Verified Robust Optical Flow Guided Local Block-Matching and 4D Filtering

Xueqing Liu[2], Xinghao Jiang[2], Tanfeng Sun[1, 2], Ajing Xu[3, 4]
[1]State Key Laboratory of Software Engineering, Wuhan University, China
[2]School of Electronic Information and Electrical Engineering Shanghai Jiaotong University, Shanghai, China
[3]School of Medicine, Shanghai Jiaotong University, Shanghai, China
[4]Department of Pharmacy, Xinhua Hospital
Email: {lxqlxq21, xhjiang, tfsun}@sjtu.edu.cn, {jeta xu}@hotmail.com

***Abstract*—Up to now, a lot of researches have been done on video denoising. In this paper, a new scheme of video denoising method, verified robust Optical Flow Guided Local Block-matching and 4D Filtering (OL-BM4D), is proposed. In this scheme, first of all, a novel algorithm is introduced to calculate verified robust optical flow. Then we propose a new method of gathering 3D blocks to form 4D groups. After the 4D groups are formed, two-stage 4D filtering is applied to them. And the video is reconstructed from the groups after denoising. Comparing our method to the cutting edge video denoising method VBM3D, the result shows OL-BM4D provides better denoising results according to both objective evaluation (PSNR) and subjective evaluation (stability between frames, the ability of retaining details and texture).**

***Index Terms*—video denoising, optical flow, local block-matching and 4D filtering, objective evaluation, subjective evaluation**

I. Introduction

Video can be compressed with a very high ratio to save storage space. That means there is a great deal of redundant information in a video. Inspired by video compressing methods, it is known to us that videos have self-similarity both in one frame and between neighbor frames. The redundant information and self-similarity make it possible for us to restore a video of severe noise without losing much information.

When a video is being denoised, the denoised result is acquired by applying the denoising method to a video, which is the combination of an original video without noise and noise added artificially. Then the denoised result is evaluated through two aspects-objective evaluation and subjective evaluation. Objective evaluation methods are achieved by comparing the denoised result with the original video without noise and calculate the error. Subjective evaluation, on the other hand, is given by human beings. The denoised results are played to volunteers, and they will judge the quality of the video. To improve the subjective evaluation, we should take the perception of human beings into consideration.

There has been a great deal of denoising methods for videos and images.

BM3D [1] is a well known image denoising strategy based on an enhanced sparse representation in transform-domain. The enhancement of the sparsity is achieved by grouping similar 2D image fragments into 3D data arrays called “groups”. VBM3D [2] and BM4D [3] are its extension to video denoising and volumetric data denoising respectively. These methods are very popular for their good performance. Still, they overlook the importance of human perception.

One of the very important factors that affect people’s evaluation of restored video is the time-domain stability. It was first taken into consideration in [4]. Before that, it is believed that there is no need to pay attention to the relationship between frames. In [4], a robust optical flow based on a coarse-to-fine algorithm is introduced and used to gain higher inter-frame stability.

Optical flow is a big research area[1]. Different optical flows are of different characteristics [5]. Most optical flows like the widely used Pyramid LK optical flow are sensitive to noise and light variance. This is a good character when being applied to applications like inter-frame forgery detection [6]. But when comes to denoising, robust optical flow is needed. The algorithms based on coarse-to-fine are often robust to noise.

In [4], although the author successfully beat VBM3D in subjective evaluation, in objective evaluation VBM3D is never really beaten. Motion estimation is used both in VBM3D and VBM4D [7]. However, being a motion estimation method of good performance optical flow is decided by most researchers not a good choice for motion estimation. They believe it can’t improve objective evaluation in video denoising. Meanwhile, the reason is never explained.

BM4D is intuitively good at gaining high inter-frame stability, but it can’t work well on video of high speed objects. In [8], the author introduced a segmentation based method based on VBM3D, optical flow came into

Manuscript received September 16, 2014; revised December 30, 2014.

[1] http://vision.middlebury.edu/flow/

doi: 10.12720/ijsps.4.2.155-161

his sight. But in [8] the optical flow is only used as an assist to line up patches between frames.

In this paper, we present a novel method called verified robust optical flow guided local block-matching and 4D filtering. Our main contributions include:

1) Design a verification algorithm for robust optical flow. This verified optical flow can be put into use in video denoising. It gains good performance on both objective evaluation and subjective evaluation (the origin robust optical flow can only show good performance on subjective evaluation).
2) Propose a novel video-denoising system. It improves BM4D's performance on videos of high speed objects. The results show its performance is better than the cutting edge video denoising method VBM3D.

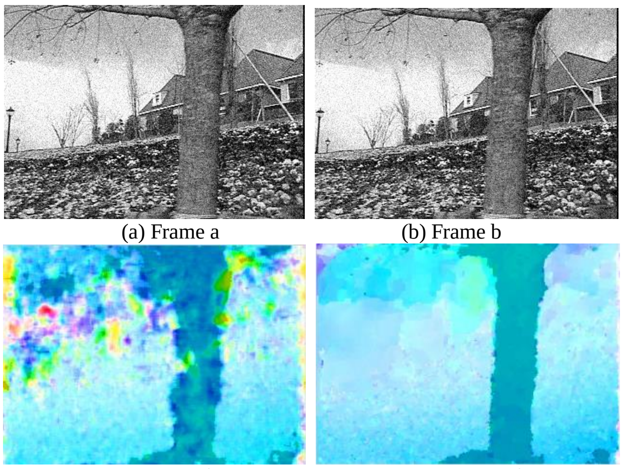
The paper is organized as follows: Section II briefly reviews robust optical flow and the BM4D approach. Section III proposes the novel video denoising system we design and the alter we make on robust optical flow to improve the denoising performance. Experiment setup and results are provided in Section IV, and finally Section V includes the concluding remarks.

II. Robust Optical Flow and BM4D

A. Robust Optical Flow and BM4D

There have been a lot of studies on optical flow. Before coarse-to-fine [4] method is introduced into optical flow, most optical flows are very sensitive to noise (Fig. 1(c)). The coarse-to-fine method provides us optical flows robust to noise (Fig. 1(d)).

In [4], the author used robust optical flow to achieve denoising task. The result shows good performance on subjective evaluations, but not objective ones. That isn't right to instinct, since objective evaluation should be consistent to subjective evaluation. But the reason for this is never revealed. In Section III, we will explain the reason and propose a novel algorithm to calculate the verified robust optical flow.



(a) Frame a (b) Frame b

(c) L-K optical flow (d) Robust optical flow

Figure 1. (a) and (b) are two frames from a noisy video, (c) shows the L-K optical flow (a widely used optical flow) between the two frames which is very sensitive to noise, (d) shows the robust optical flow which is affected little by noise.

B. Local Block-Matching and 4D Filtering (BM4D)

BM4D is a denoising method for volumetric data, it is first introduced in [3]. It is implemented in two cascading stages, namely a hard-thresholding and a Wiener-filtering stage.

1) Hard-Thresholding stage: In this stage, cubes taken from the volumetric data are clustered with a distance measured via the photometric distance

$$d(C_{x_i}^{t_i}, C_{x_j}^{t_j}) = \frac{\left\| C_{x_i}^{t_i} - C_{x_j}^{t_j} \right\|_2^2}{L^3} \tag{1}$$

The cubes that are similar to one cube $C_{x_0}^{z_0}$ are put together to build a 4-D group $G_z(x_0, t_0)$. Collaborative filtering is realized by hard thresholding in 4-D transform domain it:

$$\hat{G}_y^{ht}(x,t) = \tau_{4D}^{ht^{-1}}(\gamma^{ht}(G_z(x_0,t_0))), (x,t) \in X \times T \tag{2}$$

where $\tau_{4D}^{ht^{-1}}$ is the 4-D transform and γ^{ht} is the hard-threshold operator. The outcome of hard-thresholding stage, is obtained by aggregation of all the estimated groups.

2) Wiener-Filtering stage: In the second stage, the groups are formed in two ways: some are formed by the cubes extracted from the outcome volumetric data from stage one, the others are the groups after the hard-thresholding from stage one.

In this stage, the shrinkage coefficients are calculated as

$$W(x_0,t_0) = \frac{\left|\tau_{4D}^{wie}(G_{\hat{y}^{ht}}(x_0,y_0))\right|^2}{\left|\tau_{4D}^{wie}(G_{\hat{y}^{ht}}(x_0,y_0))\right|^2 + \sigma^2} \tag{3}$$

Winer filtering is realized by

$$\hat{G}_y^{wie}(x_0,t_0) = \tau_{4D}^{wie^{-1}}(W(x_0,t_0) \cdot \tau_{4D}^{wie}(G_z(x_0,t_0))) \tag{4}$$

This time we reconstruct the volumetric data with the weights

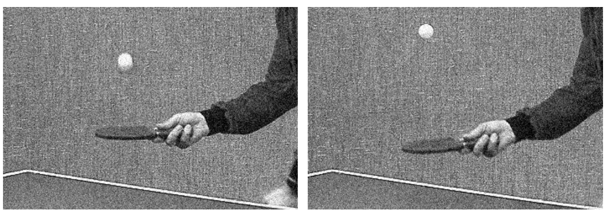
$$\omega_{(x_0,t_0)}^{wie} = \left\| W(x_0,t_0) \right\|_2^{-2} \tag{5}$$

III. Methods

A. Verified Robust Optical Flow

In this part, the novel algorithm to calculate the verified robust optical flow is proposed.

Among the robust optical flows, one is called CGL-TV. It is able to compute larger displacements in reasonable time and is of better accuracy than the optical flow used in [4].

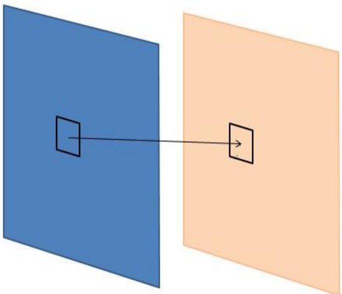


(a) Frame a (b) Frame b

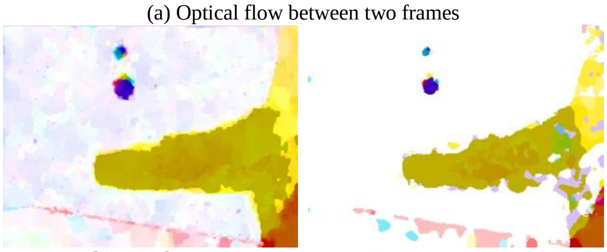
(c) L-K optical flow (d) Robust optical flow

Figure 2. (a) and (b) are two frames from a noisy video, in this video, the background is steady, (c) is the robust optical flow between them, the background should be white as there is no movement, but the movement of the foreground caused the error in the background, especially the area under the arm, we call this “leak” (d) is the verified robust optical flow, in which the “leak” is suppressed

In Fig. 2(a) and Fig. 2(b), the background is steady. As robust optical flow is a motion estimation method, the CGL-TV optical flow should be 0. But the outcome provided by CGL-TV optical flow (Fig. 2(c)) shows different result. This error causes optical flow to fail on objective evaluation. To gain good objective evaluation while denoising, the optical flow needs to be verified to make it more precise.



(a) Optical flow between two frames



(b) Search process (c) Voting process

Figure 3. Verification method: (a) shows origin robust optical flow without verification between two frames. In (b), around the patch in frame b the optical flow points to, a search window is formed. In this window we search for a patch which is the most similar to the patch in frame a. As patches overlap with each other, in (c) all the patches with a particular pixel vote for the accurate optical flow of this pixel to reduce noise

The verification method designed by us is shown in Fig. 3. The original CGL-TV optical flow calculated is shown in Fig. 3(a). The verification method is designed as follows: as error can be introduced with interpolation, the optical flow is first rounded to integers to avoid interpolation. And as optical flow can be not accurate enough, a relocating process in frame B is introduced. For every patch in A, search for a more reasonable patch in B. Suppose for the original CGL-TV optical flow, the patch in A is connected by the optical flow with a patch $P(l_j, t_j)$ in B. The dearching window is around the neighborhood of $P(l_j, t_j)$. The method is shown in Fig. 3(b). As in frame A, the patches are overlapped with each other, the optical flow for every single one of the pixels is decided by a voting process. All the patches that contains this pixel vote (Fig. 3(c)) for its optical flow. The optical flow with most votes wins. An optical flow outcome more robust to noise is ensured by this voting method. The optical flow after modification is shown in Fig. 2(d).

In the search process shown in Fig. 3(b), the similarity between a patch $P(l_i, t_i)$ in frame A and a patch $P(l_j, t_j)$ in frame B is defined through a penalized quadratic difference

$$D(P(l_i,t_i),P(l_j,t_j)) = \frac{\left\|P(l_i,t_i)-P(l_j,t_j)\right\|_2^2}{N^2} + \gamma_d \left\|distance(l_i,l_j)\right\|_2 \tag{6}$$

In (6), $l_i = (x_i,\ y_i)$ is the position the origin robust optical points to in frame B, $l_j = (x_j,\ y_j)$ is the position of one patch from frame B. γ_d is defined as follows.

$$\gamma_d(\sigma) = 0.0005\sigma^2 - 0.0059\sigma + 0.0400 \tag{7}$$

The best γ_d chosen for different deviation of the noise σ is shown in (7). This parameter fitting research is according to [7].

$distance(l_i, l_j)$ is defined according to the character of the robust optical flow. Since robust optical flow is based on coarse-to-fine method, two high speed areas of different speed have little effects on each other, but the steady areas are easily affected by the high speed area. That means the movement of the high speed areas can “leak” into the steady areas and cause errors in CGL-TV optical flow (see Fig. 2(c)). To eliminate this effect, $distance(l_i, l_j)$ is defined to be different for video of large areas of steady background and video of moving background.

Define $distance(l_i, l_j) = (dis1D(x_i, x_j), dis1D(y_i, y_j))$ and

$$dis1D(a,b) = \begin{cases} |b-a| & moving\ background \\ \left\| \left|b-\frac{a}{2}\right| - \frac{a}{2} \right| & steady\ background \end{cases} \tag{8}$$

Fig. 4 shows that for video of moving background, the 1-D distance descriptor is a single valley function. While for the video of steady background, it is a double valley function with one valley at 0 and another at *a* (as shown in Fig. 4). That means the patches without movement or with the movement estimated by optical flow are of smaller distance. This solves the leak problem (as shown in Fig. 2(d)).

This verification process is shown in the blue box in Fig. 4.

B. Framework of Video Denoising Model

Fig. 4 shows the new scheme of the video denoising method we designed. Take a single frame i, and n frames before and after this one, there are 2n+1 frames in all.

The frame needed to be denoised is in the middle, and there are n frames provides information for denoising on each side. Then for every one of the 2n frames, the robust optical flow to frame i is calculated. Take frame i+1 as an example, an initial optical flow from i to i+1 is now provided. Then by rounding the optical flow, 0 optical flow area percentage is then counted. If the percentage is larger than r, the frame is classified as of steady background; otherwise, it is classified as of moving background. Then for video of steady background or video of moving background, modification of the optical flow is applied on the frame according to the classification. And "voting" process is taken to get rid of the effects of noise and false matching afterwards. Up to this step, the verification of optical flow is done. Then frame i+1 is shifted according to the modified optical flow, and stuffed back to its position in the 2n+1 frame series. By doing these to every single frame of the 2n neighbors, a 3-D data cuboid is gained. Then Block-matching and 4D Filtering is applied to the cuboid as shown in Fig. 4. This provides the denoised result of frame i. By doing this to all the frames in the video, the denoised result of the whole video is completed.

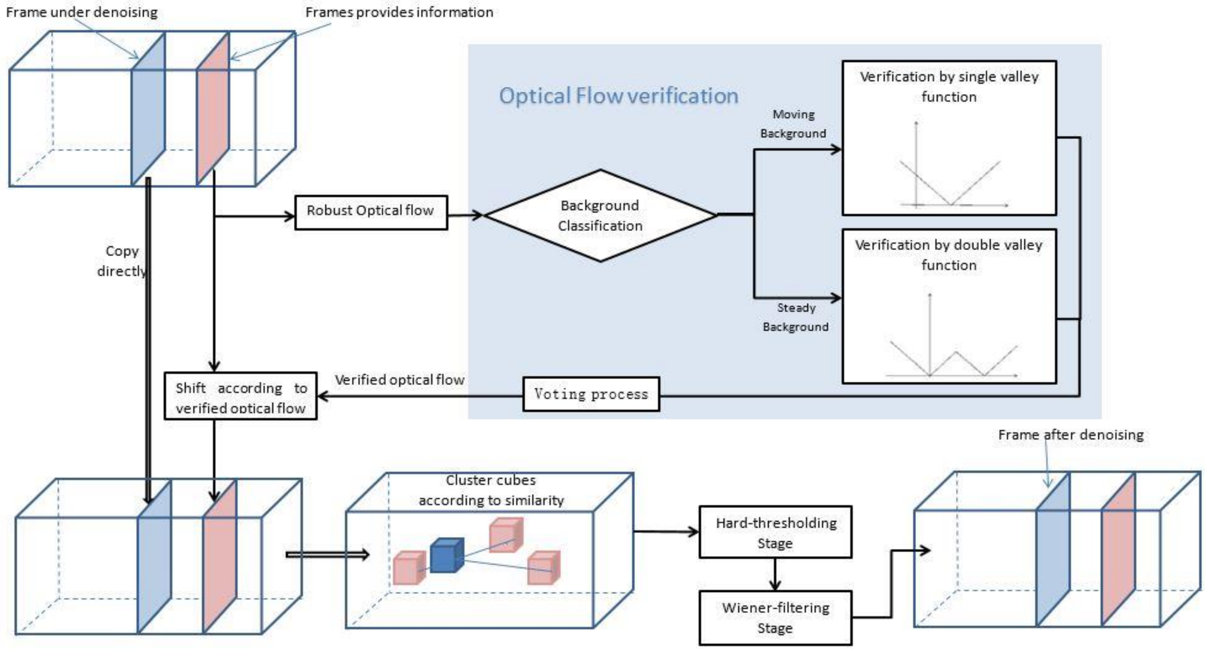


Figure 4. Framework of the denoising system

IV. Experiment and Results

A. Parameter Selection

Here the neighbors on each side of the frame to be denoised is defined to be n=8. Patch size is 7*7, the percentage of 0 optical flow area, that distinguishes steady background video and moving background video, is 20 percent. Searching window for a similar patch in the modification process is 7*7. It leads to the voters in the voting process also be 7*7 for each pixel.

B. Results

In this section we present the experimental results obtained with a MATLAB implementation of the OL-BM4D algorithm, and we compare it against V-BM3D, as it represents the state of the art in video denoising.

1) Objective evaluation: The denoising performance is measured using the PSNR as a global objective measure for the whole processed video:

$$PSNR = -10log_{10}(255^{-2}|X||T| \times \sum_{(x,t)\in X\times T}(y(x,t)-\hat{y}(x,t))^{-2}) \tag{9}$$

The objective evaluation is shown in Table I, the bold numbers are of better performance. Our method can beat VBM3D in most videos, except for some short ones. For example, "salesman" is a video of very little movement and also very short. Our method can work better than VBM3D on this video from frame 9 to frame 42, but lost on the first and last several frames a little.

TABLE I. Objective Evaluation of Denoising Results

σ	Video:	Salesm.	Tennis	Fl.Gard.	MissAm.
	Res.:	288*352	240*352	240*352	288*360
	Frames:	50	150	150	150
15	OL-BM4D	35.48	**32.67**	**30.26**	**38.98**
	VBM3D	**35.51**	32.61	29.91	38.43
20	OL-BM4D	34.10	**31.33**	**28.53**	**38.13**
	VBM3D	**34.11**	31.17	28.26	37.74
σ	Video:	Coastg.	Foreman	Bus	Bicycle
	Res.:	144*176	288.352	288*352	567*720
	Frames:	300	300	150	30
15	OL-BM4D	**33.90**	**35.64**	**32.43**	37.11
	VBM3D	32.95	34.64	31.05	**37.15**
20	OL-BM4D	**32.32**	**34.10**	**30.80**	**35.78**
	VBM3D	31.69	33.32	29.54	35.72

The bold numbers in the table are of better performance.

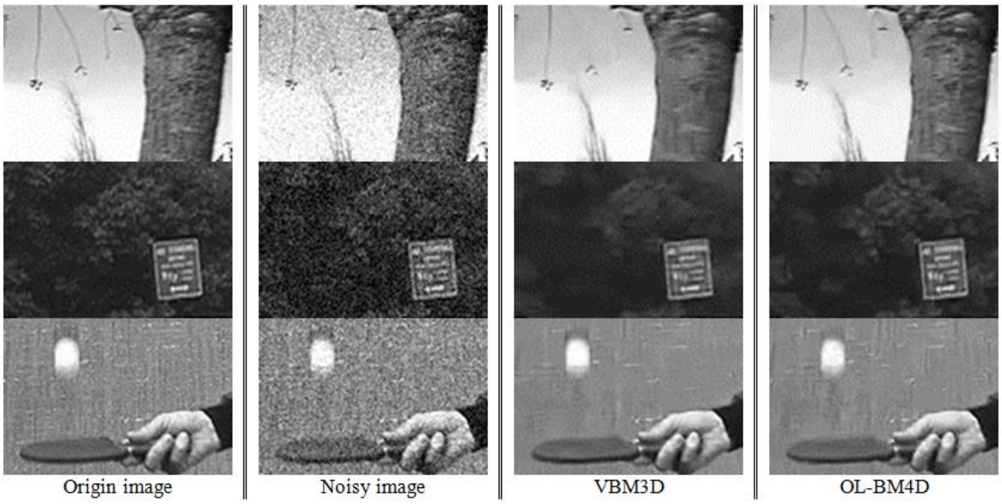
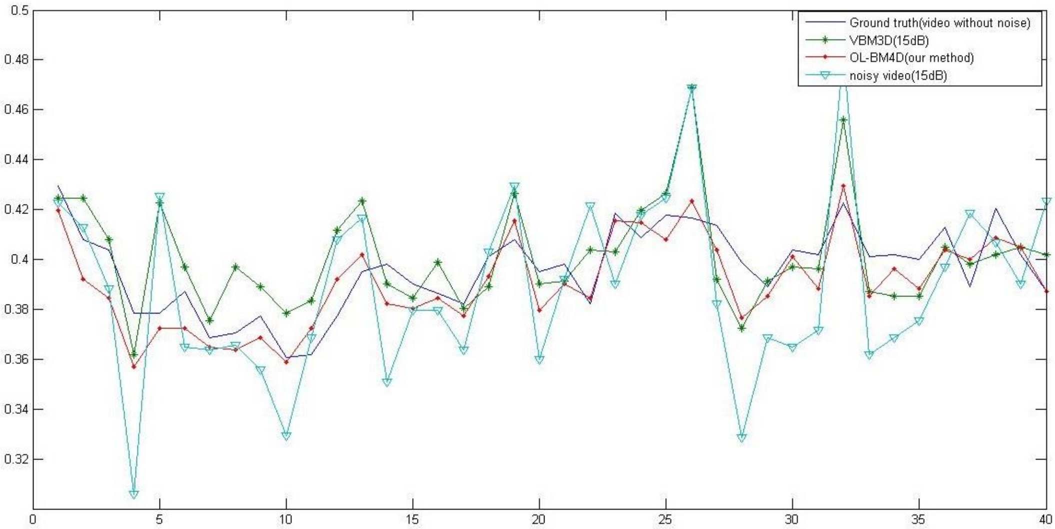
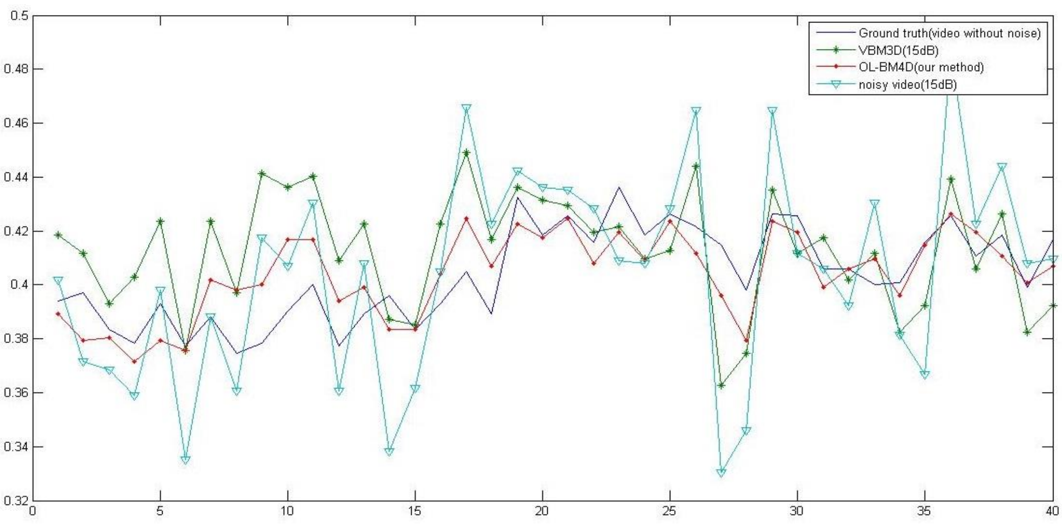


Figure 5. Visual comparison of the sequences, from top to bottom, flower garden, bus and tennis corrupted by white Gaussian noise with standard deviation $\sigma = 20$, denoised by the proposed algorithm OL-BM4D and the VBM3D algorithm. Our method clearly provides better result and recovers more detail



(a)



(b)

Figure 6. Stability between frames: the standard deviations of pixel intensities in (a) are $\sigma_{Groundtruth}$=0.0168, σ_{VBM3D}=0.0212, $\sigma_{OL\text{-}BM4D}$=0.0180 and $\sigma_{noisyvideo}$=0.0358 respectively; the standard deviations of pixel intensities in (b) are $\sigma_{Groundtruth}$=0.0173, σ_{VBM3D}=0.0215, $\sigma_{OL\text{-}BM4D}$=0.0157 and $\sigma_{noisyvideo}$=0.0381 respectively

2) Subjective evaluation: The denoising result is also shown in Fig. 5. Judge by human eye, the result shows our method can recover more detail and texture than the VBM3D.

Another important subjective evaluation criterion is the inter-frame stability. The improvement on inter-frame stability can be observed on the denoised video. Here, to show the performance of our method on stability between frames, an evaluation method from [4] is put into use. Here in Fig. 6, pixel intensities along motion paths over frames are used for demonstrating the stability between frames. Two motion paths are shown here. In both figures, judged by human eyes our method shows better stability between frames. Also, our method gain standard deviations of pixel intensity closer to the origin ones, much lower than VBM3D. This means our denoised video is more s between frames.

C. Conclusion

In this paper, we analysed and resolved the cause of normal optical flow's incapable of gaining good objective evaluation when denoising a video, propose a novel algorithm to calculate the verified robust optical flow. And introduced a novel denoising method for video. With this method, better result is gained than the cutting edge method VBM3D. The evaluation method for this result comes from both objective evaluation (PSNR) and subjective evaluation (stability between frames, the ability of retaining details and texture). Verifying optical flow properly to get accurate motion estimation is one of the keys in video denoising. And BM4D's nature character of providing inter-frame stability helps improving subjective evaluation performance. In the future, more work can be done on improving the denoise algorithm's performance on the frames on both ends of the video.

ACKNOWLEDGMENT

The work of this paper is sponsored by the National Natural Science Foundation of China (No.61272439, 61272249), the Specialized Research Fund for the Doctoral Program of Higher Education (No.20120073110053), and the Fund of State Key Laboratory of Software Engineering, Wuhan University (No.SKLSE2012-09-12).

This project is also partially supported by Cross-fund of Biomedical Engineering of Shanghai Jiaotong University (No.YG2012MS57), by a grant from the Science and Technology Commission of Shanghai Municipality (No.13ZR1426900) and the Innovation program of Shanghai Municipal Education Commission (No.12ZZ117).

REFERENCES

[1] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising with block-matching and 3D filtering," in *Proc. SPIE*, 2006.

[2] K. Dabov, A. Foi, and K. Egiazarian, "Video denoising by sparse 3D transform-domain collaborative filtering," in *Proc. 15th European Signal Processing Conference*, 2007.

[3] M. Maggioni, V. Katkovnik, K. Egiazarian, and A. Foi, "Nonlocal transform-domain filter for volumetric data denoising and reconstruction," *IEEE Transactions on Image Processing*, vol. 22, no. 1, pp. 119-133, 2013.

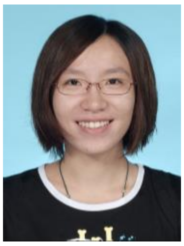
[4] C. Liu and W. T. Freeman, "A high-quality video denoising algorithm based on reliable motion estimation," in *Computer Vision–ECCV 2010*, Springer Berlin Heidelberg, 2010, pp. 706-719.

[5] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. J. Black, and R. Szeliski, "A database and evaluation methodology for optical flow," *International Journal of Computer Vision*, vol. 92, no. 1, pp. 1-31, 2011.

[6] J. Chao, X. Jiang, and T. Sun, "A novel video inter-frame forgery model detection scheme based on optical flow consistency," in *Digital Forensics and Watermaking*, Springer Berlin Heidelberg, 2013, pp. 267-281.

[7] M. Maggioni, G. Boracchi, A. Foi, and K. Egiazarian, "Video denoising using separable 4D nonlocal spatiotemporal transforms," in *Proc. IS&T/SPIE Electronic Imaging*, 2011.

[8] S. Zhang, J. Zhang, Z. Yuan, S. Fang, and Y. Cao, "A novel segmentation-based video denoising method with noise level estimation," in *Advances in Multimedia Modeling*, Springer Berlin Heidelberg, 2013, pp. 272-282.



Xueqing Liu received her B.S. degree in Information and Communication Engineering from Zhejiang University, Hangzhou, P.R. China in 2012.
She is now a dual-master student of Georgia Institute of Technology, Georgia, USA and Shanghai Jiao Tong University, Shanghai, China.
Her current research interests include computer vision, image and video processing and medical image processing.

Xinghao Jiang received the PhD. Degree in Electronic Science and Technology from Zhejiang University, Hangzhou, P.R. China in 2003.
He is a professor at the School of Information Security Engineering at Shanghai Jiao Tong University, Shanghai, P.R. China.
His current research interests include multimedia security and image retrieval, intelligent information processing, cyber information security, information hiding and watermarking.
Dr. Jiang is an IEEE member.

Tanfeng Sun was born in Jilin Province of P.R. China at June 9th 1975. He had his PH.D degree on information and communication system from Jilin University in Jilin Province of China at the end of 2003.
He worked at Information and Communication Specialty as a Post-doctor in Shanghai Jiao Tong University from Nov. 2003 to Nov. 2005. He works at School of Electronic Information and Electrical Engineering in Shanghai Jiao Tong University in China as an Assistance Professor from 2005 to 2012. He promoted the title of associate professor in Shanghai Jiao Tong University at Dec. 2012 and he still works as a faculty of SJTU now. He had been working at Department of Electrical and Computer Engineering in New Jersey Institute of Technology as a Visiting Scholar from July 2012 to Dec. 2013.
His research interests are mainly on Digital Forensics on Video Forgery, Digital Image and Video Watermarking, and Video's content recognition and understanding.
Professor Sun is an IEEE Member from 2011 to 2014. He had been invited by IEEE North Jersey Section Computer Society Seminar to give a talk on Enormous Challenge to Image and Video Tampering Detection in Fairleigh Dickinson University at Nov. 14, 2013.

Ajing Xu received the M.D. Degree from Tohoku University, Sendai, Japan in 2008.
She now works as a Staff Researcher/ clinic pharmacist at Department of Pharmacology, Xinhua Hospital and School of Medicine, Shanghai Jiaotong University, Shanghai.
Her current research interests include clinic pharmacy and Individualized medication, Pharmacological studies by using live-animal model and study of MAP kinase (MAPK) and phosphotase signaling pathways in the innate immune system; role of Nurr1 gene in the pathogenesis and treatment in Parkinson's disease.