IJSTE - International Journal of Science Technology & Engineering | Volume 3 | Issue 10 | April 2017
ISSN (online): 2349-784X

An Efficient Video Denoising using Patch-based Method, Optical Flow Estimation and Multiresolution Bilateral Filter

Rakhi.S
PG Student
Department of Electronics & Communication Engineering
Marian Engineering College, Trivandrum, India

Dr. M Sasikumar
Professor
Department of Electronics & Communication Engineering
Marian Engineering College, Trivandrum, India

Abstract

Video denoising is a longstanding area of research. The core problem is to remove noise without affecting image details. The paper presents a new algorithm which combines motion compensation, patch based method and multiresolution bilateral filter for denoising. The proposed approach takes advantage of the self-similarity and redundancy of adjacent frames. Motion compensation by optical flow method permits the use of spatio-temporal patches for a more robust comparison. The use of singular value decomposition (SVD) for patches preserves fine details such as edges and texture. Multiresolution bilateral filter framework, which integrates both bilateral filter and soft thresholding, has the potential of eliminating low-frequency noise components. Experimental results shows that the proposed method eliminate most of the artifacts due to noise and improves texture and detail preservation.
Keywords: Bilateral filter, motion compensation, optical flow, patch processing, singular value decomposition (SVD), video denoising, wavelet thresholding

I. INTRODUCTION

Digital video is becoming ubiquitous with widespread applications in information technology, telecommunications, consumer electronics and entertainment. TV and cinema have gone all-digital and high definition, and most movies and some TV broadcasts are now in 3D format. Digital video brings broadcasting, cinema, computers, and communications industries together in a truly revolutionary manner, where telephone, cable TV, and Internet service providers have become fierce competitors. A single device can serve as a personal computer, a high-definition TV, and a videophone. We can now capture live video on a mobile device, apply digital processing on a laptop or tablet, and/or print still frames at a local printer.

Video signals are subject to noise contaminations during acquisition and transmission. Removal of noise is one of the preprocessing tasks in several video processing techniques. Techniques for noise removal in digital images [1]comprise transform thresholding, local averaging, patch based methods and variational techniques. The sliding window DCT and wavelet thresholding are the main examples of thresholding methods. These methods decompose the original data in a predefined basis and cancel coefficients under a certain threshold related to noise statistics. Anisotropic filtering and bilateral filter or neighborhood filtering aim at averaging close pixels belonging to the same object, thus reducing the noise amplitude and preserving the main object boundaries. Variational techniques share the same objective but use a variational framework, the total variation minimization being the main example. NL-means introduced patch based methods into image denoising. The algorithm groups similar patches all over the image and averages them in order to reduce noise. The method is able to preserve texture and fine details additionally to the main boundaries of the image.

In video sequences, motion is a key source of information. Motion arises due to moving objects in the 3D scene, as well as camera motion. Apparent motion, also known as optical flow [2], captures the resulting spatio-temporal variations of pixel intensities in successive images of a sequence. The purpose of motion estimation techniques is to recover this information by analyzing the image content. Efficient and accurate motion estimation is an essential component in the domains of image sequence analysis, computer vision and video communication.

Noise is in general space varying and channel dependent. Blue channel is typically the noisiest channel due to the low transmittance of blue filters. An often neglected characteristic of image noise is the spatial frequency. Noise may have low frequency (coarse-grain) and high-frequency (fine-grain) fluctuations. High-frequency noise is relatively easier to remove; on the other hand, it is difficult to distinguish between real signal and low-frequency noise. Many denoising methods have been developed over the years; among these methods, wavelet thresholding [3] is one of the most popular approaches. In wavelet decomposition, a signal is decomposed into its approximation (low-frequency) and detail (high-frequency) sub bands; since most of the image information is concentrated in a few large coefficients, the detail sub bands are processed with hard or soft thresholding operations. The critical task in wavelet thresholding is the threshold selection. Various threshold selection strategies have been proposed, for example, VisuShrink, Sure Shrink, and BayesShrink, etc.

A new algorithm is proposed which make use of motion estimation, patch based methods[1] and multiresolution bilateral filter framework[17] for denoising which not only improves the denoising performance but also preserves important detail features such as edges and textures, and helps eliminates the coarse-grain noises.

II. Related work

Nowadays, state-of-the-art methods actually combine two or three of these techniques. For example variational and patch based techniques were combined for both denoising and deblurring or image segmentation. BM3D [5] combined patch based grouping and thresholding methods, using a 3D DCT transform. In [6] the authors introduced sparse representation in a redundant dictionary for denoising purposes. In [7] the authors explored the use of multi-scale combined with this dictionary learning techniques. Several methods appeared combining the grouping of similar patches and the learning of an adapted basis via PCA or SVD decomposition [8]. State-of-the-art results are obtained using Gaussian models for the group of similar patches or adapting the shape of the patch before learning a PCA model.

Local average methods, as the bilateral filter [9], or patch based methods as NL-means[10] or BM3D [5]and NLBayes [11] can be easily adapted to video just by extending the neighboring area to the adjacent frames. The performance of local average methods is improved by introducing motion compensation. Boulanger et al. [12] extended NL-means to video by growing adaptively the spatio-temporal neighborhood .Xu et al. [13] separated the temporal from the spatial filtering using NL-means, and then combined them using a motion indicator. In [14] the authors make use of an external database for still image denoising. The most similar images are retrieved from an external database and used for patch-based denoising which is combined with an internal denoising stage inspired by BM3D.

Mairal et al. [15] learnt multi-scale sparse representations for video restoration. VBM4D, the state-of-the-art for white noise removal in image sequences, exploits the mutual similarity between 3D spatio-temporal volumes constructed by tracking blocks along trajectories defined by the motion vectors. Liu and Freeman [16] also use motion vectors and group patches across adjacent frames but in a different manner. Instead of comparing patches to the reference patch, these are compared in each frame with the compensated patch of the reference one. NL-means is applied to this group of collected patches. The proposed algorithm adaptively computes the noise model of the sequence, which is an important issue for real applications.

Although the bilateral filter was first proposed as an intuitive tool, recent papers have pointed out the connections with some well-established techniques. In [19], it is shown that the bilateral filter is identical to the first iteration of the Jacobi algorithm (diagonal normalized steepest descent) with a specific cost function. The bilateral filter can also be viewed as an Euclidean approximation of the Beltrami flow, which produces a spectrum of image enhancement algorithms ranging from the linear diffusion to the nonlinear flows. Buades et al [20] proposes a nonlocal means filter, where similarity of local patches is used in determining the pixel weights. When the patch size is reduced to one pixel, the nonlocal means filter becomes equivalent to the bilateral filter. [21] extends the work of [20] by controlling the neighborhood of each pixel adaptively.

III. Proposed method

First, the whole patch is denoised and not only the central pixel, which permits an increase in the noise reduction by taking the average of all estimates per pixel (aggregation). Second, the denoised image is used as guide ("oracle") for a second iteration. The similarity between two patches is computed in the first denoised image, and the transformed coefficients are used to drive the thresholding in the second iteration. Third, the output of first stage is given to multiresolution bilateral filtering framework which combines bilateral filtering and soft thresholding. The block diagram of proposed system is shown in Fig.1

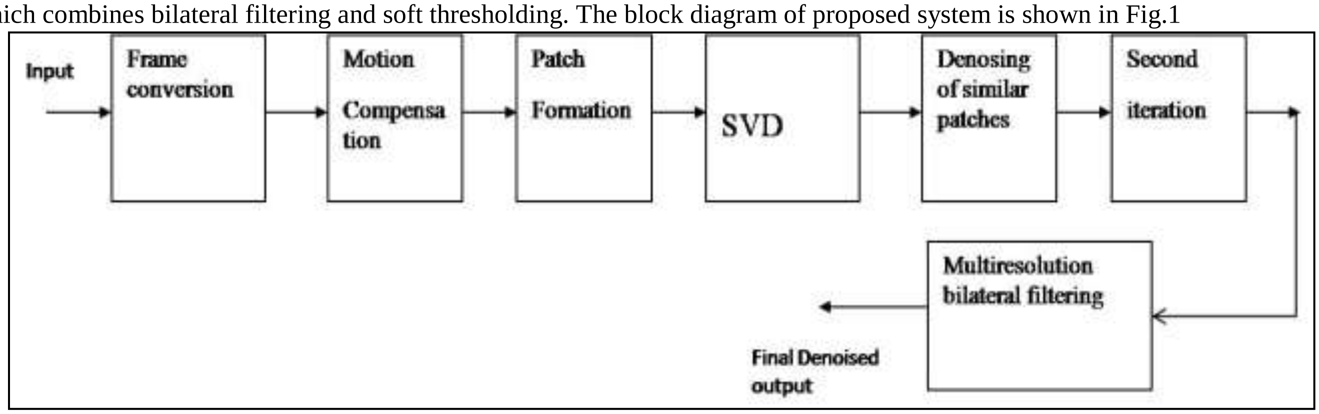

Fig. 1: Block diagram of proposed algorithm

A. Motion Compensation by Optical Flow method

Optical flow is the pattern of apparent motion of image objects between two consecutive frames caused by the movement of object or camera [4]. It is 2D vector field where each vector is a displacement vector showing the movement of points from first frame to second. The motion estimation used in the proposed method relies on the optical flow constraint (OFC), that is, we assume that the pixel intensity remains unchanged along a motion trajectory. This assumption is known as the brightness constancy constraint. In other words, the variations in time of the pixel intensity are due to the displacements of different objects present in the scene. The brightness constancy constraint implies that the illumination is uniform and the scene is Lambertian. The optical flow is used to warp adjacent frames and not only for compensating neighborhoods. Thus, the sub pixel accuracy improves the patch comparison and averaging.

First, the optical flow between I_k and adjacent frames in a temporal neighborhood is computed and used for warping these frames onto I_k . If registration was accurate and the sequence free of occlusions, a temporal average in this aligned data would be optimal, even if the noise reduction would slowly decrease as 1/M, where M is the number of adjacent frames involved in the process. Generally, this will not be the case, inaccuracies and errors in the computed flow and the presence of occlusions make this temporal average likely to blur the sequence and have artifacts near occlusions. The proposed approach tends to solve these limitations. Occlusions are detected depending on the divergence of the computed flow: negative divergence values indicate occlusions [17].

B. Choice of Similar Patches

Let $\{ I_{K-1}^{W}, \cdots, I_{K+1}^{W} \}$ be the set of adjacent frames to I_K after warping with the computed optical flow, and let M_j be the occlusion mask between frames I_K and I_K^W , j ∈ {k − t, ⋯ , k + t }. The algorithm uses a 3D volumetric approach to search for similar patches, while still 2D image patches are used for denoising. For each n × n patch P of the reference frame I_K , we consider the patch P referring to its extension to the temporal dimension, having M times more pixels than the original one (assuming M patches in the temporal neighborhood, M = 2t + 1), P = (P_{K-1} , ⋯ , P_{K+1}). The algorithm looks for the K extended patches closest to P. These extended patches are centered at frame $I_k$and the distance is written as

$$d(P, Q) = \sum_{i\in\{k-t,\ldots,k+t\}} \| P_i - Q_i \|^2 \qquad (1)$$

As each extended patch contains M 2D image patches, the group contains K · M selected 2D patches of size n × n.

C. Singular Value Decomposition (SVD)

The Singular Value Decomposition (SVD) of these K · M patches is computed and their denoised counterparts are obtained by thresholding of the coefficients.SVD [8] is a matrix decomposition technique. Using SVD, any real matrix X can be decomposed into a product of three matrices U, Σ and V as

$$X=U\Sigma V^T \qquad (2)$$

where U and V are orthogonal matrices and Σ is diagonal matrix. If X is of size m x n then, U is of size m x m orthonormal matrix and it's columns are called as left singular vectors of X and V is of size n x n orthonormal matrix and it's columns are called right singular vectors of P. Some properties of SVD which are useful in image processing are:

- The singular values are unique for a given matrix.
- The rank of matrix A is equal to its nonzero singular values. In many applications, the singular values of a matrix decrease quickly with increasing rank. Using this property we can reduce the noise or compress the matrix data by considering only some higher singular values or the lower ranks.
- The singular values of an image have very good stability i.e. when a small perturbation is added to an image; its singular values don't change significantly.

D. Denoising of Similar Patches

The decision of canceling a coefficient of a certain patch is not taken depending on its magnitude, but the magnitude of the associated singular value. A more robust thresholding is obtained by comparing the singular values to the noise standard deviation and canceling or maintaining the coefficients of all the patches associated to a singular value. The denoised set of patches can be computed as

$$\tilde{X} = FU\Sigma V^T \qquad (3)$$

where F is a $n^2 \times n^2$ diagonal matrix such that Fii = 1 if Σii > τσ and zero otherwise. The whole patch is restored in order to obtain the final estimate by aggregation

E. Second "Oracle" Iteration

A second iteration of the algorithm is performed using the "oracle" strategy. Once the whole sequence has been restored, we re-apply the algorithm on the initial noisy sequence, but motion estimation and patch selection are performed on the result of the first iteration.

Let $\{ I_{K-t}^{W} , \cdot \cdot \cdot , I_{K+t}^{W} \}$ and $\{ I_{K-t}^{0W} , \cdot \cdot \cdot , I_{K+t}^{0W} \}$ be the warped noisy and initially restored images in a temporal neighborhood of Ik where the optical flow has been computed using initially restored images $I_K^0$and I_J^0 . For each patch P of the reference frame Ik , we consider the extended patches P and P^0referring to the extension to the temporal dimension of the patch and its counterpart in the already denoised sequence. The K extended patches that will be selected as similar are the ones minimizing the distance

$$d(P^0, Q^0) = \sum_{i=\in\{k-t,\dots,k+t\}} \| P_i^0 - Q_i^0 \|^2 \qquad (4)$$

Now we have two different sets containing each one K · M 2D patches of size n × n. One set is formed by the patches of the noisy sequence and the other one by the corresponding patches of the already denoised sequence. The SVD is computed in the set of already denoised patches. Let X denote the matrix containing the selected patches of the noisy sequence as rows and X^0 the corresponding matrix with the same patches of the already filtered sequence. We compute the basis associated to X^0 making use of the SVD,

$$X^0 = U^0 \Sigma^0 V^{0^T} \qquad (5)$$

This basis is adapted to the already denoised patches which are noise-free. The coefficients of the noisy patches are computed in this new basis and modified by a Wiener filter before reconstruction. This is written

$$\widetilde{X} = F(XV^0)V^{0^T} \qquad (6)$$

where XV^0 are the coefficients of the noisy patches in the computed SVD basis and F is now a diagonal matrix implementing a optimal wiener filter.

F. *Multiresolution bilateral filter framework*

Finally, multiresolution image denoising framework [18] is applied to each of the denoised images(second stage output). Multiresolution analysis has been proven to be an important tool for eliminating noise in signals; it is possible to distinguish between noise and image information better at one resolution level than another. A signal is decomposed into its frequency sub bands with wavelet decomposition; as the signal is reconstructed back; bilateral filtering is applied to the approximation sub bands. Unlike the standard single-level bilateral filter [9], this multiresolution bilateral filter has the potential of eliminating low-frequency noise components. Bilateral filtering works in approximation sub bands; in addition, it is possible to apply wavelet thresholding [3] to the detail sub bands, where some noise components can be identified and removed effectively. This new image denoising framework combines bilateral filtering and wavelet thresholding. This framework produces results better than the individual applications of the wavelet thresholding or the bilateral filter, or successive application of the wavelet thresholding and the bilateral filter.

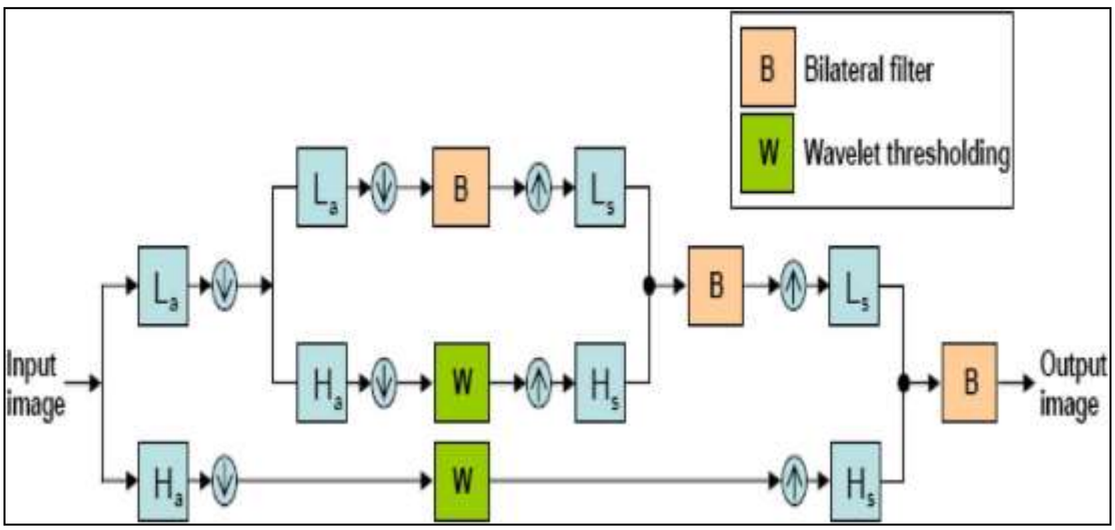

Fig. 2: Multiresolution bilateral filtering framework for an image.

Fig 2 shows one approximation sub band and one detail sub band at each decomposition level; this would be the case when the data is 1-D. For a 2-D data, there are, in fact, one approximation and three (horizontal, vertical, and diagonal) detail sub bands at each decomposition level. Also, in the illustration, there are two levels of decomposition; the approximation sub bands could be decomposed further in an application.

The resultant output is the final denoised images. These frames are then fused to get the output video.

IV. RESULTS

The proposed method is applied to the image sequences with the same parameters for all the sequences, except the temporal neighborhood consisting of M frames. For the shorter sequences we use the 8 frames as temporal neighborhood for the denoising of each one, while for the longer sequences the temporal neighborhood is M = 13 frames. The size of the patches is 8× 8. Patch based algorithms have also been adapted to other kinds of noise than additive white noise.

Performance measures such as Peak Signal to Noise Ratio (PSNR) and Root Mean Square Error (RMSE) values are calculated for the final denoised video. Table 1 shows the average PSNR and RMSE values for different standard deviation (σ) for additive white noise .

Table – 1
Comparison of performance measures for different values of σ

Standard deviation (σ)	*PSNR*	*RMSE*
σ = 10	*49.07*	*6.42*
σ = 20	*44.01*	*7.59*

Comparison of the proposed approach with the existing algorithms such as VBM3D, VBM4D and ND-SAIFIR shows that the purposed algorithm recovers better all image details even in the presence of strong noise. Table 2 shows comparison of proposed approach with existing ones.

Table – 2
Comparison of RMSE values of proposed method with existing algorithms

Standard deviation (σ)	*ND-SAFIR*	*VBM3D*	*VBM4D*	*Proposed method*
σ = 10	*5.14*	*3.88*	*4.06*	*3.57*
σ =20	*7.69*	*5.83*	*6.24*	*5.26*
σ =30	*9.72*	*7.39*	*7.94*	*6.55*
σ =40	*13.85*	*8.57*	*9.38*	*7.64*

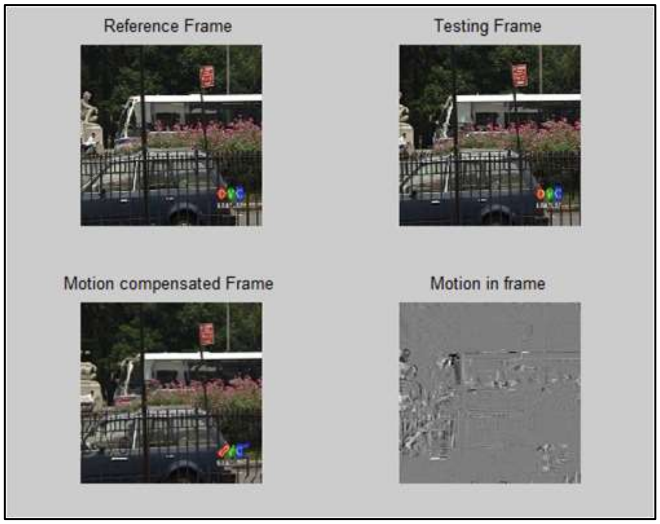

Fig. 3 shows the motion compensation and motion in the frame

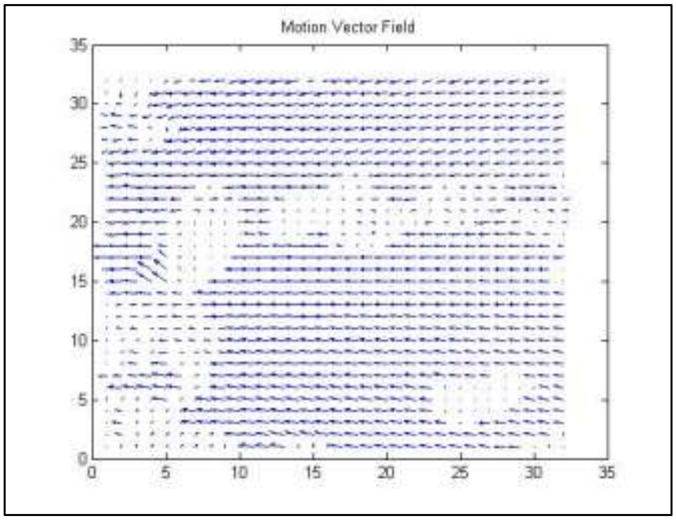

Fig. 4: Motion Vector Field

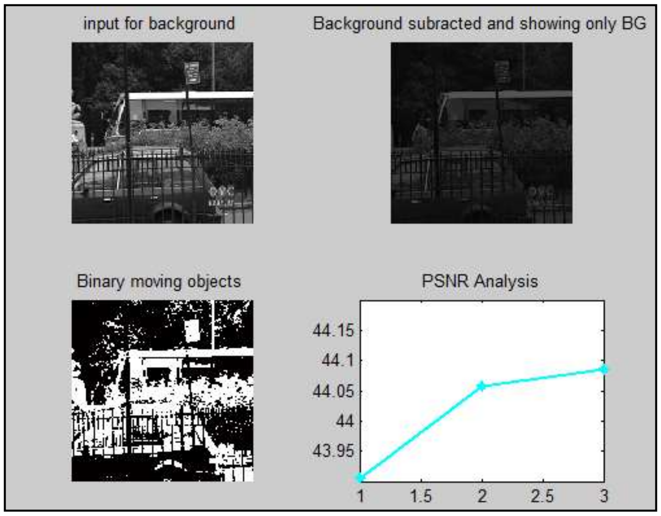

Fig. 5: Background subtraction

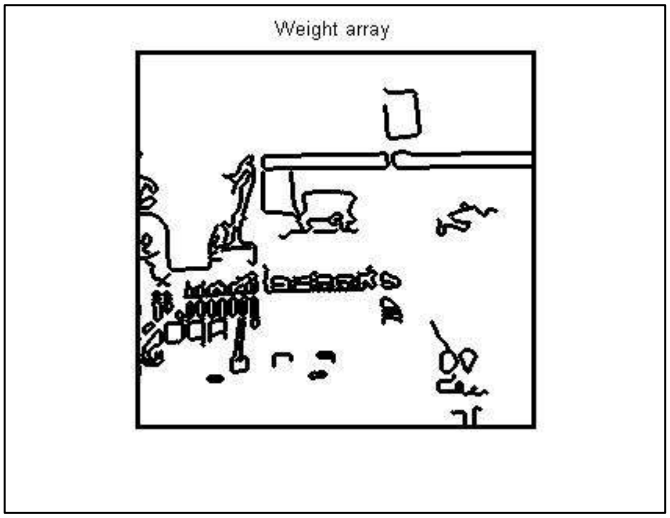

Fig. 6: Weight array

Fig 7: Denoised frames after second iteration

Fig. 8: Final denoised video after Multiresolution Bilateral filter stage

V. CONCLUSION

A novel denoising algorithm is proposed combining motion estimation, patch based denoising algorithms and multiresolution bilateral filtering. Motion compensation permits the robust patch comparison in a spatio temporal volume while the use of SVD for patch denoising preserves texture and details. The comparison with state-of-the-art algorithms illustrates the gain on performance of the proposed approach. Multiresolution image framework helped eliminating the coarse-grain noise in images. The wavelet thresholding adds power to the proposed method as some noise components can be eliminated better in detail sub bands. The proposed framework will inspire further research towards understanding and eliminating noise in real image sequences.

REFERENCES

[1] Antoni Buades, Jose-Luis Lisani, and Marko Miladinovic,"Patch-Based Video Denoising With Optical Flow Estimation", IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 25, NO. 6, JUNE 2016.
[2] C. Ballester, L. Garrido, V. Lazcano, and V. Caselles, A TV-L1 Optical Flow Method With Occlusion Detection (Lecture Notes in Computer Science), vol. 7476. Berlin, Germany: Springer-Verlag, 2012, pp. 31–40.
[3] S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," IEEE Trans. Image Process., vol. 9, no. 9, pp. 1532–1546, Sep. 2000.
[4] Ayvaci, M. Raptis, and S. Soatto, "Occlusion detection and motion estimation with convex optimization," in Proc. Adv. Neural Inf. Process.Syst., 2010, pp. 100–108.
[5] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "BM3D image denoising with shape-adaptive principal component analysis," in Proc. Signal Process. Adapt. Sparse Struct. Represent., Saint-Malo, France, Apr. 2009.
[6] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," IEEE Trans. Image Process.,vol. 15, no. 12, pp. 3736–3745, Dec. 2006.
[7] R. Yan, L. Shao, and Y. Liu, "Nonlocal hierarchical dictionary learning using wavelets for image denoising," IEEE Trans. Image Process., vol. 22, no. 12, pp. 4689–4698, Dec. 2013.
[8] J. Orchard, M. Ebrahimi, and A. Wong, "Efficient nonlocal-means denoising using the SVD," in Proc. 15th IEEE Int. Conf. Image Process.,Oct. 2008, pp. 1732–1735.
[9] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in Proc. 6th Int. Conf. Comput. Vis., Jan. 1998, pp. 839–846.
[10] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 2. Jun. 2005, pp. 60–65.
[11] M. Lebrun, A. Buades, and J.-M. Morel, "A nonlocal Bayesian image denoising algorithm," SIAM J. Imag. Sci., vol. 6, no. 3, pp. 1665–1688, 2013.
[12] J. Boulanger, C. Kervrann, and P. Bouthemy, "Space-time adaptation for patch-based image sequence restoration," IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 6, pp. 1096–1102, Jun. 2007.
[13] Q. Xu, H. Jiang, R. Scopigno, and M. Sbert, "A new approach for very dark video denoising and enhancement," in Proc. 17th IEEE Int. Conf. Image Process., Sep. 2010, pp. 1185–1188.
[14] H. Yue, X. Sun, J. Yang, and F. Wu, "Image denoising by exploring external and internal correlations," IEEE Trans. Image Process., vol. 24, no. 6, pp. 1967–1982, Jun. 2015.
[15] J. Mairal, G. Sapiro, and M. Elad, "Learning multiscale sparse representations for image and video restoration," Multiscale Model. Simul., vol. 7, no. 1, pp. 214–241, 2008.
[16] C. Liu and W. T. Freeman, "A high-quality video denoising algorithm based on reliable motion estimation," in Proc. 11th Eur. Conf. Comput. Vis., 2010, pp. 706–719.
[17] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping," in Proc. 8thEur. Conf. Comput. Vis., 2004, pp. 25–36.

[18] Ming Zhang and Bahadir K. Gunturk:"Multiresolution Bilateral filtering for image denoising", IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 17, NO. 12, DECEMBER 2008.
[19] M. Elad, "On the origin of the bilateral filter and ways to improve it,"IEEE Trans. Image Process., vol. 11, no. 10, pp. 1141–1151, Oct. 2002.
[20] A. Buades, B. Coll, and J. Morel, "Neighborhood filters and PDE's," Numer. Math., vol. 105, pp. 1–34, 2006.
[21] C. Kervrann and J. Boulanger, "Optimal spatial adaptation for patch based image denoising," IEEE Trans. Image Process., vol. 15, no. 10, pp. 2866–2878, Oct. 2006