ISSN (Online) 2394-6849

International Journal of Engineering Research in Electronics and Communication Engineering (IJERECE)
Vol 4, Issue 8, August 2017

Video Denoising using Adaptive Transform Domain Approach

[1] Lina Desale, [2] Prof. S. B. Borse
[1] M.E Student, [2] HOD, E&TC Dept
[1][2] LGNSCOE, Nashik

***Abstract* - Utmost prevailing practical digital video denoising techniques depend upon a traditional statistical model of image noise, such as Independent and identically distributed random variables- Gaussian noise, which is often violated in real life scenarios. For example, following foremost sources of video noise with dissimilar statistical distributions have been identified: photon shot noise, fixed pattern noise, amplifier noise, dark current noise and quantization noise. Performance of prevailing video denoising algorithms will severely degrade when the algorithm is applied on those images with noises evident from multiple sources. In this paper, a dictionary learning based scheme is proposed which computes the basis function adaptively from the first input image frames per fifty frames. Unlike other classical approaches like wavelet or contourlet transforms where the mother wavelet/basis functions are constant. If the mother wavelet/basis function is constant it is more likely that it will fail to capture the minuscule noise details from real life images. Therefore, the basis function is learnt from first frame itself. The dictionary learning method provides sparse representation of the image. Here, hard thresholding algorithm is applied to compute the denoised frame.**

Index Terms— video denoising, basis formation, SVD, post processing.

I.INTRODUCTION

With scientific developments in sensor design, the matrix data in image/video is relatively noiseless for superior resolution digital cameras at low sensitivities, but frames are noisy for low cost cameras at high sensitivities. The video matrix data tends to contain more noise as compared with single frame because of video capturing at high speed using the video camera. The process of video denoising primarily focuses at eliminating existing noise from all frames in the video. It employs pixel level matrix data in both the spatial and temporal domains. This unified method is finest than independently smearing a frame denoising method on each individual frame of the video, as there subsists great temporal redundancies in a video which is collection of image frames.

Most usual video denoising techniques employ a single statistical model of image noise. This model is built from Independent and identically distributed random variables Gaussian noise. Major sources of video noise with dissimilar statistical distributions are: photon shot noise, immobile pattern noise, intrinsic amplifier noise, dark current noise and digital quantization noise. Results of most recent video denoising algorithms degrade on real life noisy frames containing noises from multiple sources.

Even if video is a collection of 3-D numbers set, 3-D data manipulators are usually not used for its transformation. Undoubtedly, the typical 3-D data manipulators do not yield appropriate transformations for most video data. In general, maximum 3-D data manipulations are 1-D separable. Notwithstanding, multidimensional (M-D) separable implementations contain artifacts that can heavily hamper their utility, specifically for M-D image data (real life scenarios). Illustratively, the de facto separable 3-D manipulator is scarcely employed for digital video transformation as it combines sub-bands related to 3-D orientations. Thus for a 3-D video manipulator to provide a faithful representation of M-D video matrix, the mathematical manipulation must be non-separable. But, non-separable data manipulations are computationally heavy as well as difficult to design and implement. In the case of video denoising, the principal to be followed is to explore the sparse representation of video frames. If one considers frequency (transform) domain representation, sparse representation is related to high-kurtosis marginal distribution of band-pass filtering. For the representation in spatial domain the non-local means (NLM) algorithm [1] was presented to reduce noise by computing mean of pixels in a video frame equivalently weighted by neighboring patch similarities. Sparse matrices are major part of video representation. Amongst two noise-free set of frames (videos) that yield the equal mean peak signal-to-noise ratio (PSNR), we adopt the one with higher temporal coherence. Maximum real life image sequences incorporate structured noise which makes it a challenging problem to ensure temporal coherence.

ISSN (Online) 2394-6849

International Journal of Engineering Research in Electronics and Communication Engineering (IJERECE)
Vol 4, Issue 8, August 2017

II. LITERATURE SURVEY

Buades et al. [2] suggested a method that utilizes the self-similarity and redundancy amongst adjacent video frames. The algorithm is inspired by data fusion procedures, and subsequently the frame count increases, results in a pure temporal mean. The conventional video raster motion compensation using classical optical flow methods allows hefty patch resemblance in the desired spatiotemporal data volume. Application of PCA guarantees the exact preservation of fine texture and minute details.

Chen et al. [3] presented a Robust Kronecker Product Video Denoising (RKPVD) method using fractional-order total variation. The video model is projected to remove classical Gaussian-impulse noises from input video matrix data. Conclusive under-determined minimization problem, which includes of nuclear norm, Kronecker product sparse L1 norm and FTV, determined by a two-stage algorithm pooled with alternating direction method (ADM). The resultant RKPVD video denoising algorithms for irradiating intrinsic mixed Gaussian impulsive noise are authenticated in the video frame experiments. Rigorous experiments exhibited that the RKPVD algorithm has a superlative PSNR (peak signal-to-noise ratio) and modified visual minute detail restoration.

Almahdi et al. [4] described a novel iterative recursive Non-Local means (RNLM) method for video restoration. Besides, they have protracted this method by integrating a Poisson-Gaussian noise model. Their newfangled RNLM scheme delivers a computationally light approach for video denoising, and produces better performance benchmarked with the single image NLM and BM3D algorithms. Non-Local means (NLM) based methods are considered as classical approaches for image and video denoising related applications. Extension of this method from 2-D (image) to 3-D (video) for accessible video frame processing is multifaceted. The RNLM approach is heavily dependent upon recursion for the purpose of computational savings, and spatio-temporal correlations for enhanced performance. In their approach, the first frame is given as input to single frame NLM. Subsequent video frames are estimated with a weighted combination of the current frame NLM, and the preceding frame estimate. Experimental results demonstrate superiority of the presented approach. They have proved that the new approach outperforms classical single frame NLM and BM3D algorithms.

Redha et al. [5] proposed a pressingly effective algorithm for video denoising that explores temporal and spatial redundancy. The proposed algorithmic technique by Redha et al. [5] is heavily based upon intrinsic non-local means. NLM methods are smeared in video denoising applications. Each output video pixel is formed as a weighted summation of the center pixels within a given video search window. The SVD weights are established on the image patch intensity consisting of algebraic vector distances. The iterative process necessitates computing feature vector Euclidean distances for all of the video patches in the intrinsic search window. Straight extension of this technique from 2-D to 3-D, for video processing, can be computationally severe. The size of a 3-D search window is the size of the 2-D search window multiplied by the number of frames being used to form the output. They have presented a novel recursive NLM (RNLM) algorithm for video processing. Their RNLM method takes benefit of recursion for computational savings, as benchmarked with the direct 3D NLM. Conversely, unlike the 3D NLM, their proposed method is capable to feat both spatial and video temporal redundancy for improved performance, compared with 2D NLM. In their method, the first frame is processed with single-frame NLM algorithm. Subsequent video frames are algorithmically estimated using an actual weighted sum of internal pixels from the current video frame and a pixel from the previous video frame estimate. Numerous experimental results are obtained to validate the efficacy of proposed approach in terms of subjective and quantitative image quality.

Khalid et al. [6] considered the benefits of Surfacelet transform which were combined with Artificial Bee Colony (ABC) optimization method for adaptive threshold optimization. The resultant video sequences in denoised form have conveyed astonishing results as measured using peak signal to noise ratio (PSNR) and structural similarity (SSIM) index.

Guo et al. [7] developed a novel approach to video denoising that is built upon the notion that in general, videos corrupted with noise can be segregated into two components - the approximate low-rank component and the sparse component. Initially they split the given video sequence into these two components, and subsequently applied prevailing state-of-the-art denoising algorithms on each component. They have

ISSN (Online) 2394-6849

International Journal of Engineering Research in Electronics and Communication Engineering (IJERECE)
Vol 4, Issue 8, August 2017

showed, with the help of extensive experiments, that their denoising approach outperforms the state-of-the art denoising algorithms.

III. SYSTEM DESIGN AND OVERVIEW

A. Dictionary Learning

Sparse data exemplification has made its place as an tremendously useful tool for securing, demonstrating, and compressing high-dimensional signals. The authoritative categories of existing real life data values such as audio and videos have in built real life sparse representations which are explored with fixed standard bases (i.e., Fourier, Wavelet). In addition, singular methods dependent upon protuberant optimization or avaricious pursuit are heavily employed for high performance computing such as decompositions with high fidelity applications. In the field of computer vision, the matrix or spatial and variation semantics of a video frame are vital as compared with a dense, compressed video format. In the recent decades, L1 norm minimization have been extensively employed as main technology for vision tasks, which comprise of facial recognition, matrix super-resolution, video motion and matrix segmentation, machine learning based supervised video denoising and video in-painting and formation of background noise model and video frame classification.

The skill of classical sparse exemplifications is to reveal video frame semantic matrix data is holdings of the existing video data. Which conclusively indicate that, they fit upon or lie near low-dimensional subspaces and sub manifolds. If a assortment of illustrative samples are established for the dissemination, a characteristic sample have a very sparse exemplification with respect to learned basis function. A sparse exemplification, if calculated extrinsically, is able to encrypt the graphical semantic data of the video frame image. In addition, in order to admirably employ video frame sparse exemplification to image processing and computer vision jobs, researchers needed to address a supplementary challenge of choosing the correct basis function for available data. This is dissimilar to the classical practice in video processing which assumes a basis function with good property. In image processing and computer vision, the program has to learn from given video frames a task-specific dictionary.

The challenge of discovering the sparse representation of a video frame in a form of dictionary can be expressed as follows:

For the $N \times M$ matrix B which has values of over-complete dictionary arranged in its columnar positions, satisfying the condition $M >> N$ for the signal $g \in R^N$, the aim of sparse representation aims at defining a solution as finding a $M \times 1$ matrix element vector x which obeys $g = Bx$ and subsequently $\|x\|_0$ is to be minimized using combinational optimization which is a NP hard problem.

$$x = \min \|\bar{x}\|_0 \text{ Such that } g = Bx \quad (1)$$

Here, the $\|\bar{x}\|_0$ is related to ℓ_0 norm and it is equal to non-zero matrix elements in vector x.

Since this is a NP hard problem, an approach to solve this problem is yielded by replacing ℓ_0 norm with ℓ_1 norm. Therefore the equation (1) is transformed to equation (2) as shown below.

$$x = \min \|\bar{x}\|_1 \text{ Such that } g = Bx \quad (2)$$

A generalized format of equation (2) that is used to compute x so that the mathematical objective functions is minimized:

$$G_1(x;\partial) = \|g - Bx\|_2^2 + \partial \|x\|_1 \quad (3)$$

Here, the factor $\partial > 0$. It is assumed that the signal g is resulting from following model:

$$g = Bx + \nu \quad (4)$$

Here ν is classical white Gaussian noise. The probability distribution of x is presented by equation (5).

$$p(x) \sim \exp\left(-\partial \sum_{j=1}^{M} |x_j|^p\right) \quad (5)$$

The MAP value of estimation is presented by equation (6).

$$x_{MAP} = \arg\min(\|g - Bx\|_2^2 + \partial \|x\|_p) \quad (6)$$

Equation (6) solves the complete dictionary representation problem.

B. Discrete Wavelet Tight Frames

Wavelet tight frames, or over-complete dictionary expansions, comes with a variety of striking structures. Using extracted frames, improved time-frequency localization is achieved than what is possible with bases. Some discrete wavelet tight frames are time invariant, but wavelet bases cannot be. Frames yield additional degrees of freedom to carry out experiment. Numerous applications have been benefited with the

ISSN (Online) 2394-6849

International Journal of Engineering Research in Electronics and Communication Engineering (IJERECE)
Vol 4, Issue 8, August 2017

use of tight frames. There are numerous techniques for generation of frames in real life scenarios. A discrete wavelet tight frame can be obtained by computing the union of minimum two bases. For this, two independent filter banks implemented in parallel are utilized. The Un-decimated DWT (UDWT) computes a discrete wavelet tight frame with the help of a prevailing wavelet basis by eradicating the subsampling from a prevailing analytically sampled filter bank. A discrete wavelet tight frame is computable by recapitulating an appropriately designed oversampled filter bank. The un-decimated DWT produces a strictly shift-invariant discrete transform. But, the UDWT builds upon an expansion factor of : it expands an -element matrix to elements. For real life scenario data, like videos, this matrix expansion yields its use difficult in practice. On the contrary, the dual-tree DWT, and the oversampled DWT, give an -sample matrix to elements liberated of scales for which the video decomposition is accomplished.

IV. PROPOSED APPROACH

In this paper, a dictionary learning based scheme is proposed which computes the basis function adaptively from the input image frames. Initially, the input video is decomposed into image frames. Unlike other approaches like wavelet or contourlet where the basis functions are constant. If the basis function is constant it is more likely that it will fail to capture the minute noise details from natural images. Therefore the basis function is computed from image itself using dictionary learning. Subsequently, a set of dictionary images is learnt using the basis function on the image. The dictionary learning method provides sparse representation of the video frame. Here hard thresholding algorithm is applied to compute the denoised frame. Fig. 1 depicts the proposed scheme.

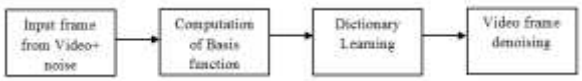

Fig 1: Proposed scheme for video denoising

Let us assume that denotes a video frame corrupted with noise and denotes noiseless video frame. The subsequent dictionary is computed which is denoted by which must be a set of real numbers. In the next step, an original video frame is split into parts and each contains pixels. By calculating this minimization as stated in equation (7), the video frame can be denoised.

$$\min_{d,O} \frac{1}{2}\|c-d\|_2^2 + \sum_u \mu_u \|c_u\|_0 - \partial \sum_u \|Oy_u - N_u d\|_2^2 \quad (7)$$

The convex optimization using this equation is computationally heavy. Thus it is further resolved using advanced mathematical approach. Few substitute iterative methods have been proposed in the recent past, which helps to obtain a crisp solution of the minimization problem. Instead of simultaneously minimizing d and O, if d and O get minimized in an iterative manner, the problem becomes simpler.

Let Ψ contains m matrix elements which is used to build the dictionary. Let the noisy input image be denoted with c. In order to construct dictionary using machine learning method, equation (8) is to be followed.

$$\min \|\ell - \Psi \bullet c\|_2^2 - \xi^2 \|\ell\|_0 \quad (8)$$

The main aim is to compute the elements of Ψ.

$$\ell^{(y)} := \min \left\|\ell - \Psi^{(y)} \bullet c\right\|_2^2 - \xi^2 \|\ell\|_0 \quad (9)$$

After solving equation (9) with Singular Vector Decomposition, the final tight frame Ψ is obtained.

Note that this method works in the intervals of 50 frames. Initially the dictionary is computed for first frame. The same dictionary is used for video frame denoising of the same frame and subsequent 49 frames. This process is repeated after each 50 frames.

V. RESULTS

For sigma=20 the noisy frames are as shown for the video gsalesman.avi in the following Fig.2:

Fig.2. Noisy frames from gsalesman sequence

ISSN (Online) 2394-6849

International Journal of Engineering Research in Electronics and Communication Engineering (IJERECE)
Vol 4, Issue 8, August 2017

The denoising using proposed approach gives following frames in Fig.3.

Fig.3. Denoised frames using proposed technique The experimentation is done on two popular video denoising sequences.

A. Experimentation on salesman video

The results are benchmarked as shown in table 1.

Table 1: Comparison of PSNR with other approaches

Sr.	Sigma	10	15	20	25	30	35	40
1	[8]	28.29	24.71	22.16	20.17	18.62	17.27	16.11
2	[9]	28.37	26.61	24.79	23.16	21.73	20.51	19.4
3	[10]	30.02	28.25	25.84	23.58	21.68	20.28	19.01
4	[11]	30.03	28.88	27.72	26.56	25.5	24.47	23.62
5	Proposed	33.55	31.29	29.74	28.53	27.50	26.62	25.83

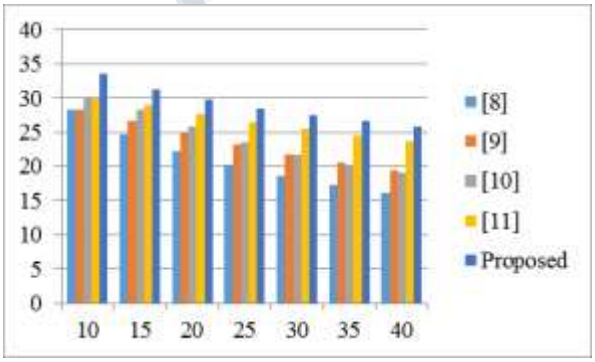

Fig.4. Performance comparison with other approaches

B. Experimentation on miss america video

The results are benchmarked as shown in table 2.

Table 2: Comparison of PSNR with other approaches

Sr. No.	Sigma	10	15	20	25	30	35	40
1	[8]	28.22	24.66	22.11	20.15	18.6	17.26	16.09
2	[9]	26.72	25.26	23.75	22.38	21.18	20.16	19.21
3	[10]	29.73	28.06	25.82	23.76	22.03	20.62	19.4
4	[11]	29.74	28.79	27.65	26.52	25.5	24.55	23.69
5	Proposed	38.27	36.52	35.23	34.16	33.13	32.11	31.03

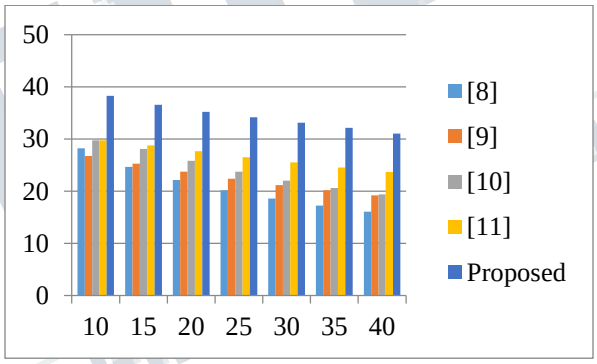

Fig.5. Performance comparison with other approaches

CONCLUSION

In this paper, a dictionary learning based scheme is proposed which computes the basis function adaptively from the input image frames. Unlike other approaches like wavelet or contourlet where the basis functions are constant. If the basis function is constant it is more likely that it will fail to capture the minute noise details from natural images. Therefore the basis function is learnt from image itself. Subsequently a set of dictionary images is learnt using the basis function on the image. The dictionary learning method provides sparse representation of the image. Here hard thresholding algorithm is applied to compute the denoised frame. The frames are recollected together to form the video again. The proposed technique achieves competitive performance on the standard video sequences.

REFERENCES

[1] Buades, A., Coll, B. and Morel, J.M., 2005, June. A non-local algorithm for image denoising. In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on (Vol. 2, pp. 60-65). IEEE.

[2] A. Buades, J. L. Lisani, and M. Miladinović, "Patch-Based Video Denoising With Optical Flow Estimation," IEEE Transactions on Image Processing, vol. 25, pp. 2573-2586, 2016.

[3] G. Chen, J. Zhang, D. Li, and H. Chen , "The Robust Kronwecker product video denoising based on fractional-order total variation model," Signal Processing, vol. 119, pp. 1-20, 2016.

[4] R. Almahdi and R. C. Hardie, "Recursive non-local means filter for video denoising with Poisson-Gaussian noise," in 2016 IEEE National Aerospace and Electronics Conference (NAECON) and Ohio Innovation Summit (OIS), 2016, pp. 318-322.

[5] R. A. Ali and R. C. Hardie, "Recursive non-local means filter for video denoising," EURASIP Journal on Image and Video Processing, vol. 2017, p. 29, 2017.

[6] M. Khalid, P. S. Sethu, and R. Sethunadh, "Optimised surfacelet transform based approach for video denoising," in 2016 International Conference on Inventive Computation Technologies (ICICT), 2016, pp. 1-5.

[7] H. Guo and N. Vaswani, "Video denoising via online sparse and low-rank matrix decomposition," in 2016 IEEE Statistical Signal Processing Workshop (SSP), 2016, pp. 1-5.

[8] J.-S. Lee, "Digital image smoothing and the sigma filter," Computer Vision, Graphics & Image Processing, vol. 24, no. 2, pp. 255–269, 1983.

[9] D. Zhang, J.-W. Han, O.-J. Kwon, H.-M. Nam, and S.-J. Ko, "A saliency based noise reduction method for digital TV," in Proceedings of the IEEE International Conference on Consumer Electronics (ICCE '11), pp. 743–744, Las Vegas, Nev, USA, January 2011.

[10] A. A. Yahya, J. Tan, and L. Li, "An amalgam method based on anisotropic diffusion and temporal filtering for video denoising," Journal of Computational Information Systems, vol. 11, no. 17, pp. 6467–6475, 2015.

[11] Ali Abdullah Yahya, Jieqing Tan, and Lian Li, "Video Noise Reduction Method Using Adaptive Spatial-Temporal Filtering," Discrete Dynamics in Nature and Society, vol. 2015, Article ID 351763, 10 pages, 2015. doi:10.1155/2015/351763