

Comprehensive Survey on Video Denoise Methods

Zhanming Bi

*International College, Guangxi University, Nanning, China
261052@qq.com*

Abstract

Today we have entered a smart information age on a variety of carriers, especially images and videos as carriers of information are widely used in our lives. There is a strong demand for clearer images and more visual enjoyable videos. However, in modern society, images and videos are inevitably affected by noise, which leads to a deterioration in image quality. We therefore need to reduce noise without losing as much of the image characteristics (edges, corners, resolution) as possible. So far, researchers have proposed a considerable number of classes of noise removal methods, each with its own advantages and disadvantages. In this paper, we present several traditional denoising methods as well as neural network-based denoising methods, supplemented by additional denoising methods based on human vision. We present the main formulations of each algorithm and give experimental results, as well as analyzing the advantages and disadvantages of them. We conclude with ideas for future trends in denoising methods finally.

Keywords

Video and Image Noise, Denoise Methods, CNN, HVS, Perceptual Quality

1. Introduction

Video noise is the random variation in brightness or color of an image produced by a sensor, scanner circuit or digital camera. Video noise also originates from point noise in film grain size and invariant quantum detectors. Video noise is often seen as an unwanted component of image acquisition. According to the different characteristics and distributions of different noise, video noise is divided into the following categories: Gaussian noise, Poisson noise, multiplicative and additive noise and pretzel noise [1]. For example, Gaussian noise is a class of noise whose probability density function follows a Gaussian distribution, and Poisson noise is a noise model that follows a Poisson distribution. In general, most video noise from natural signals satisfies the Gaussian distribution. Therefore, lots of research are conducted on the image and video Gaussian type noise. In order to reduce the image noise, numerous works are proposed.

To address the problem of image denoising, by presenting a novel approach on using adaptive principal components for image denoising, D. Darian Muresan et.al improved the image visual fidelity [2]. Antoni Buades has proposed a new method, the non-local means (NL-means), which is an image denoising method based on the non-local averaging of all pixels in the image. This method effectively improves local smoothing filters [3]. Chunwei Tian uses machine learning methods in image denoising, and he proposed a novel method called enhanced convolutional neural denoising network (ECNDNet). This new approach uses residual learning and batch normalization techniques to address training difficulties and accelerate network convergence. [4].

In addition to solving the image denoising problem, video denoising is also a point of interest. Michele Claus proposed ViDeNN (CNN for Video Denoising): a CNN (Convolutional neural network) for video denoising that combines spatial and temporal filtering, learning to perform spatial denoising for this method without prior knowledge of the noise distribution (blind denoising) [5]. It has been shown through extensive experimental data that the algorithm has the same excellent results as other state-of-the-art algorithms. Andrea Petreto also proposed a new real-time embedded video denoising algorithm(RTE-VD), which can denoise the video in real-time [6], while its output image quality is worse than the other advanced algorithms including the VBM3D [7]. Matias Tassano et al. developed DVDnet (Deep Video Denoising Network): a fast network for deep video denoising based on a CNN

architecture, an approach that significantly reduces computation time as well as memory footprint and is able to handle a wide range of noise levels with a single network model [8].

Although there are many noise reduction methods for images and videos above, all these methods have their own application characteristics. On the one hand, in the traditional video image filtering methods, some methods do not take into account the different content of different regions and use the same filtering parameters in the whole image and video, which may lead to large differences in regional filtering effects. Then, for example, some videos use time-domain filtering methods, but the methods are too complex not easy to implement in real-time. On the other hand, the noise reduction methods based on neural network learning are usually only applicable to limited data sets and are difficult to be adaptive according to the content. In addition, most of the methods do not start from the perspective of human eye vision, and most of them start from the perspective of statistical signals, and the filtering results may not be the best noise reduction for human eyes.

This paper investigates the video denoising methods proposed in the past three years by analyzing the above methods and deeply analyzes the advantages and disadvantages as well as the applicability of each video denoising method. And based on this research, we predict the future development trend of video denoising algorithms and point out the possible future development direction of video denoising to help solve the above problems now.

The remainder of this paper is organized as follows. In Section II, we present and analyze some traditional video and image denoising methods, while showing experimental results. In Section III, neural network-based denoising methods and show experimental results are presented and analyzed. In the IV Section, we show some human-eye vision-based denoising methods and demonstrate the experimental results. Finally, we conclude with a summary of the above algorithms and give our expectations for the future in the Section IV.

2. Traditional Video Denoise Methods Analysis

In this section, we will analyze the traditional methods of image denoising, namely the NL-means, and the deep analysis is given finally. Firstly, NL-means is defined by the formula,

$$NL[u][x] = \frac{1}{C(x)} \int_{-\Omega}^{\Omega} e^{-\frac{(Ga * |u(x+) - u(y+)|)^2}{h^2}} u(y) dy \quad (1)$$

where $x \in \Omega$ and $C(x) = \int_{-\Omega}^{\Omega} e^{-\frac{(Ga * |u(x+) - u(z+)|)^2}{h^2}} dz$, respectively. $C(x)$ is the normalization function, Ga is the Gaussian kernel, and h is the filtering parameter. This equation means that the denoised value at x is the average of all points in the Gaussian neighborhood of x . Experimental results have proven that the NL-means method can denoise efficiently as shown in Figure 1. Figure 1(a) shows the original image, and Figure 1(b) shows the image after processing by NL-means algorithm. From the intuitive perspective of the human eye, we can find that the NL-means algorithm is still at a high level of detail retention and structure restoration, and handles noise very well.



(a) Original Image



(b) NL-means Filtered Image

Figure 1: The filter results of NL-means method [3]

Secondly, expect for the image denoise method above, we will introduce a video denoising algorithm as the RTE-VD method. The RTE-VD is defined by the formula,

$$I_f(x) = \frac{1}{W_p} \sum_{x_i \in \Omega} I_t(x_i) e^{-\frac{[I_t(x_i) - I_t(x)]^2}{2\sigma_i^2}} * e^{-\frac{[x_i - x]^2}{2\sigma_d^2}} \quad (2)$$

where,

$$W_p = \sum_{x_i \in \Omega} I_t(x_i) e^{-\frac{[I_t(x_i) - I_t(x)]^2}{2\sigma_i^2}} * e^{-\frac{[x_i - x]^2}{2\sigma_d^2}} \quad (3)$$

and,

$$I_t = I_p(x) e^{-\frac{[I_t(x_i) - I_t(x)]^2}{2\sigma_i^2}} + I(x) (1 - e^{-\frac{[I_t(x_i) - I_t(x)]^2}{2\sigma_i^2}}) \quad (4)$$

where I_f is filtered image, I_p is the previous image, I is the current image, x is the coordinates of the image, Ω is the filter kernel, σ_i , σ_d and σ_t are the smoothing parameters. Experimental results have proven that the RTE-VD method can denoise efficiently as shown in Figure 2. Figure 2 (a) shows the original image, and Figure 2 (b) shows the image after processing by RTE-VD algorithm. What we can know from the experiments is that it is able to recover details and achieve performance improvements on very noisy videos.



Figure 2: The filter results of RTE-VD method [6]

As a conclusion, among the image and video denoising methods above, although they achieve acceptable results, there still exist weakness. On one hand, NL-means algorithm can play a good role in resisting noise while processing details and displaying fine structures. But nonlocal algorithms work only in similar windows and in restricted scales with special characteristics. On the other hand, RTE-VD is not only able to recover details perfectly on very noisy videos in real-time denoising, but also has 70 times better performance than other algorithms. However, its denoising results are worse than other advanced filters.

3. Learning Based Video Denoise Methods Analysis

In this section, we will analyze the methods of image denoising which are based on machine learning including ECNDNet and ViDeNN. Firstly, we will elaborate the ECNDNet method. The main algorithm of ECNDNet,

$$l(p) = \frac{1}{N} \sum \|f(y_j; p) - (y_j - x_j)\|^2 \quad (5)$$

where p is the parameters, y_j is the j th noisy image patch and x_j is the j th label image patch. Expanding the convolutional network at layers 2, 5, 9, and 12 not only captures the information but also

reduces the computational cost. The whole network architecture is shown as Figure 3. In addition, the residual learning enables this neural network to perform better in denoising.

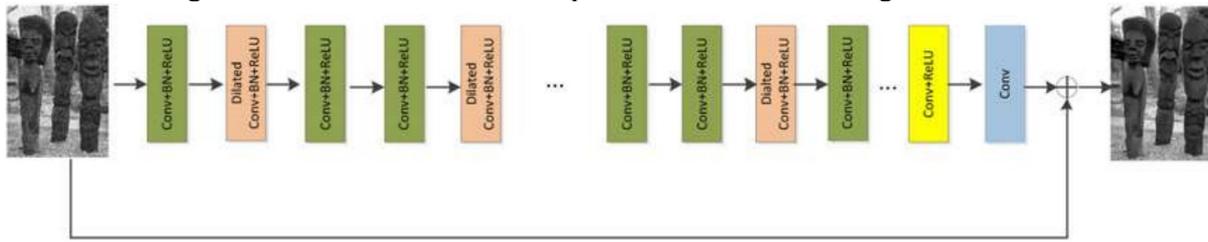


Figure 3: Architecture of ECNDNet [4]

Experimental results have proven that the ECNDNet can denoise efficiently as shown in Figure 4. Figure 4 (a) shows the original image, and Figure 4 (b) shows the image after processing by the ECNDNet. The experimental results can show that the ECNDNet algorithm not only improves the network performance and makes the network easier to train while reducing the computational cost, but also obtains clean images from the noisy images.

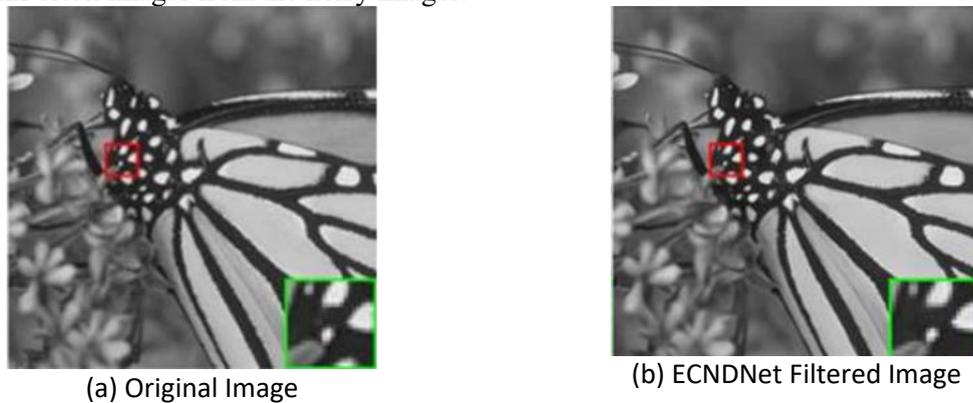


Figure 4: The filter results of ECNDNet method.

Secondly, neural networks can also be applied to video denoising, for example the ViDeNN is a good case study. The structure of the proposed ViDeNN network is shown as Figure 5. Each frame is passed through a spatially denoised CNN. The temporal CNN takes the three spatially denoised frames as input and outputs a final estimate of the central frame. These two CNNs first estimate the noise residuals, and then reduce them from the noisy input signal. ViDeNN consists of only convolutional layers. The number of feature maps is written at the bottom of each layer. The number of feature maps is written at the bottom of each layer.

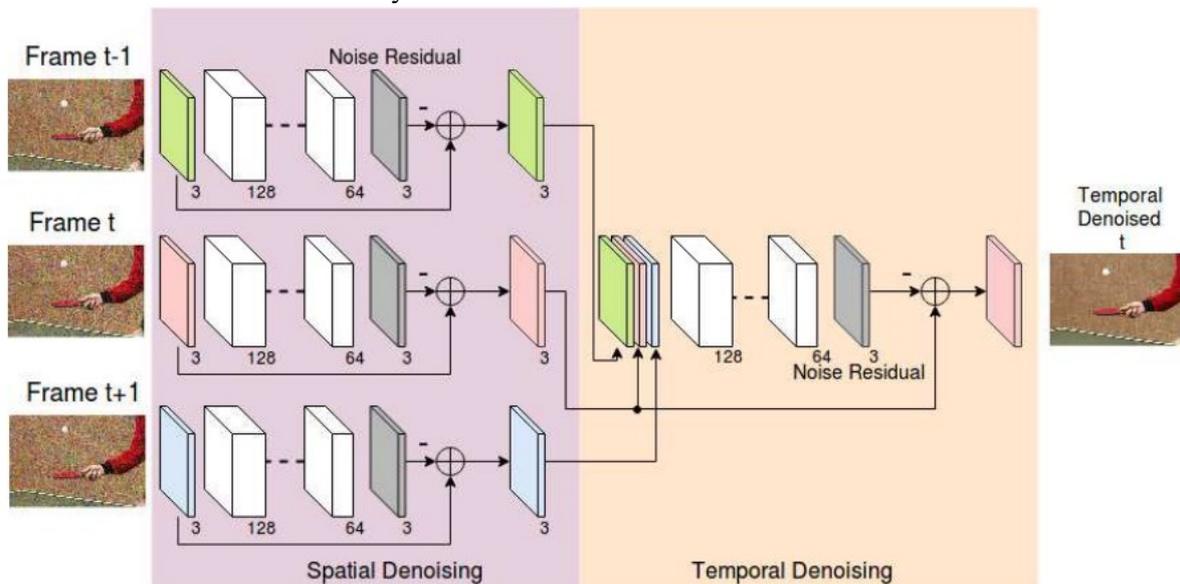


Figure 5: Architecture of ViDeNN [5]

Experimental results have proven that the ViDeNN algorithm can denoise efficiently as shown in Figure 6. Figure 6 (a) shows the original image, and Figure 6 (b) shows the video after processing by the ViDeNN algorithm. The data show that videos using the ViDeNN algorithm can be processed for noise reduction without pre-analysis of the signal. And with ViDeNN using a simplified automatic framework, video denoising can reduce the time and cost required to obtain clean video images simply and efficiently.



Figure 6: The filter results of ViDeNN method

Neural networks can be of great advantage in both video and image denoising. ECNDNet is able to improve network performance for fast image processing, and ViDeNN is able to blindly denoise any video almost unconditionally. In short, they both perform the denoising task very well and stand out among similar algorithms. However, they have great data dependency, which means they can only work efficient when the training and test data are available for the applications. When the situation is changed, the network should be re-trained, which will limit their practical application.

4. HVS Based VIDEO Denoise Methods Analysis

In this section, we will analyze HVS (Human Visual System) based video denoise methods including Multi-Scale multiscale structural similarity for image quality assessment (MS-SSIM) [9] and JND (Just-Noticeable-Difference) based perceptual filter methods. Firstly, MS-SSIM is defined by the formula as,

$$L = \alpha L_1 + \beta L_{MSSSIM} \quad (6)$$

where,

$$L_1(p) = \sum_{p \in P}^N |y(p) - \hat{y}(p)| \quad (7)$$

and,

$$L_{MSSSIM}(P) = [1 - MSSSIM(y(p), \hat{y}(p))] \quad (8)$$

where p represents a pixel in the whole image patch P of the image patch, there are P pixels in that patch. y and \hat{y} represent the real image and the restored image, α and β are two scalars. The MS-SSIM metric can be used evaluate the image structure similarity, which will help to reserve the structure and edge well. Experimental results have proven that the MS-SSIM method can denoise efficiently as shown in Figure 7. Figure 7 (a) shows the original image, and Figure 7 (b) shows the image after processing by MS-SSIM algorithm. From the experimental results, the method can retain the contrast in the high frequency region while denoising in the structures and edges.

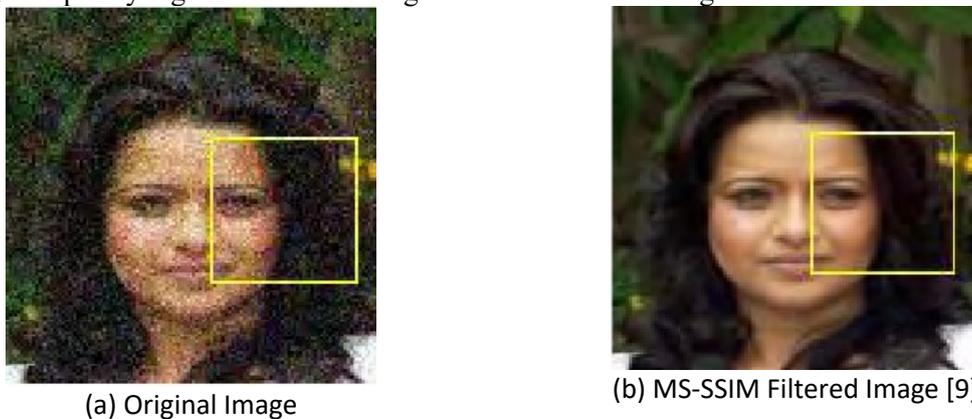


Figure 7: The filter results of MS-SSIM method

Secondly, we can present the JND model to conduct the perceptual denoise [10]. The JND means the just noticeable difference which human eyes can perceive. We can use JND model to enhance the perceptual noise with contrast masking. filter the un-noticeable perceptual noise to reduce the image high frequency information and complexity. The algorithms based on JND model is defined by the formula,

$$d_{out}(a, b) = e * VM(a, b) * LA(a, b) * d(a, b) \quad (9)$$

where $d_{out}(a, b)$ and $d(a, b)$ are the output of the noise reduction and the output of the contrast enhancement of the noise-aware detail layer, respectively. e is a control parameter for the degree of noise reduction. $LA(a, b)$ is the ratio of JND thresholds to calculate and $VM(a, b)$ is the visual masking. Experimental results have proven that the JND model method can denoise efficiently as shown in Figure 8. Figure 8 (a) shows the original image, and Figure 8 (b) shows the image after processing by JND model. Experimental results show that the model enhances the contrast of low-light images while successfully minimizing distortion and preserving detail. It means that we can get a low-filtered image with almost the same visual quality.



Figure 8: The filter results of JND model method [10]

Different from the filter methods above, the HVS base filter methods mainly focus on the human perceptual noise rather than the statistic noise. Usually the human visual characteristics are considered and explored to guide the filter model and procedure. There methods can help to obtain filtering results for better subjective quality rather than the objective quality focused by the other methods.

5. Conclusion

In this paper, we present the NL-means method, this algorithm is a non-local average image denoising method based on all pixels in the image, which effectively improves the local smoother. Also for image denoising, we can use the non-traditional neural network denoising method ECNDNet, which uses neural networks as well as residual learning to solve difficulties and speed up the fusion of the network.

For research in video denoising, we present the RTE-VD algorithm, a method that can denoise video in real-time, although its output image quality is slightly inferior to other advanced algorithms. However, the use of neural networks in the field of video denoising can lead to even better results, and to this end we present ViDeNN. The algorithm that uses CNNs for video denoising, which combines spatial and temporal filtering and has excellent results in the field of blind denoising.

We also present and analyze denoising methods in human vision, including the MS-SSIM algorithm and the JND model algorithm. The algorithms that focus on human vision both achieve the visual friendly low-filtered image quality, instead of only the objective filter performance.

From the research in this paper, we can see that each class of noise reduction methods has its own characteristics, but these methods still do not achieve enough generality and have their own limitations. Combined with the current research results, we believe that in the future, video and image noise reduction methods can be further studied in depth from the following aspects. On the one hand, for videos and images viewed by the human eye, the filtering methods should take into account the visual model as much as possible, so that the final filtering result obtains the best subjective quality. On the other hand, for videos and images that are analyzed by machines, the filtering method should be combined with the purpose of machine analysis such as face recognition and so on. In this way the best balance of model complexity and machine analysis efficiency can be achieved.

References

- [1] Sun P, Wang C, Li M, et al. Partial Differential Equations-Based Iterative Denoising Algorithm for Movie Images[J]. *Advances in Mathematical Physics*, 2021, 2021.
- [2] Muresan D D, Parks T W. Adaptive principal components and image denoising[C]//*Proceedings 2003 International Conference on Image Processing (Cat. No. 03CH37429)*. IEEE, 2003, 1: I-101.
- [3] Buades A, Coll B, Morel J M. A non-local algorithm for image denoising[C]//*2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. IEEE, 2005, 2: 60-65.
- [4] Tian C, Xu Y, Fei L, et al. Enhanced CNN for image denoising[J]. *CAAI Transactions on Intelligence Technology*, 2019, 4(1): 17-23.
- [5] Claus M, van Gemert J. Videnn: Deep blind video denoising[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 2019: 0-0.
- [6] Petreto A, Romera T, Lemaitre F, et al. A new real-time embedded video denoising algorithm[C]//*2019 Conference on Design and Architectures for Signal and Image Processing (DASIP)*. IEEE, 2019: 47-52.
- [7] Makkar H, Lamba O S. An improved VBM3D filtering technique for removal noise in video signals[J]. *European Journal of Advances in Engineering and Technology*, 2017, 4(8): 584-591.
- [8] Tassano M, Delon J, Veit T. Dvdnet: A fast network for deep video denoising[C]//*2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019: 1805-1809.
- [9] Tassano M, Delon J, Veit T. Dvdnet: A fast network for deep video denoising[C]//*2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019: 1805-1809.
- [10] Uddin A F M S, Chung T, Bae S H. A perceptually inspired new blind image denoising method using L_1 and perceptual loss[J]. *IEEE Access*, 2019, 7: 90538-90549.
- [10] Su H, Jung C. Perceptual enhancement of low light images based on two-step noise suppression[J]. *IEEE Access*, 2018, 6: 7005-7018.