Frame By Frame Digital Video Denoising Using Multiplicative Noise Model

M. Morshed, M. M. Nabi, N. B. Monzur

Department of Electrical and Electronic Engineering, Ahsanullah University of Science and Technology, 141-142, Love Road, Tejgaon I/A, Dhaka 1208, Bangladesh.
Email: monjurm@aust.edu

ABSTRACT:The sparse representations of images have achieved outstanding demising results in recent days. But noise reduction in digital videos remains a challenging problem. In this communication we considered the coherent nature of the video frames for image processing. The imaging model shows that the video frames are corrupted by multiplicative noise. Simulation results carried out on artificially corrupted videos' frames and demonstrated performances of five previously available filtering approaches.

Keywords : Digital Video Denoising; Multiplicative Noise Models; Peak Signal to Noise Ratio; Synthetic Aperture Radar; Structural Similarity Index etc.

1 INTRODUCTION

VIDEO signals are considered as a sequence of two-dimensional images, projected from a dynamic three-dimensional scene onto the image plane of a video camera. Luminance and chrominance are two attributes that describe the color sensation in a video sequence of a human being. Luminance refers to the perceived brightness of the light, while chrominance corresponds to the color tone of the light. Numerous still images and video denoising algorithms have been developed to enhance the quality of the signals over the last few decades [1]. Many of the algorithms are based on probability theory, statistics, partial differential equations, linear and nonlinear filtering, spectral and multiresolution analysis. However, image denoising can be extended to a video by applying it to each video frame independently. Depending on various signal-processing problems various algorithms have been proposed mainly for image denoising [2]. A human observer cannot resolve fine details within any image due to the presence of speckling. The available techniques are mostly based on noise suppression techniques in the post-image formation type; use computer simulation to suppress the signal as well as its speckles. The property of image sparsity is an important key to denoise image and video signals as well. Sparsity also resides in videos. Most videos are temporally consistent; a new frame can be well predicted from previous frames. The idea of combining multiple images to get a desired one is called image fusion and can be used to produce a denoised video. Video signals are often corrupted by additive noise or motion blur. Often, the noise can be modeled effectively as a Gaussian random process independent of the signal. Although the state of the art video denoising algorithms often satisfy the temporal coherence criterion in removing additive white Gaussian noise (AWGN) [3]-[6]. Normally all coherent imaging processes, such as, synthetic aperture radar (SAR) and narrow-band ultrasound suffer from speckle noise. The SAR images are available in two formats. One is amplitude format and the other one is in intensity format. The magnitude of the speckles follows the Rayleigh probability distribution and corresponding phases follow uniform distribution [7]. The speckle intensity is described by a negative exponential distribution. By multilook averaging the undesirable effects of speckle in a SAR image can be easily reduced. For the case of intensity data the statistical distribution of the resultant speckle in the degraded resolution image is given by the Gamma distribution and in case of amplitude data multiconvolution of the Rayleigh

probability density is considered. Speckle in SAR images is generally modeled as multiplicative random noise [8], whereas most available filtering algorithms were developed for additive white Gaussian noise (AWGN) in the context of image denoising and restoration, as additive noise is most common in imaging and sensing systems.
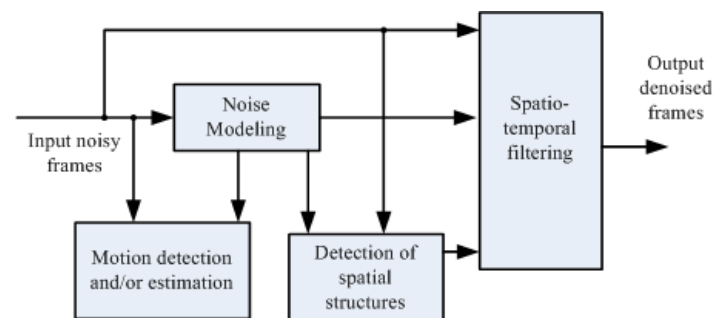

Fig. 1. General video denoising framework.

Video denoising is normally done with some linear or non-linear operation on a set of neighboring pixels and the correlation between those pixels available in spatio-temporal sense. The best video denoising can be achieved by exploiting information from both future and past frames. But this leads to an additional delay of at least one frame which is undesirable in some real-time applications. For this reason, many algorithms exploit information from usually the current frame and one or two previous frames. In image frame denoising algorithms focus to find the best compromise between noise removal and preservation of important denoised image frames. Here each frame is independently processed. That is why for optimal filter performance the spatio-temporal properties of the processed noisy image frames are taken into consideration. The general framework for video denoising is illustrated in Fig. 1. An accurate modeling of noise is necessary in order to estimate noise-free spatio-temporal sequence structures. To distinguish between the noise and the noise-free spatio-temporal correlations in the image frames the information concerning the noise and the noisy input frames are combined together. In this way, the spatio-temporal structures can be estimated in a noise-insensitive manner and consequently enable an efficient noise removal with the preservation of all the important spatio-temporal sequence features [9]. In this communication, we considered multiplicative noise models for video signals

denoising. We applied the same concept for denoising the video signals as was proposed for SAR image denoising. This paper is organized as follows. In the next sections the available literatures on multiplicative noise models have been reviewed. Section II provides a brief review of the filtering approach and describes the expressions for pdf of the noise models. An adaptive filtering recipe and its implementation details are described in Section III. Finally, simulation results to assess the effectiveness among the methods are presented in Section IV, whereas some conclusions are drawn in Section V.

2 LITERATURE REVIEW AND THEORIES

Bilateral filter is a widely used nonlinear filter [10]-[12] introduced by Tomasi and Manduchi [10] which smoothes an image while preserving the edges. This filter replaces each pixel by weighted mean of nearby pixels considering both geometrical and photometrical distance using the domain and range filters. However, in [10], the corresponding domain and range filter parameters are set rather arbitrarily from image to image at different noise levels. For SAR image despeckling with bilateral filter, a limited number of approaches have been reported in the literature for adapting the parameters [11]- [12]. Adaptation of the parameters has been performed in [12] based on Equivalent Number of Looks (ENL) and the Edge Save Index (ESI). But the method takes the parameters in a nonadaptive and approximate basis in the iterative bilateral filtering steps and is also time consuming requiring ten iterations of filtering for each parameter estimation in the initial step. The method in [10] adapts the domain parameter using local Coefficient of Variation. However, it accounts photometrical similarity from a joint probability density function based model rather than utilizing the conventional range filter of the bilateral filtering framework. Hence, adaptation of range parameter is disregarded in [13]. However, the range parameter is more sensitive to noise compared to the domain parameter and thus entails its adaptation with underlying noise [11]. Recently, a bilateral filter is introduced to despeckle SAR images with an adaptive estimation of both filter parameters. The range parameter is tuned according to estimated noise level in SAR image implementing the Intensity Homogeneity Measurements based noise estimation method of [14]. Furthermore, domain parameter (σ_d) has been adjusted according to Coefficient of Variation that takes into account the local homogeneity of center to neighbouring pixels [13]. An iterative scheme is employed with adaptively estimated parameters in each iteration and the continuation of this process is determined from comparative difference in Structural Similarity Index (SSIM) [11]. Extensive simulations have been carried out to investigate their effectiveness and results have been reported for the noisy video signals' frames. The results are compared among five available multiplicative natured noise models filtering methods.

2.1 Intensity Format

In coherent imaging system the images use electromagnetic waves which are backscattered from the targeted object. Considering this, the frames are considered in two formats. One is in intensity format and the other one is in amplitude format. Let Y, X and N denote the image intensity, backscattering coefficient and normalized fading random variable in the noisy video frame respectively. The noise corrupted image can be considered as

$$Y = X \times N \qquad (1)$$

where Y is in intensity format.

The pdf of X is given by [15]

$$P_X(X) = \frac{L^L X^{L-1} e^{-LX}}{\Gamma(L)} , \quad X \geq 0, \ L \geq 1 \qquad (2)$$

where L denotes look (independent pixels) and $\Gamma(\cdot)$ denotes gamma function. After taking natural logarithmic transformation (1) becomes

$$y = x + n \qquad (3)$$

where $y = \ln(Y)$, $x = \ln(X)$ and $n = \ln(N)$

The transformation leads to a new pdf [15]-

$$P_x(x) = \frac{L^L e^{-xL} e^{-Le^x}}{\Gamma(L)} \qquad (4)$$

The mean and variance of $P_x(x)$ are given by

$$E_x(x) = \psi(L) - \ln(L) \qquad (5)$$

and

$$\mathrm{var}(x) = \psi(1, L) \qquad (6)$$

where $\psi(\cdot)$ is digamma function and $\psi(1, L)$ is the first order polygamma function of L. Besides this, we could also presume these frames as L look images by taking square root operation. The pdf of which are also given here for the readers but this was beyond the scope of our focus and could be found elsewhere later on. It is clearly demonstrated that for intensity data log-transformed noise approaches Gaussian pdf faster than that of the original speckle [16].

$$P_X(X) = \frac{2X^{2L-1} L^L e^{-LX^2}}{\Gamma(L)} , \ X \geq 0 \qquad (7)$$

The mean and variance are [15]

$$E_X(X) = \frac{\Gamma(L+\frac{1}{2})}{\sqrt{L}\Gamma(L)} \quad \mathrm{var}(X) = 1 - E^2(X) , \qquad (8)$$

After logarithmic the pdf, mean and variance become

$$P_x(x) = \frac{2L^L e^{-2Lx}}{\Gamma(L)} \exp\left(-Le^{2x}\right) \qquad (9)$$

$$E_x(x) = \frac{1}{2}\left(\psi(L) - \ln(L)\right), \ \mathrm{var}(x) = \frac{1}{4}\psi(1, L) \qquad (10)$$

2.2 Amplitude Format

In amplitude format the multiplicative model is also represented by (1). For a single look image the normalized Rayleigh distributed random variable is [15]

$$P_X(X) = \frac{\pi X}{2} \exp\left(-\frac{\pi X^2}{4}\right), \quad X \ge 0, \ L = 1 \quad (11)$$

The mean and the variance are [15]

$$E_X(X) = 1, \ \mathrm{var}(X) = \frac{4}{\pi} - 1 \quad (12)$$

After taking logarithmic transform the statistics becomes Fisher-Tippet density function and is represented by-

$$P_x(x) = \frac{\pi e^{2x}}{2} \exp\left(-\frac{\pi e^{2x}}{4}\right), \quad L = 1 \quad (13)$$

For the case of multilook the mean will be always be unity and the variance will be decreased by a factor of L.

3 ADAPTIVE BILATERAL FILTERING OF THE VIDEO FRAMES

The bilateral filter replaces every pixel by a weighted sum of neighboring pixels. The weights depend on both the spatial distance and the photometric similarity. The output of the bilateral filter is expressed as

$$I_d(i) = \frac{\displaystyle\sum_{i\in[-w,w]} \sum_{i\in[-w,w]} I(j) e^{-\left[\frac{\|i-j\|\times|I(i)-I(j)|}{2\sigma_d\sigma_r}\right]^2}}{\displaystyle\sum_{i\in[-w,w]} \sum_{i\in[-w,w]} e^{-\left[\frac{\|i-j\|\times|I(i)-I(j)|}{2\sigma_d\sigma_r}\right]^2}} \quad (14)$$

where w is the span of the filter, indicating a window of size (2w+1)×(2w+1), and I_d, I,σ_d,σ_r denote the denoised frames, noisy frames, domain and range parameters respectively. Among the two filtering parameters, σ_d is relatively insensitive to noise variations [12]. As the main purpose of denoising is to reduce noise and σ_r is more sensitive to noise variations, it is adaptively estimated in this method. Let S_{kl} denote a W× W-sized block centered on the (k, l)-th pixel. The homogeneity of the block is measured along eight directions as shown in Fig. 2. For a given direction, a weighted sum of the corresponding pixels gives the homogeneity measure in that direction, the weights assigned for a block being
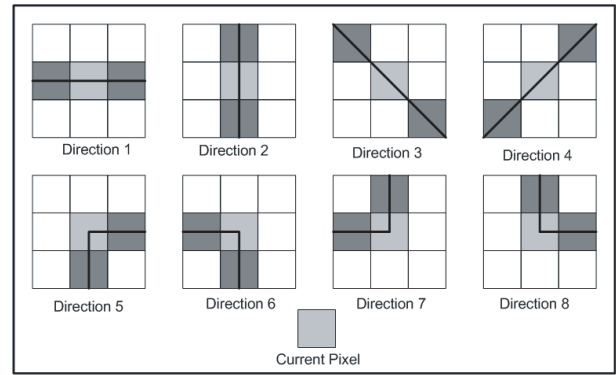
$$\{-1,-1,..,(W-1),..,-1,-1\}.$$

Fig. 2. Directions of the homogeneity measurement for a 3×3 block.

Assuming the pixels in a block to be independent and identically distributed (i.i.d.), the corresponding mean and variance, denoted by μ_{kl} and σ_{kl}^2, respectively, are computed as

$$\mu_{kl} = \frac{\displaystyle\sum_i^{W-1}\sum_j^{W-1} I(i,j)}{W^2} \quad (15)$$

$$\sigma_{kl}^2 = \frac{\displaystyle\sum_i^{W-1}\sum_j^{W-1} (I(i,j) - \mu_{kl})^2}{W^2} \quad (16)$$

where I(i, j) represents the (i, j)-th pixel of a noisy frame. Since the domain parameter is related to spatial closeness of pixels, it is adapted according to Coefficient of Variation (Cov). σ_d is related to Cov according to the following formula [13]

$$\sigma_d(Cov) = \frac{FA}{1 + e^{K_d(Cov - C_d)}} \quad (17)$$

Where

$$Cov(i) = \frac{\sqrt{V(i)}}{M(i)}, \ C_d = \frac{0.714}{\sqrt{L}}, \ K_d = \frac{\ln\left(\frac{W-1}{2}\right)}{0.7\sqrt{L}} \quad (18)$$

where V(i) and M(i) are the variance and mean of pixels in the local window, A, K_d, and C_d are parameters determining the upper limit, the decay speed, and the symmetry center of σ_d as a function of Cov , respectively with a factor F=0.3. The bilateral filtering operation is performed in an iterative manner in this method. The purpose of having iteration is to reduce the residual noise through repeated filtering. After each iteration, the parameters are estimated from the resultant denoised frame. As repeated filtering may lead to the removal of frame structures with noise, using the percentage difference in the structural similarity index (SSIM) [17] obtained at the n-th and n+1-th iteration, a stopping criterion is defined as

$$\frac{S(n+1) - S(n)}{S(n)} \times 100 < T \quad (19)$$

where S represents the SSIM [17], T represents threshold and n the index of iteration. When the percentage difference be-

tween the SSIM of the noisy frame from the two consecutive iterations reaches at the threshold T = 5, the two filtered frames are nearly similar. Hence, the filtering process is stopped since further filtering may lead to blurring of the edges, removal of frame structure with noise etc.

4 SIMULATION RESULTS

Three Additive White Gaussian Noise (AWGN) corrupted videos namely, gflowersg15.avi (flower garden), gsalesmang15.avi (salesman) and gstennisg15.avi (tennis player [18] are used for extensive simulations. Each one contains 50 frames. The pdf of the noise models are considered in amplitude format. Simulations were carried out by considering multiplicative noise nature as are used for denoising the speckle noises of the SAR images. For the Bayes-shrink [13], five level decomposition is carried out with the Daubechies' wavelet of order 8. The ENL and ESL based adaptive bilateral in [19] was initialized with $\sigma_d = 3$ and $\sigma_r = 0.05$ with a step size of 0.05. Table I, II and III show the average value of PSNR(dB), structural similarity index (SSIM) and edge preservation index (β) of the video frames. Although from a PSNR point of view our method is not always significantly better than the other methods, it invariably performs best visually as can be seen in Figs. 4, 6 and 8. Three noisy frames are shown in Fig. 3, 5 and 7 to compare with the corresponding denoised frames. Amongst these filters Frost performs the worst in terms of visual quality in terms of spatio-temporal blur. The visual quality of Bayes-Shrink, Bilateral, Adaptive Bilateral and ENL Bilateral is nearly same. The standard deviation of PSNR, SSIM and β of the methods are obtained as 1.42, 0.08 and 0.046 respectively except the Frost filter.

TABLE I. FLOWER GARDEN (gflowersg15.avi)

Method	PSNR(dB)	SSIM	β
Bayes-Shrink	25.51141	0.800954	0.898216
Bilateral	25.62475	0.788338	0.936662
Adaptive Bilateral	25.55384	0.764352	0.905316
Frost	16.9392	0.3895	0.41486
ENL Bilateral	22.7178	0.6274	0.828694

TABLE II. SALESMAN (gsalesmang15.avi)

Method	PSNR(dB)	SSIM	β
Bayes-Shrink	24.91301	0.67457	0.5962
Bilateral	25.88039	0.653032	0.833696
Adaptive Bilateral	25.03207	0.641336	0.674246
Frost	22.15655	0.338266	0.218848
ENL Bilateral	25.70528	0.604316	0.65549

TABLE III. TENNIS PLAYER (gstennisg15.avi)

Method	PSNR(dB)	SSIM	β
Bayes-Shrink	24.0291	0.522752	0.627794
Bilateral	25.20242	0.601848	0.814058
Adaptive Bilateral	24.13672	0.50076	0.65928
Frost	20.68966	0.200206	0.219756
ENL Bilateral	24.52833	0.509458	0.686952

5 CONCLUSION

For high-quality video denoising a real and structured noise model is essential. Simulation results show that the systems can play significant role with the state of the art in removing AWGN and structured noise. Robust motion estimation is essential for high-quality video denoising which be area for further investigation along with multiplicative noise model. In this paper, an adaptive bilateral filter has been introduced for video frames denoising. The method suppresses speckle noise well, while retaining the structure and edges of the frames. The domain parameter is adjusted according to Coefficient of Variance of local neighborhood. The range parameter is adapted with present spackle noise in the frames which are estimated from Intensity Homogeneity Measurements of the frames. The method outperforms all the other methods regarding objective and subjective criteria such as, PSNR, SSIM, edge preservation index as well as in terms of visual quality of the frames.

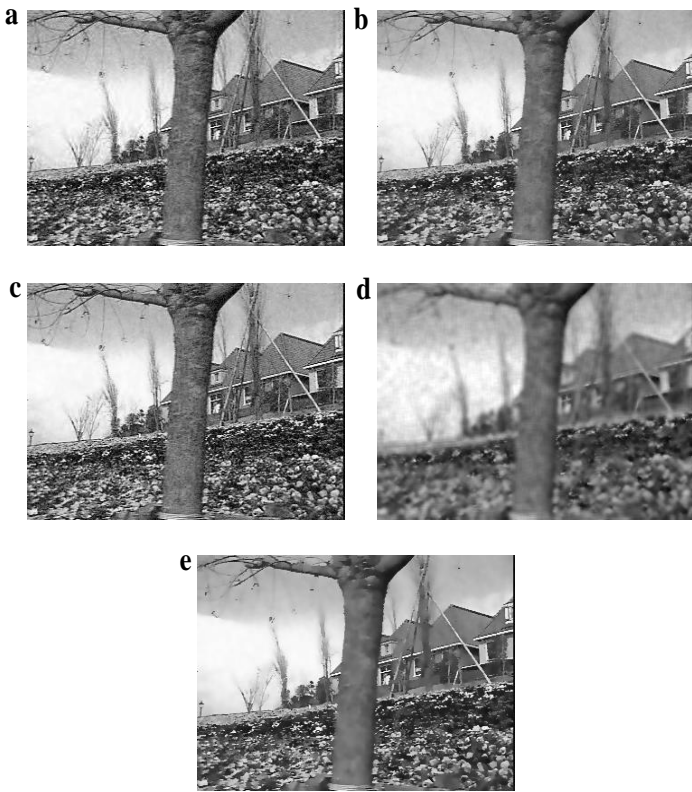**Fig. 3**. Noisy frame of Flower Garden (10-th frame).

Fig. 4. Denoised frame of Flower Garden (10-th frame). (a) Bayes-Shrink (b) Bilateral (c) Adaptive Bilateral (d) Frost (e) ENL Bilateral.
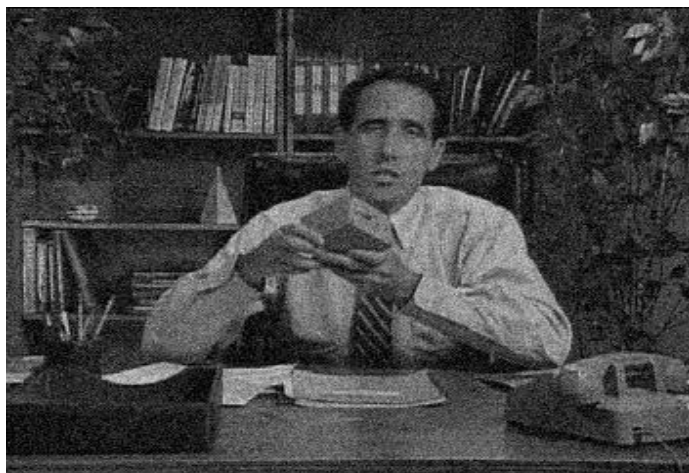

Fig. 6. Denoised frame of Salesman (10-th frame). (a) Bayes-Shrink (b) Bilateral (c) Adaptive Bilateral (d) Frost (e) ENL Bilateral.

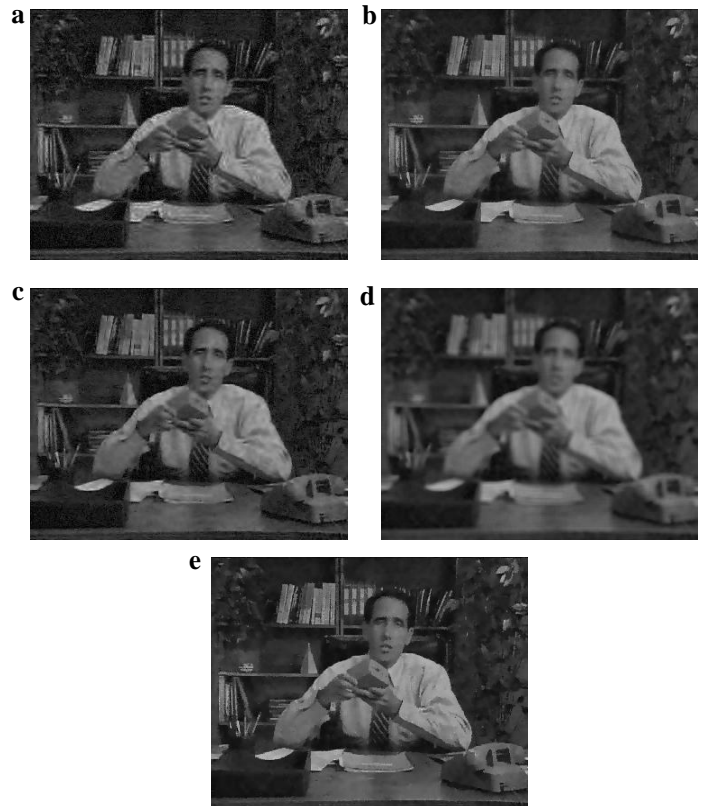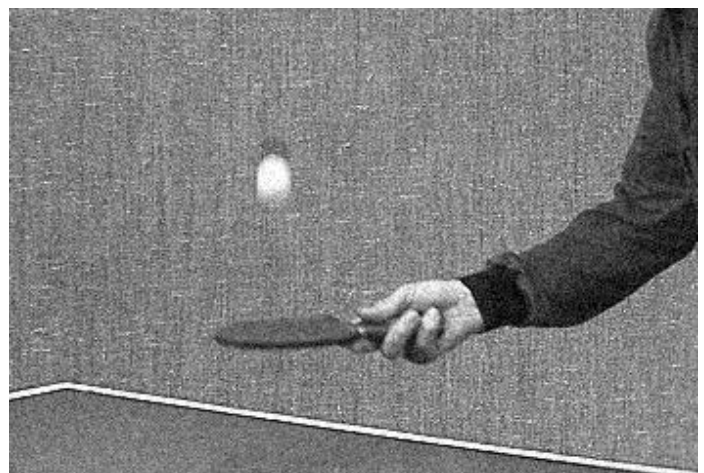

Fig. 5. Noisy frame of Salesman (10-th frame).
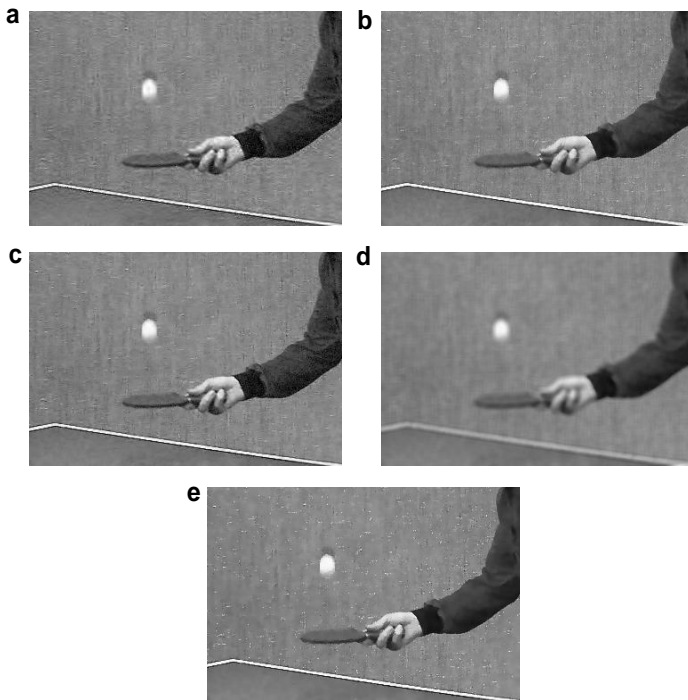

Fig. 7. Noisy frame of Tennis Player (10-th frame).

Fig.8. Denoised frame of Tennis Player (10-th frame). (a) Bayes-Shrink (b) Bilateral (c) Adaptive Bilateral (d) Frost (e) ENL Bilateral.

.

REFERENCES
[1] C. Liu and W. T. Freeman, "A High-Quality Video Denoising Algorithm Based on Reliable Motion Estimation." Computer Vision – ECCV 2010. Ed. Kostas Daniilidis, Petros Maragos, & Nikos Paragios. LNCS Vol. 6313. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp.706–719.

[2] P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using generalized Gaussian and complexity priors," IEEE Trans. on Info. Theory, Vol. 45, No. 3, 1999, pp.909-919.

[3] M. Maggioni, G. Boracchi, A. Foi and K. Egiazarian, "Video Denoising, Deblocking, and Enhancement Through Separable 4-D Nonlocal Spatiotemporal Transforms," , IEEE Trans. on Image Process., Vol. 21, No. 9, 2012, pp.3952 - 3966.

[4] A. Amer and E. Dubois, "Fast and reliable structure-oriented video noise estimation," IEEE Trans. on Circuits and Systems for Video Technology, Vol. 15, No. 1, 2005, pp.113–118.

[5] H. Cheong, A. Tourapis, J. Llach and J. Boyce, "Adaptive spatio-temporal filtering for video de-noising," IEEE, Inter. Conf. on Image Process., 2004, pp.965–968.

[6] G. de Haan, P. Biezen, H. Huijgen, and O. Ojo, "True-motion estimation with 3-d recursive search block matching," IEEE Trans. on Circuits and Systems for Video Technology, Vol. 3, No. 5, 1993, pp.368–379

.
[7] J. S. Lee, "Speckle analysis and smoothing of synthetic aperture radar images," Comput. Graph. Image Process., Vol-17, 1981, pp.24-32.

[8] J. A. Stiles, K. S. Shanumugan, and J. C. Holtzman, "A model for radar images and its application to adaptive digital filtering of multiplicative noise," IEEE Trans. Pattern Anal. Mach. Intell., Vol. 4, No. 2, 1986, pp. 157-165.

[9] V. Zlokolica, "Advanced Nonlinear Methods for Video Denoising," PhD Thesis, Faculteit Ingenieurswetenschappen Academiejaar, 2005-2006, Universiteit Gent.

[10] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," IEEE Inter. Conf. on Comp. Vision, 1998, pp.839–846.

[11] M. Zhang and B. K. Gunturk, "Multiresolution bilateral filtering for image denoising," IEEE Trans. on Image Process., Vol. 17, No. 12, 2008, pp. 2324–2332.

[12] W. G. Zhang, F. Liu and L. C. Jiao, "SAR image despeckling via bilateral filtering," IET Electronics Letters, Vol. 45, No. 15, 2009, pp.781-783.

[13] G. Ting Li, C. Le Wang, P. Ping Huang and W. Dong Yu, "SAR Image Despeckling Using a Space-Domain Filter With Alterable Window", IEEE Geo. and Remote Sensing Let., Vol. 10, No. 2, 2012, pp.263-267.

[14] A. Amer and E. Dubois, "Fast and Reliable Structure-Oriented Video Noise Estimation," IEEE Trans. on Circuits and Systems for Video Tech., Vol. 15, No. 1, 2005, pp.113-118.

[15] F. Ulaby and M. C. Dobson, "Handbook of Radar Scattering Statistics for Terrain. Norwood," MA: Artech House, 1989.

[16] H. Xie, L. E. Pierce, and F. T. Ulaby, "Statistical Properties of Logarithmically Transformed Speckle," IEEE Trans. on Geo. and remote sensing, Vol. 40, No. 3, 2002, pp.721-727.

[17] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli,"Image quality assessment: from error visibility to structural similarity," IEEE Trans. on Image Process., Vol. 13, No. 4, 2004, pp. 600-612.

[18] http://telin.ugent.be/~vzlokoli/Results_J/noisy/

[19] S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for Image Denoising And Compression," IEEE Trans. on Image Process., Vol. 9, No. 9, 2000, pp.1532-1546.