A Novel Segmentation-based Video Denoising Method with Noise Level Estimation

Shijie Zhang[1], Jing Zhang[1], Zhe Yuan[1], Shuai Fang[2], Yang Cao[1]

[1] University of Science and Technology of China
[2] Hefei University of Technology
{sjz1901,zjwinner,zheyuan}@mail.ustc.edu.cn ,fangshuai@hfut.edu.cn ,forrest@ustc.edu.cn

Abstract. Most of the state of the art video denoising algorithms consider additive noise model, which is often violated in practice. In this paper, two main issues are addressed, namely, segmentation-based block matching and the estimation of noise level. Different from the previous block matching methods, we present an efficient algorithm to perform the block matching in spatially-consistent segmentations of each image frame. To estimate the noise level function (NLF), which describes the noise level as a function of image brightness, we propose a fast bilateral median filter based method. Under the assumption of short-term coherence, this estimation method is consequently extended from single frame to multi-frames. Coupling these two techniques together, we propose a segmentation-based customized BM3D method to remove colored multiplicative noise for videos. Experimental results on benchmark data sets and real videos show that our method significantly outperforms the state of the art in removing the colored multiplicative noise.

1 Introduction

Noise reduction (denoising) is of crucial importance in multimedia applications, as digital images and videos are often contaminated by noise during acquisition, compression, and storage. Accordingly, it is desirable to reduce noise for visual improvement or as a preprocessing step for subsequent processing tasks.

A number of previous video denoising methods are directly extended from image denoising, such as block matching and 3D filtering [1], wavelet shrinkage [2], PDE based methods [3] and non-local means (NLM) [4]. Several methods integrate motion estimation with spatial filtering. For example, a NLM framework integrating with robust optical flow is introduced in [5]. Besides, the idea of sparse coding in a patch dictionary has also been applied on video denoising (e.g. [6, 7]), where the denoised image patches are found by seeking for the sparsest solution in a patch dictionary. In [8], the problem of denoising patch stacks is converted to the problem of recovering a complete low rank matrix from its noisy and incomplete observation.

The performance of a denoising method is highly dependent on how close the real noise fits the noise model assumed by the method. Most of the state of the art video denoising works consider additive noise model. However, this is often violated in practice. According to [9], there are five primary noise sources, fixed pattern noise, dark current noise, shot noise, amplifier noise, and quantization noise. The practical image noise is very likely to be the colored multiplicative noises, which are not independent of the pixel value. Therefore, most current denoising approaches cannot truly effectively estimate and remove the real noise. This fact prevents the noise removal techniques from being practically applied to multimedia applications.

The main contribution of this paper is that we propose a segmentation-based customized BM3D method that can effectively reduce colored multiplicative noise introduced by today's digital cameras. Two main issues are addressed in this paper: segmentation based block matching and the estimation of noise level. Different with the previous block matching methods which search similar blocks in a fixed-size neighborhood, we present an efficient method to perform the block matching in spatial-temporally consistent segmentations for each image frame. The estimation of the noise level function (NLF), which describes the noise level as a function of image brightness, is the key to ensure the removing of the colored multiplicative noise. Here in our method, a bilateral median filter is exploited to estimate NLF by fitting a lower envelope to the discrete samples measured from image segmentations. To obtain a more reliable NLF, we further extend the estimation method from single frame to multi-frames, under the assumption of short-term coherence of segmentations.

2 Related work

Plenty of video denoising methods have been proposed in the recent few years. As it is beyond this paper to provide a comprehensive and detailed review, we will just focus on reviewing the work closest to ours.

In the past several decades, a variety of denoising methods have been developed in the image area. The significant part of denoising processing is the way to exploit image sparsity [10]. In frequency domain, when a natural image is decomposed into multiscale-oriented subbands [11], we can observe the highly kurtotic marginal distributions [12]. With this case, image sparsity leads to coring algorithms [13] to suppress low-amplitude values while retaining high-amplitude values. In spatial domain, the image sparsity is formulated as image self-similarity, namely patches in an image are similar to one another. NLF methods was proposed by finding the similar patterns to a query patch and take the mean or other statistics to estimate the true pixel value in [4]. Other denoising algorithms, such as PDE [3] and region based denoising [14] also implicitly formulate sparsity in their representation. Sparsity also resides in videos. A number of previous video denoising methods are directly extended from image denoising. Since video sequences usually have very high temporal redundancy, a new frame can be well predicted from previous frames by motion estimation. Therefore, several methods integrate motion estimation with spatial filtering for better performance [5]. Besides, the advances in the sparse representations have achieved outstanding denoising results [6, 7, 15]. The frequency and spatial forms of the image sparsity are introduced by BM3D method [1]. This method produces high-quality results and has been extended for video denoising by exploiting the much more redundancy information in video.

Although many video denoising methods achieve the state of the art results (e.g. [1, 4, 8]), they mainly consider additive noise model which is often violated in practice. As stated in [9], there are five primary noise sources, fixed pattern noise, dark current noise, shot noise, amplifier noise, and quantization noise. The noise model of a CCD camera is proposed in [14] as:

$$I = f(L + n_s + n_c) + n_q \tag{1}$$

where I is the observed image luminance, f is camera response function, n_s denotes all the noise components that are dependent on irradiance L, n_c is the independent noise before gamma correction, and n_q is additional quantization and amplification noise.

Therefore, the practical image noise is very likely to be the colored multiplicative noises, which are not independent of the pixel value. Most existing denoising methods are sensitive to the noise model violation. The existence of other type of noises will severely degrade the denoising performance. This inspires us to develop a robust denoising algorithm capable of removing color multiplicative noise from the given video data. In this paper, we propose a segmentation-based customized BM3D method to remove colored multiplicative noise in real world videos. Firstly, to improve the efficiency of the grouping process in BM3D, we propose a segmentation-based k-nearest neighbors searching approach. Secondly, we propose a fast bilateral median filter based method to estimate the noise level function (NLF). Once the NLF is properly estimated, the noise parameters in those collaborate filtering steps of BM3D method can be adaptively selected.

3 A segmentation-based customized BM3D method for video denoising

3.1 Overview

Figure 1 is the flowchart of our proposed method. This novel method consists of three modules. The first module aims at establishing the spato-temporal consistent segmentation [16, 17]. Based on the segmentation, we can group similar blocks together efficiently by using our proposed k-nearest neighbors searching approach. In addition, we can estimate the NLF through a large number of discrete samples under the assumption of short-term coherence between consistent segmentations. Functionally, this module is designed as an auxiliary part to improve the performance of the classical BM3D algorithm when dealing with the colored multiplicative noise in real world videos. Thus, we have our customized BM3D algorithm shown as the second module. The third module is the output module which is to separate the filtered segmentations and stitch them together to form the final denoising result.

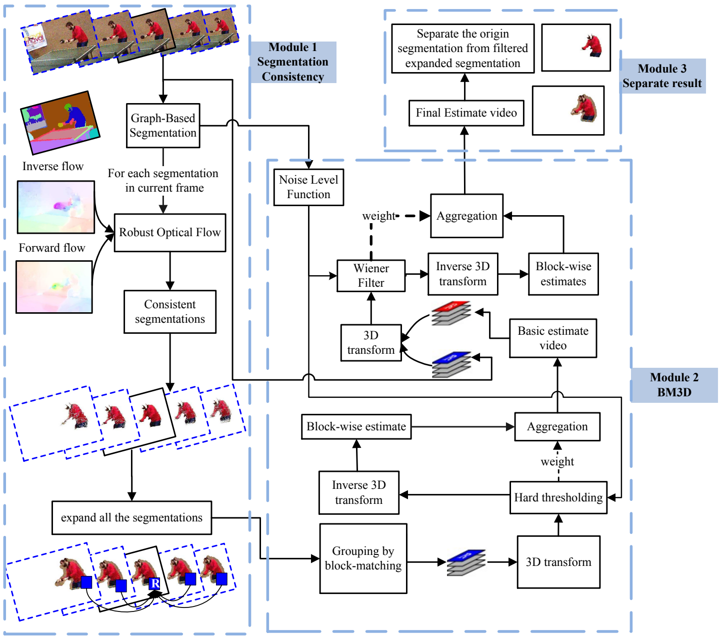

Fig. 1. Flowchart of our proposed segmentation-based customized BM3D method. The operations surrounded by dashed lines are repeated for each segmentation.

3.2 Segmentation consistency

To achieve high quality video denoising result, we should fully utilize of the spatial-temporal consistency of videos to exploiting the redundancy information properly. Firstly we use Graph-based segment method [18] to get piece-wise smooth regions. Then we adopt robust optical flow [19] to find consistent segmentations across adjacent frames.

The segmentation size affects the performance of the following BM3D algorithm a lot. On one hand, the smaller the area is, the more accurate the noise level estimate will be. On the other hand, the bigger the area is, the more similar patches to the reference one we may find. In this paper, we set the maximum size as 0.04 times of image size, and set the minimum size as 0.002 times of image size. In this way, we make a trade-off between the maximum size and the minimum size about segmentation.

We use the robust optical flow proposed by C. Liu et al. [5] to estimate motion. We estimate forward flow $w^b(z) = [v_x, v_y, 1]$ from frame I_t to I_{t+1} and backward flow $w^b(z) = [v_x, v_y, -1]$ from frame I_t to frame I_{t-1} to establish bidirectional temporal correspondence. For each pixel in one segment, we use forward and backward flow to find the corresponding points in adjacent frames. In addition, if a pixel moves out of the image boundary, we just ignore it. In this way, all the points in one segment are mapped from one frame to its neighboring frames, and the temporal consistent segmentation is established.

After finding consistent segmentations in adjacent frames, we expand all the consistent segmentations with a morphological expansion factor. This is to avoid boundary effect in the following block-matching step. In detail, expansion enables the block on the segmentation boundary higher chance to find similar blocks inside consistent segmentations. Figure 2 shows a typical example of the consistent segmentation.

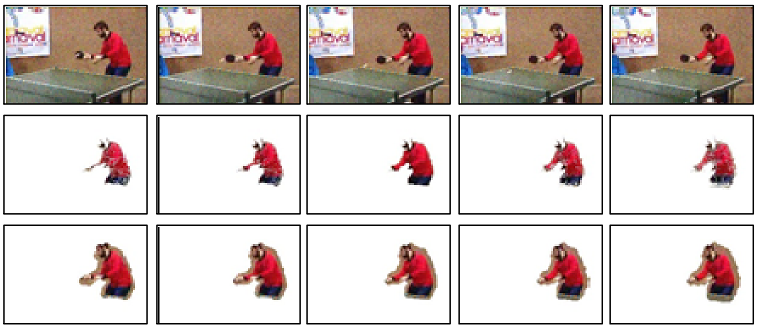

Fig. 2. A typical example of the consistent segmentation. The first row shows several consecutive frames of the tennis sequences with colored multiplicative noise. The second row shows the consistent segmentations established by robust optical flow. The third row shows the consistent segmentations after expansion.

3.3 Segmentation-based block-matching

In the original BM3D algorithm, the searching process is performed in a fixed-size window. In addition, no motion information is exploited to guide the searching process. In contrast, we implement our searching process in the consistent segmentations. This implement has superiority in two aspects: the motion information helps us to improve the searching accuracy in the temporal domain and the segmentation helps us to improve the searching accuracy in the spacial domain. Consequently, we can fully utilize the spatial-temporal redundancy information contained in the videos.

As shown in Fig. 3, for each block centered at pixel z, we want to obtain a set of segmentation-based k-nearest neighbors. For efficiency, we use the priority queue data structure to store the k-nearest neighbors. The priority queue is initialized by the k-nearest blocks to the reference one inside the segmentation in the current frame, which is shown as green boxes in Fig. 3. For each block in the priority queue, we can determine their corresponding blocks in the adjacent frames after we estimate the motion in the video. For example, for block $P\left(z\right)$ in current frame t, we find its corresponding block $P\left(z+\mathrm{w}^{f}\left(z\right)\right)$ in the next frame $t+1$. If $z+\mathrm{w}^{f}\left(z\right)$ is out of segmentation boundary, shown as the blue boxes in Fig. 3, this block $P\left(z+\mathrm{w}^{f}\left(z\right)\right)$ will be discarded. Otherwise, this block

$P\left(z{+}\mathrm{w}^{f}\left(z\right)\right)$ will be inserted into the priority queue and the last one in the queue is discarded. Then we can determine a $N_{FR} \times N_{FR}$ searching neighborhood of $z{+}\mathrm{w}^{f}\left(z\right)$, shown as the pink translucent regions in Fig. 3. And we try to insert more blocks similar to the reference one into the priority queue and update it. Consequently, we use this method to complete the priority queue after searching among those 2H adjacent frames.

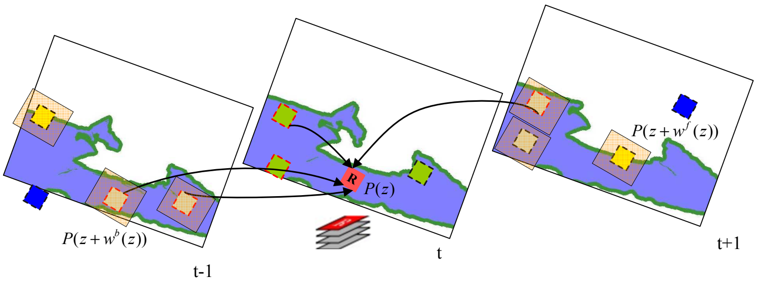

Fig. 3. Illustrations of the supporting patches in spatial-temporal domain for a block $P\left(z\right)$

3.4 Noise level estimation

The noise parameter controls the behavior of the collaborating filtering steps in BM3D. Therefore, it is important to set the noise parameter appropriately. C. Liu et al firstly introduced Noise level function (NLF) in [14] that describes the noise level as a continuous function of image intensity. They also proposed a method to rough estimate the upper bound of the noise based on Bayesian MAP method. However, the estimation method is computational expensive and needs a large number of statistical data. In this paper, we propose an easier but effective method that apply bilateral median filter to estimate an upper bound of NLF. Furthermore, we extend this method to the multi-frame situations.

Firstly, we calculate the means I_i and standard deviation σ_i of each consistent segmentations for all the three color channels (Here, the consistent segmentations are obtained according to the method shown in Section 3.2). Thus, we obtain the discrete point sets $\Omega_k = \{(I_{ik}, \sigma_{ik}), i = 1, 2, ...\}$, where i is the index of segmentations and k represents the color channel (r, g and b). In the next, the point sets Ω_k are plotting in Cartesian coordinates. The problem of estimation of NLF is then transformed into fitting the lower envelope of the point sets Ω_k by a smooth curve that is subject to some constraints. The fitting problem can be formalized as follows:

$$\arg\max_{i} \int_{i} \tau(I_i) - \alpha(\|\nabla\tau(I_i)\|^2) \qquad (2)$$

with the constraints:

$$0 \leq \tau(I_i) \leq \sigma_{\min}(I) \tag{3}$$

$$b_{\min}(I) \leq T(I_i) \leq b_{\max}(I) \tag{4}$$

Here $\tau(I_i)$ is the noise level at index i and the factor α controls the smoothness of the solution. $T(I_i)$ is the first-order derivatives of $\tau(I_i)$. We discretize the range of brightness [0,1] into uniform intervals Θ. The lower bound of the standard deviation is denoted as $\sigma_{\min} = \min_{\Omega}\{\sigma_i\}, i \in \Theta$. Similarly, $b_{\min}(I)$ and $b_{\max}(I)$ represent the lower and upper bound of the first-order derivatives on $\tau(I_i)$, which are calculated by the derivative of σ_i. Since the optimization of (2) being computational expensive, we search for another way to deal with this problem. Here we adopt a bilateral median filter to infer the noise level $\tau(I_i)$ as follows:

$$A(I_i) = median_{S_w}(\sigma_{\min}(I_i)) \tag{5}$$

$$B(I_i) = A(I_i) - median_{S_w}(|\sigma_{min}(I_i) - A(I_i)|) \tag{6}$$

$$C(I_i) = B(I_{i-1}) + \max(\min(B(I_i) - B(I_{i-1}), b_{\max}(I_i)), b_{\min}(I_i)) \tag{7}$$

$$\tau(I_i) = \max(\min(C(I_i), \sigma_{\min}(I_i)), 0) \tag{8}$$

Here S_w is the size of window used in the median filter. Figure 4 shows the example of the estimation of NLF using the above method with S_w=15 × 15.

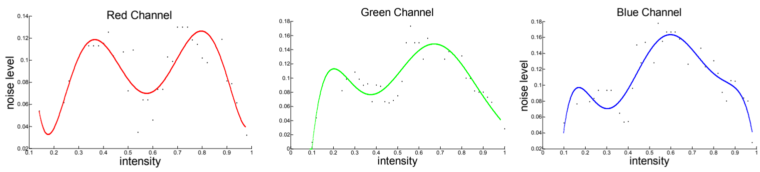

Fig. 4. Figure for noise estimation inside one frame. The red, green, blue curves stand for three channels respectively. X axis stands for intensity and Y axis stands for noise level.

3.5 Segmentation-based customized BM3D algorithm

Coupling the above two techniques together, we propose a segmentation-based customized BM3D method. Figure 1 shows the overview of this method. It can be summarized as: firstly, we obtain consistent segmentations across adjacent frames using the method in Section 3.2. Then, we group the similar blocks by using k-nearest neighbors searching approach and estimate the NLF based on the consistent segmentation results. In the next, for the two collaborate filtering steps in BM3D method, Hard-thresholding filtering and Wiener filtering, the noise parameters are adaptively selected according to the estimated NLF. Finally, we separate the filtered segmentations and stitch them together to form the final denoising result.

4 Experimental results

In this section we present and discuss the experimental results obtained by our method. In our experiments, we use 8×8 patches and 5 adjacent frames. A graph-based image segmentation algorithm is used to partition each image frame into regions. More details of the segmentation algorithm can be found in [18] and the source code is also available online [3]. For segmentation expansion, we use a hexagon morphological operator with the radius of 9 pixels. For the optical flow, we use a robust optical estimation algorithm [19] and the source code is available online [4]. We download BM3D code package from BM3D website [5] for comparison. In this package, the function VBM3D is designed to deal with gray videos, and CVBM3D is designed to deal with color videos. The gray benchmark videos are also downloaded from BM3D website, and the color benchmark videos are downloaded from website [6]. We adopt the synthetic method shown in [14] to add the colored multiplicative noise into the benchmark videos. As shown in (1), there is a camera response function (CRF) f and two noise variances σ_s and σ_c which control the noise level of n_s and n_c in the synthesized video. We download 190 CRFs from the website [7]. To achieve better synthesis results, in our experiment, we select the 60^{th} function to synthesize the colored multiplicative noise.

Firstly, We apply our method on the Foreman sequences with synthetically colored multiplicative noise (with $\sigma_s= 10,\ \sigma_c= 10$) to show its performance on gray video. We use the ground-truth video and VBM3D results (with noise parameters $\sigma= 10\sqrt{2}$) for comparison. The PSNR values of VBM3D method and our proposed method are 27.39dB and 28.32dB, respectively. As shown in the first row of Fig. 5, in terms of visual quality, the VBM3D result is highly affected by the presence of colored multiplicative noise, while our method can achieve better result no matter in the smooth region or the boundary area.

Then, we move on to the color benchmark videos. We select 6 videos and synthesize the colored multiplicative noised videos with different colored multiplicative noise level (with $\sigma_s= 5,\ \sigma_c= 5$ and $\sigma_s= 10,\ \sigma_c= 10$). The contrastive PSNR results of our method with that of CVBM3D on these videos are included in Table 1. We can see that, for most case, our proposed method can achieve better results in term of PSNR. The visual comparison results on color Tennis sequence, Hall-monitor sequence and Flower sequence are shown in the 2^{nd} to 4^{th} row of Fig. 5. It can be seen that some structured noises are still remains in the CVBM3D's results. Our method, however, can smooth out flat regions, as well as keep subtle texture details. In Fig. 5b, our method can smooth out the noise in the background and preserve the textures of table as well. In Fig. 5c and Fig. 5d, though our method performs slightly worse in terms of PSNR, our

[3] `http://www.cs.brown.edu/~pff/segment/`
[4] `http://people.csail.mit.edu/celiu/OpticalFlow/`
[5] `http://www.cs.tut.fi/~foi/GCF-BM3D/`
[6] `http://media.xiph.org/video/derf/`
[7] `http://www.cs.columbia.edu/CAVE`

method completely remove the noise in the flat region and achieve more visually pleasing results.

Table 1. We conduct two groups of experiments on different color videos and compare the PSNR results of our proposed method and the one of the best denoising algorithm: classical CVBM3D (with $\sigma = 5\sqrt{2}$ and $\sigma = 10\sqrt{2}$ respectively for the following two groups experiments).

Method	CVBM3D		Our method	
Parameters	$\sigma_s = 5, \sigma_c = 5$	$\sigma_s = 10, \sigma_c = 10$	$\sigma_s = 5, \sigma_c = 5$	$\sigma_s = 10, \sigma_c = 10$
Foreman	30.53	28.67	30.36	30.43
Newsreport	31.92	29.35	32.46	31.5
Tennis	24.02	23.36	24.21	23.91
Hall-monitor	28.51	27.14	28.49	27.71
Flower	20.82	20.43	20.34	20.15
Mother-daughter	30.71	28.27	32.39	31.74

Finally, we deal with the real video sequences Baby download from website [8], a 720P HD video clip catured by SONY HDR-XR150. One input frame and denoised frame are shown in Fig. 6(a). We compare our method with CVBM3D and C. Liu et al's denoising algorithm. For CVBM3D method, we try three noise parameters $\sigma = 20$, $\sigma = 30$ and $\sigma = 40$. The contrastive denoising results are shown in the Fig. 6(b-f). The close-up view is shown at the lower right corner. We can see that some structured noises are still remains in the CVBM3D's results. C. Liu et al's method and our method both can remove the structured noise and preserve details. Relatively, our method also reduce other types of multiplicative noise and achieve a more visually pleasing result.

5 Conclusion

In this paper, we propose an adaptive segmentation-based video denoising method to remove colored multiplicative noise from video data. We perform block-matching in spatial-temporally consistent segmentations, and integrate 3D filtering with the estimation of noise level function. The effectiveness of our proposed method is validated in various experiments on benchmark datasets and real videos. There is still a limitation of our approach: when the motion is large, the segmentation consistency becomes weak. We will investigate the use of stronger motion estimate method to compensate this case. In addition, we will try to embed our proposed method in other video denoising frameworks for further study.

Acknowledgements This paper is supported by the Fundamental Research Funds for the Central Universities of China. (No. WK2100100009), NSFC

[8] `http://research.microsoft.com/en-us/um/people/celiu/ECCV2010/`

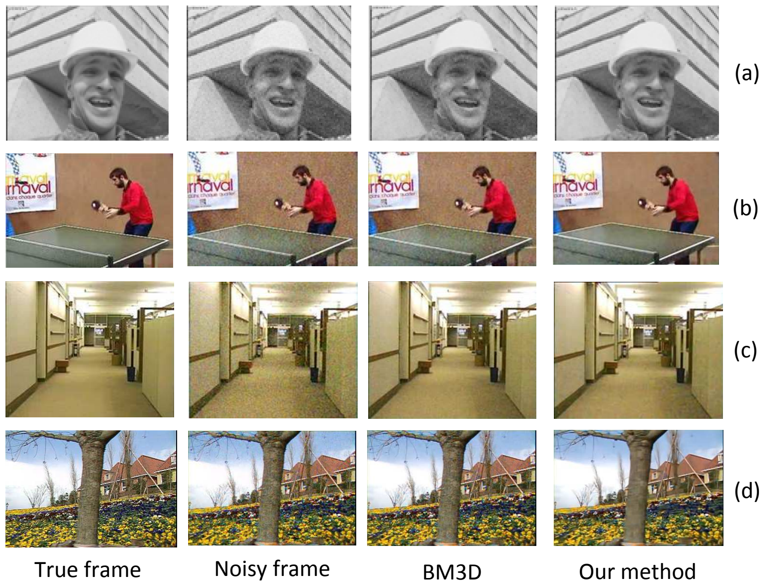

Fig. 5. Comparison results on a series of video sequences. (a) results of Foreman sequence, with σ_s= 10, σ_c= 10 and noise parameters σ= $10\sqrt{2}$ for classical VBM3D. (b) results of Tennis sequence, with σ_s= 10, σ_c= 10 and noise parameters σ= $10\sqrt{2}$ for classical CVBM3D. (c) results of Hall-monitor sequence, with σ_s= 10, σ_c= 10 and noise parameters σ= $10\sqrt{2}$ for classical CVBM3D. (d) results of Flower sequence, with σ_s= 5, σ_c= 5 and noise parameters σ= $5\sqrt{2}$ for classical CVBM3D. Please see color version of this figure in electronic edition.

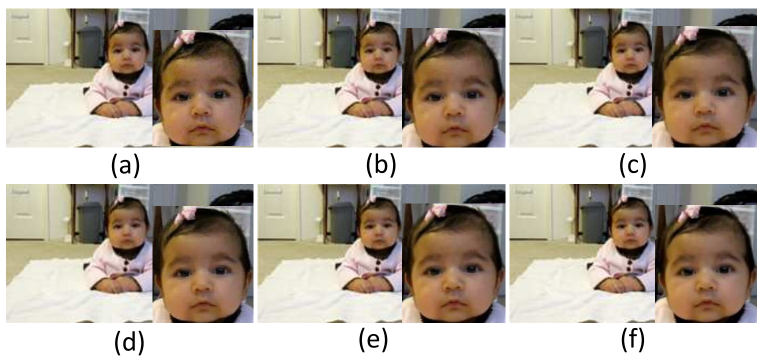

Fig. 6. Comparison results on real video sequences Baby. (a) One frame of real video. (b) CVBM3D with noise parameter σ=20. (c) CVBM3D with noise parameter σ=30. (d) CVBM3D with noise parameter σ=40. (e) Result of C. Liu et al's method. (f) result of our proposed method. Please see color version of this figure in electronic edition.

(No.61175033) and NSFY of Anhui (No.BJ2100100018) and STP (No.11010202192) of Anhui.

References

1. Dabov, K., F.A.e.a.: Image denoising by sparse 3-d transform domain collaborative filtering. IEEE Trans. on Image Processing **16** (2007) 2080–2095
2. Sendur, L., Selesnick, I.W.: Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency. IEEE Trans. on Signal Processing **50** (2002) 2744–2756
3. Tschumperle, D.: Fast anisotropic smoothing of multi-valued images using curvature-preserving pdes. International Journal of Computer Vision (IJCV) **68** (2006) 65–82
4. Buades, A., C.B.M.J.: Nonlocal image and movie denoising. International Journal of Computer Vision (IJCV)) **76** (2008) 123–139
5. Liu, C., Freeman, W.: A high-quality video denoising algorithm based on reliable motion estimation. In: European Conference on Computer Vision (ECCV). (2010)
6. Mairal, J., Bach F., e.: Non-local sparse models for image restoration. In: International Conference on Computer Vision (ICCV). (2009)
7. Protter, M., Elad, M.: Image sequence denoising via sparse and redundant representations. Image Processing, IEEE Transactions on **18** (2009) 27 –35
8. Jiy, H., L.C.S.Z., Xuz, Y.: Robust video denoising using low rank matrix completion. In: Computer Vision and Pattern Recognition (CVPR). (2006)
9. Healey, G., Kondepudy, R.: Radiometric ccd camera calibration and noise estimation. IEEE Trans. on Pattern Analysis and Machine Intelligence **16** (1994) 267–276
10. Yang, J., W.J.H.T.M.Y.: Image super-resolution as sparse representation of raw image patches. In: Computer Vision and Pattern Recognition(CVPR). (2008)
11. Mallat, S.: A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans. on Pattern Analysis and Machine Intelligence, **11** (1989) 674 –693
12. Field, D.: Relations between the statistics of natural images and the response properties of cortical cells. Journal of the Optical Society of America (JOSAA) **4** (1987) 2379–2394
13. Simoncelli, E., Adelson, E.: Noise removal via bayesian wavelet coring. In: IEEE Int. Conf. on Image Processing (ICIP). (1996)
14. Liu, C., S.R.e.a.: Automatic estimation and removal of noise from a single image. IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI) **30** (2008) 299C314
15. Wang, M., H.X.e.a.: Beyond distance measurement: Constructing neighborhood similarity for video annotation. IEEE Trans. on Multimedia **11** (2009) 465–476
16. Zitnick, C., Jojic, N., e.a.: Consistent segmentation for optical flow estimation. In: International Conference on Computer Vision (ICCV). (2005)
17. Klappstein, J., V.T.e.a.: Moving object segmentation using optical flow and depth information. In: Advances in Image and Video Technology. (2008)
18. Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient graph-based image segmentation. International Journal of Computer Vision (IJCV) **59** (2004) 167–181
19. Liu, C.: Beyond pixels: exploring new representations and applications for motion analysis. PhD thesis, Massachusetts Institute of Technology (2009)