve chroma denoising performance. Experimental results demonstrate that the proposed algorithm surpasses state-of-the-art algorithms in most cases.**

***Index Terms*— Video Denoising, Chroma denoising, Bilateral Filter, Image enhancement, Structured noise**

I. INTRODUCTION

The recent advances in image denoising can achieve appropriate denoising results. However, video denoising remains a challenge, since the origins of the degradation in videos are complicated due to the video acquisition process.

Recently, many algorithms for removing noise in videos have been proposed. Some common algorithms, such as Re-BF [1] and K-nearest NLM [2], are to simply use standard grayscale video denoising algorithms on R, G, B components of the video, in effect, treating each color component as a separate grayscale video.

Other algorithms perform denoising in the perceptually uniform luma-chroma color space [3-6]. For example, in [3], stronger noise filtering could be applied to chroma components because chroma noise is more objectionable than luma noise. In [4, 5], considering the chroma would suffer from the lower SNR and the grouping is sensitive to the level of noise, researchers proposed to perform the grouping only once for the luma and reuse exactly the same grouping when applying collaborative filtering on the chroma.

These video denoising algorithms greatly improve the performance, however, they fail to achieve satisfactory results in removing the structured noise (as shown in Fig.1), which has strong spatial correlation and can be mixed with video signals.

The structured noise of videos mainly results from the video acquisition process. After tricolor signals separated from the sensor, noise remains in all components of the signal, namely the luma component (Y) and two chroma components (Cb, Cr). Usually, chroma components are down-sampled for compression because of bandwidth limitation and up-sampled for display, which may weaken the high-frequency characteristic of chroma noise. Hence, chroma noise is coarse-grained. Note that the noise in the luma component remains high-frequency characteristic, since there is not down-sampling for the luma component. This conclusion will be confirmed by the experiment in the following section.

Fig.1 Structured noise contained in videos.

The structured noise is difficult to be suppressed due to the different characteristics between the luma component and chroma components, which can be concluded to the following two aspects.

On one hand, when comparing with the luma, chroma components mainly contain low-frequency information. Thus, the chroma noise is spatial diffusion and real chroma edge information is weakened. Stronger chroma noise filtering leads to color bleeding. Although state-of-the-art denoising algorithms [4, 5] have noticed that chroma components need more consideration and luma information is used to guide chroma aggregation, they only use the edge information within the chroma for chroma denoising. Since low-frequency scene information and coarse-grained noise are mixed with each other, there is no discrimination during the blurring step between chroma noise and chroma scene detail. The denoising algorithms have not yet reached the limit of noise removal. Video chroma denoising with inadequate edge information remains a challenge.

On the other hand, although the luma contains enough edge information, there is not an obvious correlation between the chroma and the luma. Thus, it is impossible to directly use the denoised luma edge information to guide the chroma denoising for high quality results.

Therefore, a multi-scale joint luma-chroma bilateral filter is proposed to denoise the chroma by fully considering its characteristics. Instead of independently considering edge information of the chroma, edge information within the

denoised luma is referred during chroma denoising. Moreover, given that the luma and the chroma contain different strength edges, edge information of the luma is extracted through multi-scale bilateral functions, and each scale of edge information is utilized to denoise chroma through joint filter. Each denoised chroma component maintains different intensity of edge information, and all the denoised chroma components are combined through an average gain to reconstruct the final denoised chroma. The experimental results demonstrate that the proposed algorithm improves the chroma denoising performance by comparing with state-of-the-art algorithms.

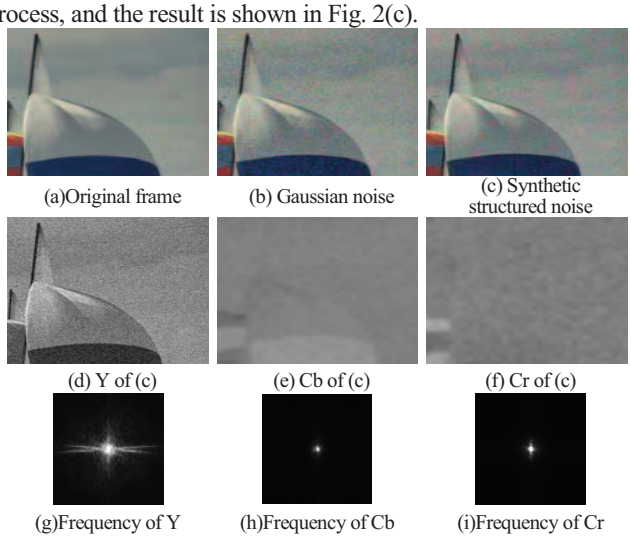
Note that since this paper focuses on the chroma denoising, 4-D collaborative filtering in transform domain [5] is directly employed for the luma denoising due to its high performance.

II. Observations and Justification

Some observation related to the above-mentioned problems and justifications to the proposed approach are discussed as follows.

Two chroma signals are usually compressed with low bandwidth in PAL/NTSC color encoding process to transmit signals conveniently. When the signal is digitized, down-sampling and other compression techniques can be used on the chroma. Thus, if the video contains noise, the luma and the chroma would degrade in different degrees.

Using a full chroma sampling frame with Gaussian noise (as shown in Fig.2 [b]), we compress the chroma with low-frequency Fourier filter and down-sampling to simulate the acquisition process, and the result is shown in Fig. 2(c).



(a)Original frame (b) Gaussian noise (c) Synthetic structured noise

(d) Y of (c) (e) Cb of (c) (f) Cr of (c)

(g)Frequency of Y (h)Frequency of Cb (i)Frequency of Cr

Fig.2 Generation of the structured noise.

After decomposition, coarse-grained noise occupies the chroma, which is different from the noise characteristic in the luma. Furthermore, compared with rich and sharp edges in the luma, edges in the chroma are seriously weakened. The noise cannot be removed with a small radius kernel. However, blurring through a large radius kernel leads to color bleeding because of the obscure edges. Figs. 2(h)–2(i) illustrate that the signal and noise are merged when chroma components are represented in the frequency domain.

III. The Proposed Algorithm

Our denoising algorithm is based on the preceding analysis. In this paper, the video is firstly converted into YCbCr color space. Inspired by the joint bilateral filter algorithm [7], the denoised luma is introduced as an auxiliary image for chroma denoising based on the observation that the luma and the chroma have similar edges.

In some situations, when the edge of luma is relatively weak, the chroma cannot be well reconstructed as expected. Furthermore, given that the joint bilateral filter is an edge-preserving operator, it can lead to a sharp edge in the chroma. To estimate the chroma accurately, joint bilateral filter should be extended to multi-scale form and a final denoised chroma can be reconstructed by combining all the information of each scale. More details are given as follows.

A. Edge Information Extraction from Luma

Let Ω be a local patch centered at (p,q), $(i,j)\in\Omega$ and l be the intensity of the denoised luma. The edge information e in the nth scale around (p,q) can be extracted through the bilateral function as follow:

$$e_n(i,j)=\frac{1}{W_n(p,q)}f_n(i,j,p,q)\,g_n\left(l_{i,}l_{j,}l_p,l_q\right),\quad (i,j)\in\Omega \tag{1}$$

The normalization factor $W_n(p,q)$ ensures pixel weights sum to 1.0:

$$W_n(p,q)=\sum_{(i,j)\in\Omega}f_n(i,j,p,q)\,g_n\left(l_{i,}l_{j,}l_p,l_q\right) \tag{2}$$

The functions f_n and g_n can be expressed by:

$$f_n(i,j,p,q)=\exp\left(-\frac{(i-p)^2+(j-q)^2}{2\sigma_{d_n}^2}\right),\quad (i,j)\in\Omega \tag{3}$$

$$g_n\left(l_{i,}l_{j,}l_p,l_q\right)=\exp\left(-\frac{(l_{ij}-l_{pq})^2}{2\sigma_{r_n}^2}\right),\quad (i,j)\in\Omega \tag{4}$$

where σ_{d_n} and σ_{r_n} are edge-stopping parameters about the spatial distance and the intensity distance. In each scale, they are set to different values. When the spatial distance of pixels in the luma is larger than σ_{d_n}, or the intensity distance of pixels in the luma is larger than σ_{r_n}, the bilateral function value is small, which means that pixels are likely to cross an edge. In other words, the bilateral function values of luma and the bilateral function values of chroma in one small patch have strong correlation. The bilateral function values imply the possibilities that whether the center pixel and the surrounding pixels are across an edge. The smaller the value is, the greater the possibility is.

B. Chroma Denoising Based on Multi-scale Joint Luma-Chroma Bilateral Filter

Based on the extracted edge information around (p,q) in the luma, a maximum extent local weighted averaging with chroma information can be carried out by using bilateral function values. Let c be the noisy chroma. The denoised chroma pixel $c(p,q)$

in the *n*th scale can be generated as:

$$c_n(p,q)=\sum_{(i,j)\in\Omega} e_n(i,j)c(i,j) \tag{5}$$

Let *m* be the number of scales. The final reconstruction of the chroma pixel $c'(p,q)$ can be generated as:

$$c'(p,q)=\frac{1}{m}\sum_{n=1}^{m} c_n(p,q) \tag{6}$$

Fig. 3 shows a 1-D example of the chroma denoising process. As shown in figure 3(left), affected by noise, the edge of the chroma and noise are indistinguishable. However, the edge of the denoised luma is clearly visible. Using the edge information e between the center pixel and surrounding pixels of the denoised luma, pixels on one side of the edge are averaged together but pixels across an edge are almost not averaged together. So, the noise of the chroma is cleaned out but edges are retained. Denoised chroma components with two different scales are illustrated in the right of Fig.3. As shown in the green rectangle, in most cases, choosing different scales, the estimate signal will fluctuate around the real signal up and down. With an average gain the chroma can be better reconstructed.

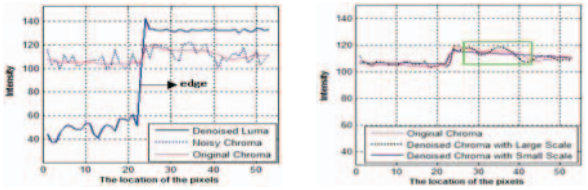


Fig. 3 1-D illustration for chroma denoising.

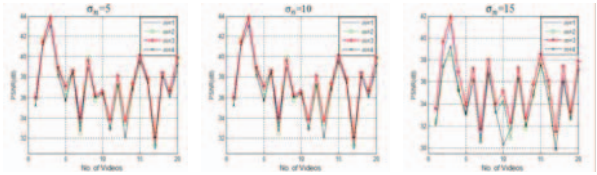


Fig. 4 Comparison results on standard test videos (σ_n is the noise standard deviation).

TABLE I AVERAGE PSNR with DIFFERENT *m* in FIG.3

σ_n	m=1	m=2	m=3	m=4
5	37.28	37.23	**37.53**	36.81
10	35.56	35.43	**35.83**	35.22
15	34.79	34.56	**35.75**	34.34

Moreover, the scale number *m* is important for the results. In order to determine the optimal value of *m*, the intermediate experiment was carried out based on the standard test videos[1] and the experiment setting was as follows.

In order to simulate the low-frequency noise in the chroma, we generated a Gaussian noise matrix which was a quarter of the size of the video and then up-sampled it. Then we divided σ_{d_n} into four grades for different scales: 5, 10, 15, 20. We applied different values for the parameters *m* and σ_{r_n}. The best average PSNR results of Cb and Cr in each scale are given in Fig.4. It shows that the joint luma-chroma bilateral filter with 3 scales achieves the best performance. Especially when the noise is more serious, the advantage is more obvious. When the scale number is bigger than 3, the performance is gradually degraded. The average PSNR of different scales in Fig.4 are included in Table1. It can improve the PSNR when compared with the best results of one scale, two scales, and four scales, respectively. Therefore, the scale number *m* is set to be 3 in this paper.

Note that although the proposed multi-scale form is an extension form of the single-scale, the selection of σ_{d_n} and σ_{r_n} is consistent. The parameter selection procedure in detail can be found in [3].

IV. Experimental Results

In this section, the proposed algorithm is evaluated by comparing with M-BF [3], VBM4D [5], shift-able bilateral Filter (St-BF) [8] and Su-NLM [9]. Reference parameters are: σ_{d_1}=5, σ_{d_2}=10 and σ_{d3}=20, σ_{r_1}=7, σ_{r_2}=15 and σ_{r_3}=20.

TABLE II COMPARISON RESULTS of CB

Frame name	σ_n	PSNR				
		M-BF	VBM4D	St-BF	Su-NLM	Ours
Coastguard 352×288	5	41.68	41.84	37.25	41.41	**43.60**
	10	37.93	37.72	35.32	37.22	**42.36**
	15	35.01	34.28	30.16	33.44	**40.92**
Container 352×288	5	38.53	36.65	36.63	38.08	38.33
	10	35.64	35.60	32.18	35.54	35.76
	15	33.37	33.50	29.71	33.07	**35.12**
Hall 352×288	5	38.13	37.82	37.68	37.73	38.64
	10	35.75	35.97	32.29	35.79	**37.16**
	15	33.76	33.45	30.07	33.09	**36.30**
MaD900 352×288	5	39.13	39.80	36.91	39.56	**40.44**
	10	36.13	36.65	31.92	36.17	39.60
	15	33.78	33.77	30.49	33.27	**38.97**
Soccer 352×288	5	38.50	38.31	37.60	37.91	38.43
	10	35.20	35.84	35.14	35.57	**38.07**
	15	33.11	33.20	30.84	33.00	**37.16**
Average		36.37	36.29	33.61	36.05	***38.72***

TABLE III COMPARISON RESULTS of CR

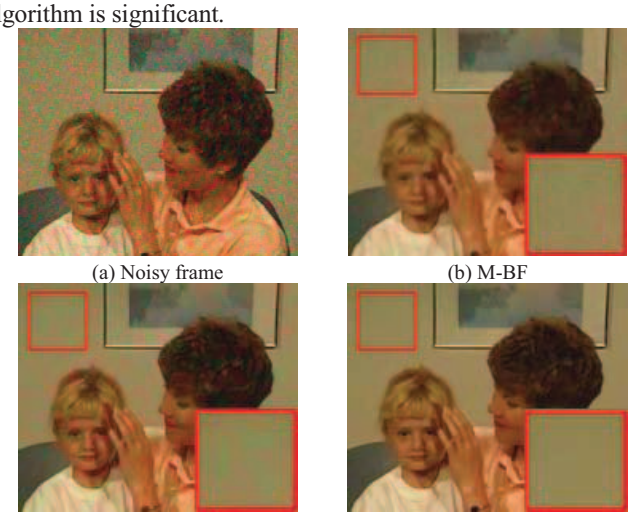
Frame name	σ_n	PSNR				
		M-BF	VBM4D	St-BF	Su-NLM	Ours
Coastguard 352×288	5	42.57	42.62	37.58	41.86	**43.49**
	10	38.15	38.02	35.40	37.12	**42.30**
	15	35.25	34.48	30.01	33.79	**40.41**
Container 352×288	5	37.83	36.22	36.15	36.81	37.06
	10	34.99	34.41	32.11	34.73	34.92
	15	32.79	33.01	30.17	32.25	**34.56**
Hall 352×288	5	40.01	40.59	38.36	38.75	**41.18**
	10	37.05	37.21	32.52	36.53	**38.23**
	15	34.27	33.89	30.27	33.55	**38.83**
MaD900 352×288	5	39.25	39.89	37.71	39.54	39.11
	10	36.47	36.57	31.62	36.37	**38.67**
	15	33.89	33.76	30.50	33.23	**37.86**
Soccer 352×288	5	39.82	39.83	38.15	39.46	39.58
	10	36.37	37.34	36.15	36.25	**38.89**
	15	34.21	33.66	30.95	33.47	**37.92**
Average		36.86	36.76	33.84	36.24	***38.86***

In the first experiment, we compare the proposed algorithm with these state-of-the-art algorithms on synthetic noisy videos.

[1] http://trace.eas.asu.edu/yuv/index.html

To do a quantitative comparison, we simulate chroma noise for standard test videos by adding white Gaussian noise as section 3. When the luma noise is set to three versions with standard deviations 5, 10 and 15, average PSNR of Cb and Cr are included in Table I and Table II.

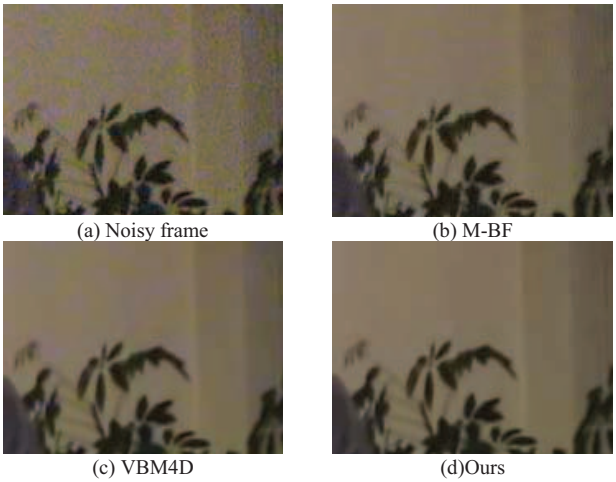
It is obvious that our algorithm has the highest average PSNR. Especially, when the chroma noise is serious, the proposed algorithm can outperform other four algorithms. Moreover, our algorithm can achieve stable results with different levels of chroma noise, whereas highest results of the other four algorithms have violent fluctuations. Fig.5 demonstrates the visual effects of the top three algorithms and the advantage of the proposed algorithm is significant.



(a) Noisy frame (b) M-BF

(c) VBM4D (d)Ours

Fig. 5 Comparison results on MaD900 (Standard deviations of noise on *Y, Cb, Cr* are 10, 5 and 15 respectively).



(a) Noisy frame (b) M-BF

(c) VBM4D (d)Ours

Fig. 6 Denoised results using various algorithms.

It is well known that the white Gaussian noise assumption is not always accurate for real noisy videos. As a result, experiments with real videos are necessary to evaluate the performance of denoising algorithms. In the second experiment, we test denoising algorithms on the video captured by the Sony DV. As shown in Fig.6, the proposed algorithm is also producing a good visual result with the real video. Although there is not quantitative comparison, it is obvious that the proposed algorithm eliminates the coarse-grained chroma noise more effectively.

V. Conclusions

In this paper, we focus on the chroma denoising of videos. The primary contribution of our work is to show how to improve denoise performance upon previous techniques. And a multi-scale joint luma-chroma bilateral filter algorithm is proposed. Experimental results demonstrate that videos enhanced by the proposed algorithm are visually pleasing. The proposed idea can deliver valuable information to improve the quality of noisy videos. Moreover, there are other problems in video denoising (e.g. temporal coherence). In the future, more sophisticated work can be performed to further improve the quality.

Acknowledgments

This work was partially supported by the National Science Fund for Distinguished Young Scholars (No.61125206), the National Natural Science Foundation of China (No. 61370121), the National Hi-Tech Research and Development Program (863 Program) of China (No.2014AA015102), and Outstanding Tutors for doctoral dissertations of S&T project in Beijing (No. 20131000602).

References

[1] Q. X. Yang, “Recursive Bilateral Filtering”, European Conference on Computer Vision (ECCV), 399-413, 2012.

[2] C. Liu and W. T. Freeman, “A high-quality video denoising algorithm based on reliable motion estimation”, European Conference on Computer Vision (ECCV), 2010.

[3] M. Zhang and B. K. Gunturk, “Multiresolution bilateral filtering for image denoising” , IEEE Transactions on Image Process (TIP), vol. 17, no. 12, pp. 2324-2333, 2009.

[4] K. Dabov, A. Foi, and K. Egiazarian, “Video denoising by sparse 3D transform-domain collaborative filtering,” Proc. 15th European Signal Processing Conference (EUSIPCO), 2007.

[5] 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 process (TIP), vol. 21, no. 9, pp. 3952-3966, 2012.

[6] G. Varghese and Z. Wang, “Video de-noising based on a spatiotemporal gaussian scale mixture model”, IEEE Transactions on Circuits and Systems for Video Technology (TCSVT), vol.20, no.7, pp.1032-1040, 2010.

[7] L.Caraffa, J. P. Tarel, and P. Charbonnier, “The guided bilateral filter: when the joint/cross bilateral filter becomes robust” IEEE Transactions on Image Processing(TIP) ,vol.24,no.4, pp.1199 – 1208,2015.

[8] K. N. Chaudhury, “Acceleration of the shiftable algorithm for bilateral filtering and nonlocal means”, IEEE Transactions on Image Processing (TIP), vol.22, no.4, pp.1291-1300, 2013.

[9] H. Bhujle and S. Chaudhuri, “Novel speed-up strategies for non-local means denoising with patch and edge patch based dictionaries”, IEEE Transactions on Image Processing (TIP), vol.23, no.1, pp.356-365, 2014.