ILLUMINATION-INVARIANT NON-LOCAL MEANS BASED VIDEO DENOISING

Jie Ren[1], *Yue Zhuo*[2], *Jiaying Liu*[1†], *Zongming Guo*[1]

[1]Institute of Computer Science and Technology, Peking University, Beijing 100871, China
[2]College of Information Science and Technology, Beijing Normal Universiy, Beijing 100875, China

ABSTRACT

In this paper, we present a robust illumination-invariant non-local means (NLM) based video denoising algorithm with special illumination handling. Illumination variances pose a challenge to the NLM-based denoising algorithms. To address this issue, we first propose several possible technical improvements, and verify their efficacy of eliminating the influence of illumination changes. Then, by analyzing and comparing these techniques, a histogram processing based technique is integrated into the non-local means denoising framework. Experimental results on synthesis and real video denoising show that the proposed method is able to fully explore the non-local self-similarity property in natural videos under variable illumination conditions.

***Index Terms*—** Video denoising, Non-local means, Illumination-invariant, Histogram processing

1. INTRODUCTION

Image and video denoising is a long-standing research area in the image processing community. It aims to recover the high-quality clean image (sequence) from its noised version, which may be taken, for example, by a low-end imaging devices and/or under limited conditions.

A key of image and video denoising is to exploit the prior information. For video denoising, which is the main focus of this paper, the temporal coherence of image sequence is an important ingredient of designing an efficient video denoising algorithm. To explore the temporal information in video denoising, motion compensated image sequence filters are proposed [1]. However, as claimed in [2], motion estimation is a difficult problem mainly because the aperture problem on textureless regions. Filtering along the inaccurate trajectories can lead to blur and information loss.

Instead of relying on a robust and reliable motion estimation, the Non-Local Means (NLM) filtering was proposed by Buades *et al.* [2], which is originally established for image denoising [3]. The idea of NLM is very simple: the patches that have similar structure patterns can be spatially far from each other, and thus one can collect them in the whole image or sequence. This simple idea motivates many follow-up research works, *e.g.*, towards the computational aspect [4, 5], rotation-invariant [6]. For structured noise, Liu and Freeman [7] recently proved that the combination of robust optical flow estimation with the NLM framework will further boost the video denoising performance.

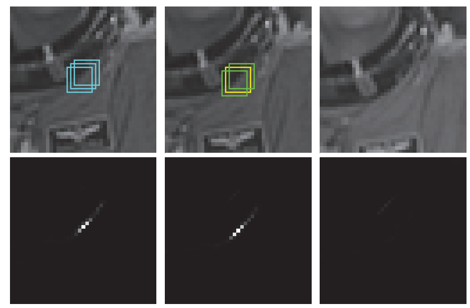


Fig. 1. Non-local patch searching in a space-temporal volume consisting of search windows located in three adjacent frames, in which the rightmost frame has an illumination change due to the flashlight. The color boxes are similar patches to the current patch (yellow box) found by the non-local search. The second row shows the weight distribution of each temporal slice of the 3D space-temporal volume.

In the original NLM video denoising framework and its variants, it is implicitly assumed that the "intrinsic" similar local structures in the image sequence have the coherent illumination condition. Therefore, intensities of image patch will not have significant change and can be directly utilized as the feature vector of local structure. However, it often occurs that the video scene has illumination changes during the capture process, *e.g.*, the flashlight effect. As shown in Fig.1, when the neighboring frame has illumination condition changes, the patch matching may miss the "intrinsic" structural similar patches although they are very visually similar.

Therefore, we argue that NLM-based video denoising, especially when illumination changes are taken into account, indeed needs special handling to make it more robust. We use the NLM as the backbone of our system and propose

†Corresponding author
This work was supported by National Natural Science Foundation of China under contract No.61101078, National Key Technology R&D Program of China under Grant 2012BAH18B03 and Doctoral Fund of Ministry of Education of China under contract No.20110001120117.

several possible techniques to address the illumination invariance issue. Finally, by further analyzing and comparing these techniques, the histogram processing based technique is integrated into the NLM framework due to its robustness to the noise, scene changes and illumination changes.

The rest of this paper is organized as follows. After introducing the NLM-based denoising framework in Sec. 2, we propose and compare three possible techniques, and extend the NLM-based denoising method into an illumination-invariant one. Experimental results are presented in Sec. 3 to verify the video denoising performance of the proposed method. Finally, Sec. 4 concludes this paper.

2. ILLUMINATION-INVARIANT NLM DENOISING

2.1. NLM-based Denoising Framework

For the purpose of clarity, we briefly review the the NLM filtering proposed by Buades *et al.* [3]. The NLM filtering is formularized as follow:

$$NL[v(i)] = \sum_{j \in \Omega} \omega(i,j) v(j), \tag{1}$$

where $v(i)$ is the noisy pixel value, and $NL[v(i)]$ is the NLM output. Ω is a window for searching pixels similar to pixel i. $\omega(i,j)$ depends on the weighted Euclidean distance $\|v(\mathcal{N}_i) - v(\mathcal{N}_j)\|^2_{2,a}$ of patch $\mathcal{N}_i$ and $\mathcal{N}_j$ centered at pixel i and j,

$$\omega(i,j) = \frac{1}{Z(i)} \exp\left(-\frac{\|v(\mathcal{N}_i) - v(\mathcal{N}_j)\|^2_{2,a}}{h^2}\right), \tag{2}$$

where h is a filtering parameter, $Z(i) = \sum_j e^{-\frac{\|v(\mathcal{N}_i) - v(\mathcal{N}_j)\|^2_{2,a}}{h^2}}$ is the normalizing constant and a is the standard deviation of Gaussian kernel.

2.2. Multi-frame vs. Single-frame

In [2], the NLM-based image denoising algorithm was extended to video denoising by aggregating patches in a space-temporal volume, which avoids the explicit motion estimation. Due to the temporal coherence of image sequence, more similar patches can be found than just searching within the current single frame. As shown in Fig. 2, the performance of multi-frame NLM-based denoising performance (using the three adjacent frames) is much better than the performance of single-frame NLM-based denoising.

However, when the adjacent frames are having different illumination conditions, like the flashlight, the redundancy of similar patches founded by the original NLM-based method is suddenly reduced, as the example illustrated in Fig. 1. It motivates us to further improve the NLM-based video denoising by making the similar patches searching much more robust to the illumination changes.

2.3. Illumination-Invariant Improvement Strategy

In this subsection, we address the issue of video denoising under variable illumination conditions. First, three possible

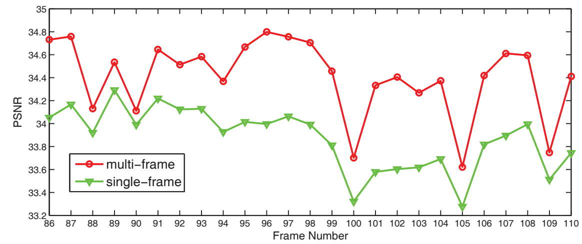


Fig. 2. PSNR curves of the multi-frame and single-frame NLM-based denoising methods on *Crew* sequence.

technical improvements are presented. Then, by further analysis and comparisons of their efficacy on eliminating the influence of illumination changes, we propose an extension to the NLM-based algorithm with special handling of the illumination changes.

Illumination condition often has influence on the local contrast of image.To make the adjacent frames around illumination changes have similar visual appearances, the local contrast should be adjusted. Based on the NLM-based framework, one can have three process scales: patch scale, search window scale, and frame scale. The patch scale has less pixels and therefore is more sensitive to the noise, while the frame scale captures the scene structures but may not be adaptive to the local illumination changes. The search window scale is a good tradeoff between the patch scale and frame scale.

Let $S(i,t)$ be the search window centered at the pixel i in tth frame. $S(i, t+k)$ denotes the search window in the kth adjacent frame. We define the filter $\mathcal{F}$ which processes $S(i, t+k)$ to make it have a similar visual appearance to $S(i,t)$. For convenience, let p_t, q_t, μ_t denote the minimum, maximum, and average intensity level of $S(i,t)$. We proposed three improvement strategies to implement the filter $\mathcal{F}$:

- Direct linear mapping function (*i.e.*, contrast-stretching transformation), $\mathcal{F}_{cst}$;

$$\begin{aligned} v_{out} &= \mathcal{F}_{cst}(v_{in}) \\ &= \left(v_{in} - \frac{q_{t+k} + p_{t+k}}{2}\right) \frac{q_t - p_t}{q_{t+k} - p_{t+k}} + \frac{q_t + p_t}{2}. \end{aligned} \tag{3}$$

- Linear mapping function preserving the average intensity level of $S(i,t)$, $\mathcal{F}_{avg}$;

$$v_{out} = \mathcal{F}_{avg}(v_{in}) = c \cdot (v_{in} - \mu_{t+k}) + \mu_t, \tag{4}$$

  where $c = \min(\frac{\mu_t - p_t}{\mu_{t+k} - p_{t+k}}, \frac{q_t - \mu_t}{q_{t+k} - \mu_{t+k}})$.

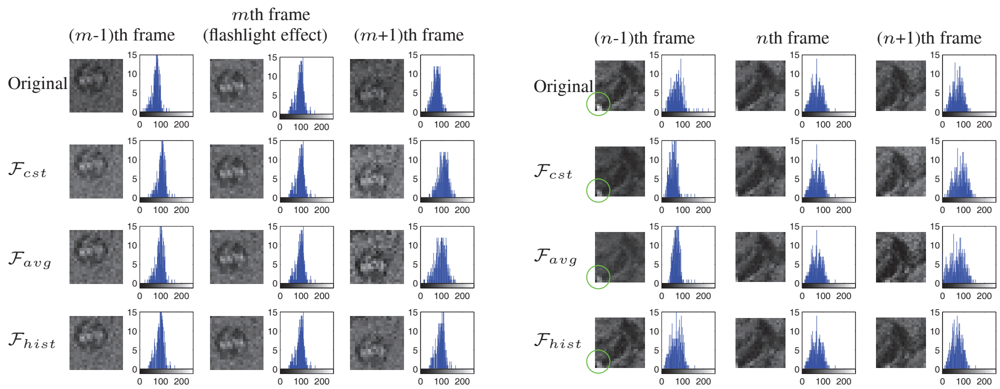
- Histogram specification processing, $\mathcal{F}_{hist}$.

$$v_{out} = \mathcal{F}_{hist}(v_{in}) = G^{-1}(T(v_{in})), \tag{5}$$

  where $T(r) = \int_0^r p_r(w)dw, G(z) = \int_0^z p_z(t)dt$, and $p_r(r), p_z(z)$ are the probability density functions corresponding to intensity level r and z.

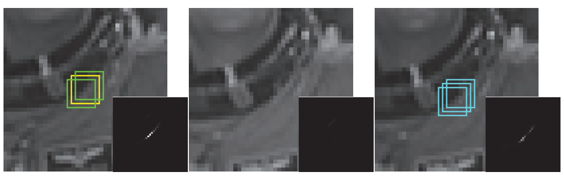
Now we study and compare the efficacy of the three strategies for reducing the influence of illumination changes. For the denoising purpose, the modification should satisfy the following requirements:

- Eliminating the influence of sudden illumination condition changes;



(a) (b)

Fig. 3. Comparison of the filtering effects by three filtering methods. (a) search windows in three adjacent frames with illumination condition changes in the middle frame; (b) search windows in three adjacent frames without illumination changes.



(a) $(k-1)$th frame (b) kth frame (c) Filtered kth frame

Fig. 4. Comparison of non-local patch searching before and after filtering of the kth frame with flashlight.

- Preserving the performance of original NLM-based method in regions without illumination changes;
- Robust to noise and scene changes in image sequence.

According to the requirements, we verify the three strategies in different scenarios in Fig.3. In each subfigure, the first row shows the original search window images at three adjacent frames. Filter results of these images by $\mathcal{F}_{cst}$, $\mathcal{F}_{avg}$ and $\mathcal{F}_{hist}$ are illustrated in the 2nd, 3rd and 4th rows, respectively. The histogram of each filtered image is also given aside.

In Fig.3(a), illumination change occurs at the mth frame caused by flashlight. The linear mapping function $\mathcal{F}_{cst}$ can adjust the local contrast, but can not guarantee to keep the same average intensity of the filtered image. Although $\mathcal{F}_{avg}$ addresses this problem, both of $\mathcal{F}_{cst}$ and $\mathcal{F}_{avg}$ are not robust to the local scene changes, *e.g.*, in Fig.3(b), the bright pixels (within green circles) appear in the $(n-1)$th frame but disappear in the nth and $(n+1)$th frames. In this situation, $\mathcal{F}_{cst}$ and $\mathcal{F}_{avg}$ produce unsatisfactory filtering results as the majority of pixel intensities are scaled into low contrast due to the existence of a few bright pixel intensities.

In both cases, $\mathcal{F}_{hist}$ can produce plausible visual effects and fulfill the above requirements. Therefore, we select $\mathcal{F}_{hist}$ as the core block of our proposed denoising system. A processing example is shown in Fig.4, comparing the non-local search results before and after the histogram processing on the kth frame with flashlight effect.

Once the search windows are processed, one can perform the non-local patch search in the filtered 3D space-temporal volume to find more "intrinsic" visually similar patches.

3. EXPERIMENTAL RESULTS

We conduct experiments to verify the efficacy of our proposed method. In setting the parameters of NLM-based framework, the patch size is set to 7×7, the searching space-temporal volume is set to $21 \times 21 \times 3$ (adjacent three frames), and the parameter h in Eqn.(2) is set to $h = k \cdot \sigma_e$ with $k \in [0.7, 1]$. In the experiments, we select the optimal k producing the best Peak Signal-to-Noise Ratio (PSNR) result for original NLM-based method. Once the parameters are selected, they keep fixed for the proposed method for fair comparison.

We test our proposed method on the *Crew* sequence, which contains the illumination changes caused by flashlight, and compare the results with original NLM-based video denoising method [2]. The original sequence is contaminated by Gaussian noise with standard deviation $\sigma_e = 10$. The PSNR gains of the proposed method (denoted as Hist-NLM) and the multi-frame NLM-based method (denoted as NLM) over the single-frame NLM-based method are illustrated in Fig.5. It is obvious that the proposed method produces much better PSNR results at the frames with illumination changes (*e.g.*, 89th, 90th, 100th, 105th, 109th.), and well maintains the performances of multi-frame NLM-based denoising method at frames with stable illumination conditions.

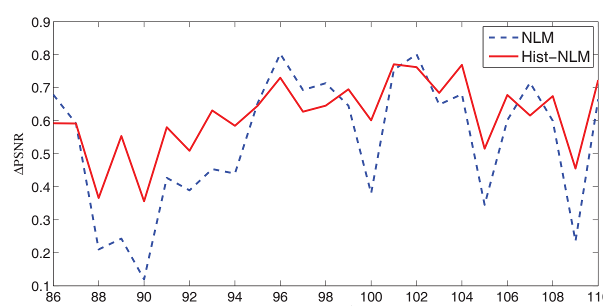
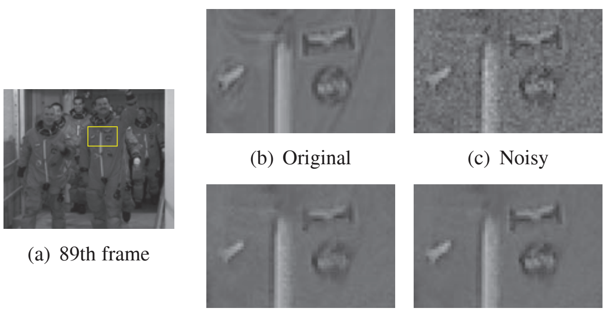


Fig. 5. The PSNR gains of the proposed method (Hist-NLM) and the multi-frame NLM-based method over the single-frame NLM-based method on *Crew* sequence frame by frame.

Table 1. The PSNR results (dB) on denoising three frames with sudden illumination changes for five noise levels.

σ_e	h	89$^{\text{th}}$ Frame			100$^{\text{th}}$ Frame			109$^{\text{th}}$ Frame		
		NLM	Hist-NLM	ΔPSNR	NLM	Hist-NLM	ΔPSNR	NLM	Hist-NLM	ΔPSNR
5	4	38.1452	**38.4282**	0.2830	37.3436	**37.6314**	0.2878	37.1847	**37.5285**	0.3438
10	7	34.5351	**34.8459**	0.3108	33.7015	**33.9227**	0.2212	33.7479	**33.9682**	0.2203
15	13	30.9084	**31.4788**	0.5704	30.8418	**31.3479**	0.5061	30.8310	**31.3753**	0.5443
20	16	29.6137	**30.2102**	0.5965	29.4825	**29.9174**	0.4349	29.3324	**29.9425**	0.6101
25	20	28.4888	**29.1919**	0.7031	28.1683	**28.6627**	0.4944	27.9787	**28.6938**	0.7151

For the 89th, 100th, and 109th frames, the denoising results of multi-frame NLM-based method and our proposed method with five noise levels ($\sigma_e = 5, 10, 15, 20, 25.$) are given in Table 1. For these frames with sudden illumination changes, our proposed method improves the original NLM-based method by 0.2-0.7dB. In Fig. 6, we compare the visual quality of two denoising methods for the 89th frame. Our proposed method removes most of the noise in the flashlight region (*e.g.*, the clothes of the right-front person) and preserves more details than the original NLM-based method, which can be seen from the zoomed images in Fig. 6 (d) and (e).



(a) 89th frame (b) Original (c) Noisy (d) Multi-frame NLM (e) Proposed

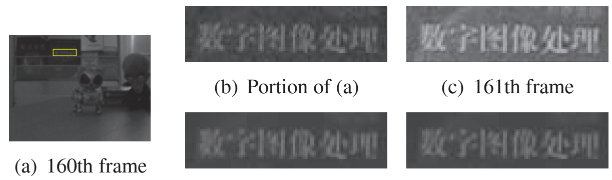
Fig. 6. Denoising results on the 89th frame of *Crew* sequence. (d) NLM, PSNR=32.9263dB, SSIM =0.8634; (e) Proposed, PSNR=**34.2063**dB, SSIM=**0.8904**.

Lastly, we run our system on a real noisy video sequence, *Indoor*, captured by Canon PowerShot-A570 under low light condition with flashlight effect. The 160th noisy frame and zoomed denoised images are shown in Fig. 7. Our method is able to collect more information from the next illumination-changed frame and recover more details while removing the noise (Notice the details of the first Chinese character).

For more experimental results, please visit our project website: http://www.icst.pku.edu.cn/course/icb/Hist-NLM.html.

4. CONCLUSION

NLM-based algorithms are popular in video denoising techniques. However, most of them ignore the influence of illumination changes. In this paper, we attempt to extend the NLM-based video denoising to handle illumination condition changes. Three possible technical improvements are presented and by a detailed analysis and comparison, the histogram specification processing is verified to produce robust results to illumination changes, noise and scene change, which is finally integrated into the NLM-based denoising system. Experimental results show that our method is efficient to video denoising where there are illumination changes. This work can be easily integrated with the illumination change detection techniques or extended to a color version handling the chrominance and illumination changes simultaneously.



(a) 160th frame (b) Portion of (a) (c) 161th frame (d) Multi-frame NLM (e) Proposed

Fig. 7. Denoising results on the 160th frame of a real noisy video *Indoor*.

5. REFERENCES

[1] M.K. Ozkan, M.I. Sezan, and A.M. Tekalp, “Adaptive motion-compensated filtering of noisy image sequences,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 3, no. 4, pp. 277–290, Aug. 1993.

[2] A. Buades, B. Coll, and J.M. Morel, “Denoising image sequences does not require motion estimation,” in *AVSS 2005. IEEE Conference on Advanced Video and Signal Based Surveillance*, Sept. 2005, pp. 70 – 74.

[3] A. Buades, B. Coll, and J.-M. Morel, “A non-local algorithm for image denoising,” in *CVPR 2005. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 2005, vol. 2, pp. 60–65.

[4] M. Mahmoudi and G. Sapiro, “Fast image and video denoising via nonlocal means of similar neighborhoods,” *IEEE Signal Processing Lett.*, vol. 12, no. 12, pp. 839 – 842, Dec. 2005.

[5] T. Tasdizen, “Principal neighborhood dictionaries for nonlocal means image denoising,” *IEEE Trans. Image Processing*, vol. 18, pp. 2649–2660, Dec. 2009.

[6] Z. Ji, Q. Chen, Q.-S. Sun, and D.-S. Xia, “A moment-based nonlocal-means algorithm for image denoising,” *Inf. Process. Lett.*, vol. 109, pp. 1238–1244, Nov. 2009.

[7] Ce Liu and W. T. Freeman, “A high-quality video denoising algorithm based on reliable motion estimation,” in *Proceedings of the 11th European conference on computer vision conference on Computer vision: Part III*, Berlin, Heidelberg, 2010, pp. 706–719, Springer-Verlag.