MULTISPECTRAL IMAGE DENOISING IN WAVELET DOMAIN WITH UNSUPERVISED TENSOR SUBSPACE-BASED METHOD

A. Zidi, K. Spinnler

Fraunhofer IIS, Röntgentechnik EZRT
Flugplatzs., 90768 Fürth, GE

J. Marot, S. Bourennane

AMU, Ecole Centrale Mars.,
Inst. Fresnel, 13397 Marseille, FR

ABSTRACT

Multiway Wiener filtering has been inserted in a wavelet framework to enhance spatial details while denoising multidimensional images. An elevated number of rank values is required. A solution is to retrieve the best rank values while minimizing a mean square criterion. In this paper, we justify the adaptation for this purpose of a stochastic optimization method, and we evaluate comparatively a genetic algorithm and particle swarm optimization. Results obtained on multispectral images in terms of signal to noise ratio and perceptual image quality permit to emphasize the performance of the obtained unsupervised method for realistic noise magnitudes.

Index Terms— Rank, Wavelets, Tensor, Multispectral

1. INTRODUCTION

In the frame of multidimensional data denoising, a now state-of-the-art method is the multiway Wiener filtering (MWF) [1]. Then, the MWPT-MWF (Multidimensional Wavelet Packet Transform -Multiway Wiener Filtering) method has been proposed [2, 3]. The main goal of MWPT-MWF is to denoise multidimensional images while preserving details. Indeed, since the interest of the remote sensing community for hyperspectral images (HSI) is growing, it is necessary to develop denoising methods: most of HSIs, acquired by Hyperspectral Digital Imagery Collection Experiment (HYDICE [4]) and Airborne Visible/Infrared Imaging Spectrometer (AVIRIS [5]) sensors, are impaired by noise from solar radiation, or atmospheric scattering [6]. We consider subsampled versions of HSI acquisitions, *i.e.* multispectral images. MWPT-MWF has yielded good results in terms of signal to noise ratio (SNR) and classification accuracy when the image contains lots of details such as small targets. The drawback of this method is that a large number of subspace rank values must be estimated to ensure accurate denoising results: there exist one rank value for each tensor mode, several modes for each wavelet coefficient, and several coefficients for each level. In [2], a study about the accurate depth of the wavelet decomposition has been performed, but the subspace ranks are estimated with AIC, which is known, in the array processing field for instance, not to perform well in a noisy environment. In [7], the authors propose instead to minimize the mean square error (MSE) between expected and estimated tensor, but no numerical justification is provided for the choice of the optimization method. The aim of this paper is to investigate optimization methods and compare their efficiency and computational load when estimating multiple ranks in MWPT-MWF. In 2 we state the denoising problem to be solved, in section 3 we justify our choice of stochastic optimization methods to minimize the previously cited MSE criterion. In section 4 we detail the steps of our algorithm for multiple rank estimation in the frame of tensor data denoising. In section 5 we show comparative results on multispectral images.

EUVIP 2016, Oct. 25-27, 2016, Marseille, France.
978-1-5090-2781-1/16/$31.00 ©2016 IEEE.

2. PROBLEM SETTING

We consider a noisy multidimensional signal, also called tensor: a signal $\mathcal{X}$ impaired by a multidimensional additive white noise $\mathcal{N}$ [8]. Tensors are multidimensional arrays of size $I_1, I_2, \ldots, I_N$, where N is the 'order' of the tensor, or equivalently the number of modes. We aim at denoising tensor $\mathcal{R} = \mathcal{X} + \mathcal{N}$ to get a denoised tensor denoted by $\hat{\mathcal{X}}$. This denoised tensor depends on the so-called 'subspace ranks' $\{K1, K2, \ldots KN\}$ which must be estimated. Following the notations in [2], $l_1, l_2, \ldots l_N$ are the number of decomposition levels in the wavelet tranform for modes $1, 2, \ldots N$. There exists 2^{l_k} coefficients for mode k, $k = 1, 2, \ldots N$. Each wavelet coefficient is a tensor of order N [2], so N rank values must be estimated for each coefficient. In total, the number of rank values to be estimated in the whole algorithm is given by $N \prod_{k=1}^{N} 2^{l_k}$, possibly elevated.
Hence the need for a non-supervised method for the estimation of the rank values. The first issue consists then in finding an appropriate scalar criterion to estimate the subspace ranks $K1$, $K2$, $\ldots$ KN, we choose:

$$J(K1, K2, \ldots KN) = ||\mathcal{R} - \hat{\mathcal{X}}||^2, \qquad (1)$$

where $||.||$ represents the Frobenius norm. The criterion J is a nonlinear function of the parameters $K1, K2, \ldots KN$.
The noise is considered as Gaussian, so we minimize a least square error to maximize the log-likelihood.

The second issue is then the following: a global optimization method is required. We compare a genetic algorithm and particle swarm optimization not only in terms of SNR and computational load, but also visual aspect of the denoised images through the MSSIM (Mean Structural Similarity) criterion [9].

3. GLOBAL OPTIMIZATION METHODS

As the considered optimization problem is highly non-linear, and we have no insurance that some constraints on the minimized function are respected, we left aside the deterministic optimization methods such as Gradient, or Dividing Rectangles (DIRECT) [10]. We focus on two types of stochastic optimization methods: particle swarm optimization (PSO) and a genetic algorithm (GA).
In their seminal work concerning particle swarm optimization, Kennedy and Eberhart [11] got inspired by [12], where the term 'particle swarm' was chosen to define the members of a population or test set. In their paradigm, the population members are mass-less and volume-less. Their evolution is described through position, speed, and acceleration parameters [11].
An often cited, now well-known reference [13] introduces genetic algorithms in the context of evolutionary computation which implies an evolution of a population of candidates which is inspired by Darwin's natural selection theory. Another largely cited reference presents basics about genetics, the hierarchical genetic algorithm, and applications to H_∞ control, neural network, and speech recognition [14].

4. PROPOSED ALGORITHM: UNSUPERVISED MWPT-MWF

We wish to adapt rank estimation to the most recent version of MWF, *i.e.*, its implementation in a wavelet framework [2], where an automatic rank estimation is required due to the high number of required rank values: we wish to obtain an unsupervised MWPT-MWF algorithm. In this paper we will further focus on multispectral images, for which $N = 3$. Hence, in the following, we restrict our study to third-order tensors. Following [2], minimizing the MSE between $\mathcal{X}$ and its estimate $\hat{\mathcal{X}}$ is equivalent to minimizing the MSE between $\mathcal{C}^{\mathcal{X}}_{\mathbf{l},\mathbf{m}}$ and $\mathcal{C}^{\hat{\mathcal{X}}}_{\mathbf{l},\mathbf{m}}$ for each $\mathbf{m}$:

$$\|\mathcal{X} - \hat{\mathcal{X}}\|^2 = \|\mathcal{C}^{\mathcal{X}}_{\mathbf{l}} - \mathcal{C}^{\hat{\mathcal{X}}}_{\mathbf{l}}\|^2 = \sum_{\mathbf{m}} \|\mathcal{C}^{\mathcal{X}}_{\mathbf{l},\mathbf{m}} - \mathcal{C}^{\hat{\mathcal{X}}}_{\mathbf{l},\mathbf{m}}\|^2 \qquad (2)$$

where $\mathcal{C}^{\mathcal{X}}_{\mathbf{l}}$ is the wavelet packet coefficient tensor for levels in $\mathbf{l} = [l_1, l_2, l_3]^T$, $\mathcal{C}^{\mathcal{X}}_{\mathbf{l},\mathbf{m}}$ is the coefficient

subtensor of $\mathcal{C}_{\mathbf{l}}^{\mathcal{X}}$ where $\mathbf{m} = [m_1, m_2, m_3]^T$ is the index vector, $1 \leq m_k \leq 2^{l_k} - 1, k = 1, \ldots, 3$.
We wish to minimize all terms of the summation in Eq. (2), knowing that the noise-free tensor $\mathcal{X}$ is not available. For this we propose Algorithm 1, *multidimensional wavelet packet transform and multiway Wiener filtering with rank estimation* (MWPT-MWF-RE). In Algorithm 1, $\mathbf{H}_{1,\mathbf{m}}, \mathbf{H}_{2,\mathbf{m}}, \mathbf{H}_{3,\mathbf{m}}$ denote the k-mode filters of MWF, which depend on rank values (K1,K2,K3) [8, 2]; $\mathcal{C}_{\mathbf{l},\mathbf{m}}^{\mathcal{R}}$ denote the wavelet coefficients of $\mathcal{R}$.

Algorithm 1 MWPT-MWF-RE

Input: noisy tensor $\mathcal{R}$.
• compute the wavelet decomposition of the noisy tensor $\mathcal{R}$: $\mathcal{C}_{\mathbf{l}}^{\mathcal{R}} = \mathcal{R} \times_1 \mathbf{W}_1 \times_2 \mathbf{W}_2 \times_3 \mathbf{W}_3$
• extract the wavelet coefficients [2]:
$\mathcal{C}_{\mathbf{l},\mathbf{m}}^{\mathcal{R}} = \mathcal{C}_{\mathbf{l}}^{\mathcal{R}} \times_1 \mathbf{E}_{m_1} \times_2 \mathbf{E}_{m_2} \times_3 \mathbf{E}_{m_3}$,
• for each wavelet coefficient $\mathcal{C}_{\mathbf{l},\mathbf{m}}^{\mathcal{R}}$:
i) estimate with a global optimization method the optimal rank values $\hat{K1}, \hat{K2}, \hat{K3}$ in terms of the criterion:
$J_{\mathbf{m}}(K1, K2, K3) = ||\mathcal{C}_{\mathbf{l},\mathbf{m}}^{\mathcal{R}} - \mathcal{C}_{\mathbf{l},\mathbf{m}}^{\hat{\mathcal{X}}}||^2$
where $\mathcal{C}_{\mathbf{l},\mathbf{m}}^{\hat{\mathcal{X}}} = \mathcal{C}_{\mathbf{l},\mathbf{m}}^{\mathcal{R}} \times_1 \mathbf{H}_{1,\mathbf{m}} \times_2 \mathbf{H}_{2,\mathbf{m}} \times_3 \mathbf{H}_{3,\mathbf{m}}$.
ii) apply MWF to each coefficient subtensor $\mathcal{C}_{\mathbf{l},\mathbf{m}}^{\mathcal{R}}$, with the optimal rank values.
• obtain $\mathcal{C}_{\mathbf{l}}^{\hat{\mathcal{X}}}$ by concatenating all coefficients $\mathcal{C}_{\mathbf{l},\mathbf{m}}^{\hat{\mathcal{X}}}$.
• reconstruct the final estimated tensor by inverse wavelet transform: $\hat{\mathcal{X}} = \mathcal{C}_{\mathbf{l}}^{\hat{\mathcal{X}}} \times_1 \mathbf{W}_1^T \times_2 \mathbf{W}_2^T \times_3 \mathbf{W}_3^T$
Output: denoised tensor $\hat{\mathcal{X}}$.

With a global optimization method, algorithm 1 is supposed to converge asymptotically towards the best set of rank values for each coefficient of each decomposition level. In practice, the total number of iterations, *i.e.*, the parameter maxit is fixed automatically: the algorithm stops when the criterion $J_{\mathbf{m}}(K1, K2, K3)$ does not vary from an iteration to another by a small parameter ϵ set by the user.

5. DENOISING RESULTS

In this section, we apply the proposed method on real-world multispectral images issued from an AVIRIS sensor. The results obtained are evaluated in terms of SNR and perceptual image quality. The perceptual image quality is measured through mean $SSIM$ [9] over all spectral bands. In [9], under the assumption that human visual perception is highly adapted for extracting structural information from a scene, the structural similarity (SSIM) criterion is proposed to quantify the degradation of structural information. Multispectral images follow the tensor model and are expressed as: $\mathcal{R} = \mathcal{X} + \mathcal{N}$. Tensors $\mathcal{R}$, $\mathcal{X}$, and $\mathcal{N}$ are of size $I_1 \times I_2 \times I_3$. For each spectral band indexed by $i = 1, \ldots, I_3$, the noise $\mathbf{N}(:, :, i)$ is assumed stationary zero-mean.
Programmes were written in $Matlab^{\circledR}$, and run on a PC running Windows, with a 3GHz double core and 3GB RAM. The denoising performance will be evaluated through $SNR = 10\ log(\frac{||\mathcal{X}||^2}{||\mathcal{X}-\hat{\mathcal{X}}||^2})$ and $SSIM(\hat{\mathcal{X}}, \mathcal{X})$ [9]. The images are artificially impaired with white, identically distributed random noise with the following input SNR values (in dB): 5, 10, 15, 20 and 25. In the wavelet decomposition, following the recommendations in [2] we choose Coiflets and Daubechies wavelet functions. Following the recommandation in [7], we choose two decomposition levels for the space modes and no decomposition in the wavelength mode: $\mathbf{l} = [2, 2, 0]^T$. This yields 16 wavelet coefficients (4 coefficients for each level), which are 3^{rd}-order tensors of size $64 \times 64 \times 16$. For a given image, the total number of rank values to be estimated is $3 \prod_{k=1}^{3} 2^{l_k} = 3 * (2^2 * 2^2 * 2^0) = 48$. When we run PSO algorithm, the swarm size is 10 and $\epsilon = 10^{-6}$. This generally yields $\mathrm{maxit} = 150$ iterations. The acceleration constants γ_{1i} and γ_{2i} are set to 2 and 3 respectively. We run a version of GA using a Lagrangian algorithm [15, 16], with 300 individuals (ranks) in the initial population. We use the fitness function to provide a measure of how individuals have performed in the problem domain. In the following subsections we present the numerical and visual results obtained with GA, PSO and AIC. For the RGB display of the multispectral images throughout the section, we select 3 representative bands in the red, green, and blue wavelength domains respectively, $R = 690$, $G = 550$, $B = 450$.

5.1. Visual results

We first present results obtained on two subimages of PAVIAU HSI. PAVIAU is interesting in the sense that it contains small details but also rather homogeneous regions. From the presence of such small local features we expect to emphasize the interest of wavelet-based processing, because it permits to process seperately high and low frequency features. The first subimage is presented in Fig. 1, and the second subimage in Fig. 2. The first image is subsampled to size $256 \times 256 \times 16$. It exhibits some details but also rather homogeneous regions (the roof of the main building for instance). The second image is subsampled to size $128 \times 128 \times 16$. It was extracted from another region of PAVIAU, where more small fetaures are present and should be preserved. We exemplify the proposed method with an input SNR of 10dB for each image. By analysis of Fig 1 and Fig 2, we can draw the following conclusions: while some noise is remaining in the results provided by AIC, the homogeneous regions are smoother when GA and PSO are used, and meanwhile the details such as parked cars are clearly visible. This can be due to an overestimation of the rank values when AIC is used.

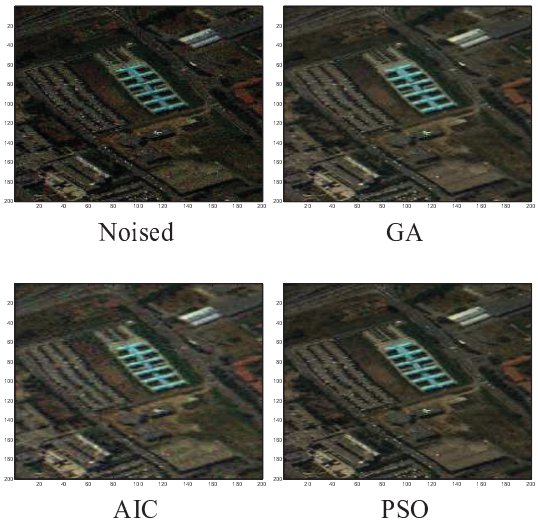

Fig. 1. example images: noised, 10 dB, and 3 methods.

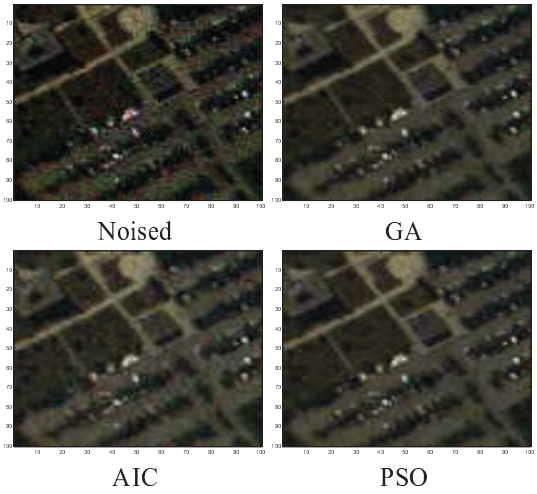

Fig. 2. example images: zoom on image: noised, 10 dB, and 3 methods.

In fig 2 we focus on the region containing many details and frontiers. Comparing the result obtained by AIC and wavelet decomposition with rank estimation by PSO and GA, we notice that the frontiers are much less blurred and that the details are better preserved when wavelet decomposition is used, and that the homogeneous regions are better denoised when PSO is used compared to the case where AIC is used.

5.2. Numerical results

In this subsection, to avoid the border issues, the numerical results are computed from images truncated to the size $200 \times 200 \times 16$ (resp. $100 \times 100 \times 16$). Tables 1 and 3 present the results obtained in terms of output SNR. Table 2 and Table 4 present the results obtained in terms of output MSSIM. The closer to 1, the better the denoising results [9]. From these Tables, we can draw the following conclusions: firstly, the output SNR and MSSIM values obtained with GA and PSO are very close, and significantly higher than those obtained with AIC, except with an input SNR of 25dB (in Tables 1, 2, and 4). This confirms the fact that, for intermediary and realistic input SNR values, AIC tends to overestimate the rank values, which is not the case when GA or PSO are

adapted. Secondly, we notice that, with an input SNR of 25dB, the output SNR is less than 25 dB when either GA, PSO, or AIC are used: all methods are anyway of limited interest for input SNR values equal to and probably higher than 25 dB. Thirdly, apart from this particular case where the input SNR is 25 dB, we can infer from all MSSIM values that the perceptual aspect of the denoised images is best when either GA of PSO are used. Last but not least, when running GA, PSO and AIC on the multispectral image of Fig. 1, we got computational times equal to 123.2 sec., 2.199 sec. and 0.3501 sec., respectively. It shows that, while similar output SNR and mean SSIM values are reached, PSO is much faster than GA. We then conclude that PSO is preferable, compared to GA.

	GA	PSO	AIC
5 db	10.4620	10.5285	11.2588
10 db	14.6130	14.8469	12.9793
15 db	18.1850	18.6587	15.9241
20 db	19.8834	21.5239	19.3406
25 db	21.7576	22.8081	23.7061

Table 1. Output SNR vs. input SNR for image in Fig. 1

	GA	PSO	AIC
5 db	0.5299	0.5455	0.4997
10 db	0.7053	0.7341	0.5624
15 db	0.8594	0.8655	0.7488
20 db	0.9041	0.9275	0.8775
25 db	0.9319	0.9478	0.9569

Table 2. Output MSSIM vs. input SNR for image in Fig. 1

	GA	PSO	AIC
5 db	11.0060	10.9490	10.9379
10 db	15.0184	15.2056	13.6347
15 db	18.5342	18.7736	16.8716
20 db	20.3681	21.2576	19.8940
25 db	21.9652	22.5749	22.5152

Table 3. Output SNR vs. input SNR for image in Fig. 2

	GA	PSO	AIC
5 db	0.5726	0.5798	0.5029
10 db	0.7877	0.7927	0.6785
15 db	0.8931	0.8956	0.8388
20 db	0.9223	0.9424	0.9147
25 db	0.9352	0.9440	0.9586

Table 4. Output MSSIM vs. input SNR for image in Fig. 2

6. CONCLUSION

The contribution of this paper is an unsupervised multiple rank estimation method inserted in a wavelet framework, for the purpose of multidimensional data denoising. We show that a stochastic optimization method is appropriate to minimize a least-squares criterion between expected and denoised wavelet coefficient. Focusing on a GA and PSO, we have illustrated the ability of the proposed method to remove noise with realistic magnitude values in an application to multispectral image denoising. Referring to SNR and perceptual quality evaluated through mean SSIM, GA and PSO perform better than AIC while estimating the rank values: the details are better preserved as well as the perceptual information. We also have shown that PSO is faster than GA for equivalent numerical results.

7. REFERENCES

[1] D. Muti and S. Bourennane, "Multidimensional filtering based on a tensor approach," *Signal Proceesing Journal, Elsevier*, vol. 85, no. 12, pp. 2338–2353, Dec. 2005.

[2] T. Lin and S. Bourennane, "Hyperspectral image processing by jointly filtering wavelet component tensor," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 6, pp. 3529–3541, 2013.

[3] J. Marot and S. Bourennane, "Recent advances on tensor models and their relevance for multidimensional data processing," in *2014 48th Asilomar Conference on Signals, Systems and Computers*, Nov 2014, pp. 586–590.

[4] R.W. Basedow, D.C. Carmer, and M.L. Anderson, "Hydice system, implementation and performance," in *SPIE Proc.*, Orlando, FL, April 1995, vol. 2480, pp. 416–428.

[5] Robert O. Green, Charles M. Sarture, Christopher J. Chovit, Jessica A. Faust, Pavel Hajek, and H. lan Novak, "Aviris: A new approach to earth remote sensing," *Opt. Photon. News*, vol. 6, no. 1, pp. 30, Jan 1995.

[6] J.P. Kerekes and J.E. Baum, "Full spectrum spectral imaging system analytical model," *IEEE Trans. on Geosc. and Remote Sensing*, vol. 43, no. 3, pp. 571–580, March 2005.

[7] Julien Marot and Salah Bourennane, *Advanced Concepts for Intelligent Vision Systems: 16th International Conference, ACIVS 2015, Catania, Italy, October 26-29, 2015. Proceedings*, chapter Improvement of a Wavelet-Tensor Denoising Algorithm by Automatic Rank Estimation, pp. 779–790, Springer International Publishing, Cham, 2015.

[8] Damien Muti, Salah Bourennane, and Julien Marot, "Lower-rank tensor approximation and multiway filtering," *SIAM Journal on Matrix Analysis and Applications*, vol. 30, no. 3, pp. 1172–1204, 2008.

[9] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 600 – 612, 2004.

[10] Donald R. Jones, Cary D. Perttunen, and Bruce E. Stuckman, "Lipschitzian optimization without the lipschitz constant," *Journal of Optimization theory and Application*, vol. 79, no. 1, pp. 157–181, October. 1993.

[11] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *IEEE International Conference on Neural Networks*, Perth, 1995, pp. 1942–1948.

[12] William T Reeves, "Particle systemsâĂŤa technique for modeling a class of fuzzy objects," *ACM Transactions on Graphics (TOG)*, vol. 2, no. 2, pp. 91–108, 1983.

[13] Melanie Mitchell, *An introduction to genetic algorithms*, MIT press, 1998.

[14] Kim-Fung Man, Kit Sang TANG, and Sam Kwong, *Genetic algorithms: concepts and designs*, Springer Science & Business Media, 2012.

[15] David E Goldberg et al., *Genetic algorithms in search optimization and machine learning*, vol. 412, Addison-wesley Reading Menlo Park, January 1989.

[16] Andrew R Conn, Nicholas IM Gould, and Philippe Toint, "A globally convergent augmented lagrangian algorithm for optimization with general constraints and simple bounds," *SIAM Journal on Numerical Analysis*, vol. 28, no. 2, pp. 545–572, 1991.