RESTORATION OF IMAGES CORRUPTED BY MIXED GAUSSIAN-IMPULSE NOISE BY ITERATIVE SOFT-HARD THRESHOLDING

M. Filipović

Rudjer Bošković Institute
Bijenička 54
10000 Zagreb, Croatia

A. Jukić[*]

Department of Medical Physics and Acoustics
University of Oldenburg
26111 Oldenburg, Germany

ABSTRACT

We address the problem of restoration of images which have been affected by impulse or a combination of impulse and Gaussian noise. We propose a patch-based approach that exploits approximate sparse representations of image patches in learned dictionaries. For every patch, sparse representation in a dictionary is enforced by ℓ_1-norm penalty, and sparsity of the residual is enforced by ℓ_0-quasi-norm penalty. The obtained non-convex problem is solved iteratively by a combination of soft and hard thresholding, and a proof of convergence to a local minimum is given. Experimental evaluation suggests that the proposed approach can produce state-of-the-art results for some types of images, especially in terms of the structural similarity (SSIM) measure. In addition, proposed iterative thresholding algorithm could possibly be applied to general inverse problems.

Index Terms— Denoising, Impulse Noise, Sparse Representation, Dictionary, Thresholding

1. INTRODUCTION

Image denoising is a fundamental task in image processing. In this paper, we are interested in a special kind of noise, which is a mixture of impulse and Gaussian noise. Generally, several types of noise affect images: Poissonian, Gaussian and impulse. Poisson-Gaussian noise mixture can generally (when no data is lost, i.e. an image is not affected by impulse noise) be transformed into pure Gaussian noise [1], or can be treated by specialized methods [2]. However, the mixture of Gaussian and impulse noise is quite challenging for denoising. Two models of impulse noise that are used in the literature are salt-and-pepper noise and random-valued impulse noise. Let us suppose that the dynamic range of an image is $[d_{\min}, d_{\max}]$. Then, in the salt-and-pepper noise model, every image pixel is replaced, with a given probability, with a value $d_{\min}$ or $d_{\max}$. The second model of impulse noise, which is much more difficult to handle, assumes that a pixel value is replaced, again with a given probability, with a random value in the range $[d_{\min}, d_{\max}]$. Since detecting salt-and-pepper noise is much easier than random-valued impulse noise, we concentrate on the more general scenario of random-valued impulse noise in this paper.

Before introducing and describing our approach, we review previous work. Most approaches for removing a mixture of Gaussian and impulse noise generally start by detecting the pixels corrupted by impulse noise. Then, after the influence of these noisy pixels is reduced, some robust noise removal method is used. In [3], a three-phase denoising approach was proposed. Firstly, the outlier candidates pixels are detected using a median-type filter. Then, the initial approximation of a clean image is computed by a variant of the KSVD denoising algorithm [4, 5] using only pixels that are declared as clean. Finally, a model enforcing small ℓ_1-norm of the error and sparse representation (measured by ℓ_0-quasi-norm, which is defined as the number of nonzero elements of its argument) of image patches in learned dictionary is approximately solved. The dictionary is adaptively updated in every step. Similar approaches were presented in [6–8]. However, these papers simultaneously tackle the problem of blur in images (as noted in [9], this makes the problem of detection of noisy pixels even easier, since the image is smoother).

In [9], a patch-based approach for removing impulse noise, based on robust statistics, was proposed. This approach has three main steps. Firstly, the fraction of (impulse) noise-corrupted pixels is estimated using some impulse noise detector. Then, for every patch a set of similar patches is found using some robust measure of similarity between patches. Finally, an estimate of clean patch is found using robust statistical estimation on the set of similar patches found in the previous step. This approach provided state-of-the-art results, as demonstrated in [9]. We perform extensive experimental comparison both with the approach in [9] and the one in [3] in Section 3.

In [10], a model similar to the one proposed here was used. Namely, ℓ_0 penalty on the noise was used, while the total variation term was used as a regularization (therefore, a global regularization was used, unlike patch-based, as used in this paper). Another recent method is described in [11].

[*]The author is supported by the Marie Curie Initial Training Network DREAMS (Grant agreement no. ITN-GA-2012-316969).

They used structured sparsity as a regularization by adapting the approach from [12] to impulsive noise removal and blind inpainting. An approach for general inverse problems in imaging was proposed in recent paper [13]. It uses a combined image regularization model based on total generalized variation and shearlets. Their approach, as demonstrated in the paper, recovers both edges and fine details much better than the regularization models based on total variation and wavelets. Computational complexity of the algorithm wasn't discussed in much detail, similarly to all the approaches discussed in this section.

In this paper, an approach for removing impulse or a combination of impulse and Gaussian noise is proposed. For each patch, an optimization problem is formulated, where a sparse representation in learned dictionary is enforced by ℓ_1-norm penalty, and a sparse residual is enforced by ℓ_0-quasi-norm penalty. The obtained non-convex problem is solved in an iterative procedure by combination of soft and hard thresholding, with established convergence to a local minimum. In the experimental section a detailed comparison against two state-of-the-art methods is performed. Obained results show that the proposed method produces good results for some types of images, especially in terms of the structural similarity (SSIM) measure (described later).

1.1. Notation and organization of the paper

Scalars are denoted by lowercase, vectors by bold lowercase, and matrices by bold uppercase letters. Operators and sets are denoted by uppercase or calligraphic letters. Transpose of matrix $\mathbf{D}$ is denoted by $\mathbf{D}^T$. The estimation of $\mathbf{x}$ at iteration k is denoted by $\mathbf{x}^{(k)}$. i-th component of vector $\mathbf{x}$ is denoted by x_i. Componentwise multiplication (for vectors and matrices) is denoted by $\otimes$.

In Section 2 we describe our approach. Numerical experiments and comparison with state-of-the-art methods are presented in Section 3. Conclusions are given in Section 4.

2. PATCH-BASED APPROACH USING A LEARNED DICTIONARY

The overall process of denoising consists of several steps: (offline) dictionary learning, impulse detection and iterative minimization algorithm; these are described in the following subsections.

2.1. Dictionary learning

In [3], dictionary for sparse representation was learned on the damaged image itself, using a variant of the K-SVD algorithm. However, in this paper we simply learn the dictionary offline, on a training set of images of natural scenes. The dictionary is then fixed during the image restoration process. The dictionary is learned using the independent component analysis (ICA) on the set of patches extracted from the images in the training set. We demonstrate that this approach gives good results. Similar results could be obtained using some other dictionary learning algorithm (for example, K-SVD, or ℓ_1-based dictionary learning from [14]), but in our simulations the dictionary learned using ICA consistently gave better results in terms of the SSIM measure.

2.2. Impulse detection

We use the well known ROAD detector [15], as also used in [9]. It can be described as follows. For every image pixel, all absolute differences between that pixel and pixels in the surrounding patch are computed. These differences are then sorted, and the value of ROAD statistic at that pixel is obtained by computing the sum of 4 smallest differences. When the ROAD is above some threshold, the pixel is considered as noisy. Parameters used in the experiments are described in Section 3.

2.3. Image restoration algorithm

We propose the following formulation of the problem. For every image patch, the following problem is solved:

$$\min_{\mathbf{x},\,\mathbf{f}} \frac{1}{2} \left\| \mathbf{\Omega} \otimes (\mathbf{u} - \mathbf{D}\mathbf{x} - \mathbf{f}) \right\|_2^2 + \lambda_1 \left\| \mathbf{x} \right\|_1 \quad \text{subject to} \quad \left\| \mathbf{\Omega} \otimes \mathbf{f} \right\|_0 \leq \lambda_2. \tag{1}$$

Here, $\mathbf{u} \in \mathbb{R}^n$ denotes noisy image patch, $\mathbf{D} \in \mathbb{R}^{n\times m}$ is the learned dictionary, $\mathbf{\Omega} \in \mathbb{R}^n$ is the vector indicating the positions of pixels not affected by *impulse* noise (as declared by the ROAD detector) in the current patch, $\mathbf{x} \in \mathbb{R}^m$ is a vector of coefficients, $\mathbf{f} \in \mathbb{R}^n$ is a sparse vector of pixels affected by impulse noise and λ_1, λ_2 are parameters. λ_1 is a parameter that depends on the level of Gaussian (bounded) noise, while λ_2 controls the level of impulsive (sparse) noise. Unconstrained formulation of the problem is also possible (in that case, λ_2 has different interpretation); however, it seems more natural to consider formulation (1) because here λ_2 has a simple interpretation: it bounds the number of pixels affected by impulsive noise. Another approach would be to use a *global* regularization on f (since the patches are generally overlapping); however, the local approach used here is simpler and gives good results. Among many possible formulations, the ℓ_1 norm could also be used as a measure of sparsity of noise; however, in our simulations better results were obtained with (1).

Formulation (1) can be solved by a combination of soft and hard thresholding. Let us suppose that the 2-norm of matrix $\mathbf{D}$ is less than 1 ($\mathbf{D}$ can be normalized if necessary, so that this assumption is realistic). Let us denote $F(\mathbf{x}, \mathbf{f}) =$

$\frac{1}{2}\left\|\Omega \otimes (\mathbf{u} - \mathbf{Dx} - \mathbf{f})\right\|_2^2$. As in [16], we introduce the surrogate function

$$Q_{L,\,\mathbf{x}^{(k)},\,\mathbf{f}^{(k)}}(\mathbf{x},\,\mathbf{f}) = F(\mathbf{x},\,\mathbf{f}) + \left(\begin{bmatrix}\mathbf{x}\\ \mathbf{f}\end{bmatrix} - \begin{bmatrix}\mathbf{x}^{(k)}\\ \mathbf{f}^{(k)}\end{bmatrix}\right)^T \cdot$$
$$\cdot \nabla F(\mathbf{x}^{(k)},\,\mathbf{f}^{(k)}) + \frac{L}{2}\left\|\begin{bmatrix}\mathbf{x}\\ \mathbf{f}\end{bmatrix} - \begin{bmatrix}\mathbf{x}^{(k)}\\ \mathbf{f}^{(k)}\end{bmatrix}\right\|_2^2 \quad (2)$$

The surrogate function in (2) majorizes $F(\mathbf{x},\,\mathbf{f})$ when L is greater than the Lipschitz constant of the gradient of F [16]. Variables $\mathbf{x}$ and $\mathbf{f}$ in $Q_{L,\,\mathbf{x}^{(k)},\,\mathbf{f}^{(k)}}$ decouple, since $Q_{L,\,\mathbf{x}^{(k)},\,\mathbf{f}^{(k)}}$ can (up to a constant) be written as

$$\frac{1}{2}\left\|\mathbf{x} - \left(\mathbf{x}^{(k)} + \frac{1}{L}\mathbf{D}^T\left(\mathbf{\Omega} \otimes \left(\mathbf{u} - \mathbf{Dx}^{(k)} - \mathbf{f}^{(k)}\right)\right)\right)\right\|_2^2$$
$$+\frac{1}{2}\left\|\mathbf{f} - \left(\mathbf{f}^{(k)} + \frac{1}{L}\left(\mathbf{\Omega} \otimes \left(\mathbf{u} - \mathbf{Dx}^{(k)} - \mathbf{f}^{(k)}\right)\right)\right)\right\|_2^2.$$

Therefore, minimization of the surrogate function can be performed separately with respect to $\mathbf{x}$ and $\mathbf{f}$. By iteratively solving the following problems

$$\mathbf{x}^{(k+1)} = \arg\min_{\mathbf{x}}\left(Q_{L,\,\mathbf{x}^{(k)},\,\mathbf{f}^{(k)}}(\mathbf{x},\,\mathbf{f}) + \lambda\left\|\mathbf{x}\right\|_1\right) \quad (3)$$

$$\mathbf{f}^{(k+1)} = \arg\min_{\mathbf{f}} Q_{L,\,\mathbf{x}^{(k)},\,\mathbf{f}^{(k)}}(\mathbf{x},\,\mathbf{f})$$
$$\text{subject to } \left\|\mathbf{\Omega} \otimes \mathbf{f}\right\|_0 \leq \lambda_2, \quad (4)$$

the objective function in (1) is decreased. Minimizations in (3) and (4) reduce to soft [17] and hard thresholding [18] operators, respectively. It should be noted that we are *not* performing *alternating* optimization.

2.4. Convergence of the iterative algorithm

The proof of convergence of the proposed algorithm is based on the proofs of convergence of the iterative hard thresholding [18] and the iterative soft thresholding (IST) [17] algorithms. Due to lack of space, we only include a short sketch of the proof here.

Proposition 1. *The algorithm converges to a fixed point, or equivalently, to a local minimum of the problem.*

Sketch of the proof. We consider two cases separately: in the first case, we suppose that there are infinitely many k such that $\mathbf{f}^{(k)}$ and $\mathbf{f}^{(k+1)}$ have different support; in the second case, we suppose that there is k_* such that for all $k > k_*$ the support of $\mathbf{f}^{(k)}$ is fixed.

The first case is treated in the same way as in [18], so we skip it here. In the second case, the algorithm reduces to a kind of iterative soft thresholding where ℓ_1 penalty is enforced only for one part of the vector. Since here we are considering only finite-dimensional setting, the same proof of convergence as presented in [17] is also valid in this case.

Table 1. Results of denoising for pure impulse noise in terms of the SSIM measure.

		[3]	PARIGI [9]	proposed
Lena	0.4	0.86	**0.91**	0.88
	0.6	0.73	**0.83**	0.81
Bridge	0.4	0.75	**0.76**	**0.76**
	0.6	0.58	0.55	**0.59**
Baboon	0.4	0.63	0.63	**0.69**
	0.6	**0.50**	0.46	**0.50**
Barbara	0.4	0.75	**0.92**	0.81
	0.6	0.60	**0.82**	0.68
Cameraman	0.4	0.80	**0.93**	0.92
	0.6	0.57	**0.86**	0.75
Boat	0.4	0.82	**0.83**	0.82
	0.6	0.67	0.70	**0.71**
Peppers	0.4	0.86	**0.88**	0.86
	0.6	0.70	**0.82**	0.80
Goldhill	0.4	**0.84**	**0.84**	**0.84**
	0.6	0.68	**0.73**	**0.73**

In [17], a weighted ℓ_1 norm $\sum_i w_i\left|x_i\right|$ in (1) was used, with the assumption $w_i > 0$ for all i. However, Lemma 3.6 in [17] is the only place where the assumption $w_i > 0$, for all i, was used; however, that lemma is trivially true in the finite-dimensional case even without that assumption. Therefore, the algorithm converges to a fixed point. □

Of course, it is known that the convergence of the IST can be very slow (accelerated version of IST was presented in [16]). While it is not obvious how to improve the speed of the proposed algorithm, another possible approach to solving (1) could be to use *alternating* optimization (with respect to $\mathbf{x}$ and $\mathbf{f}$). Recently, a general alternating optimization framework (which applies to formulation (1)) was presented in [19], with established convergence guarantees[1]. In our simulations, this approach did not bring performance improvement.

3. EXPERIMENTS

We have compared the approach described in this paper with recent state-of-the-art methods from [9] (named PARIGI) and [3] described in the Introduction.

For measuring the quality of reconstructed images, we have used Peak Signal-to-Noise Ratio (PSNR) and Structural SIMilarity index (SSIM) [20]. It was demonstrated that the SSIM is a metric that better corresponds to subjective quality of visual perception.

The methods were tested on test images also used in [9] and provided on their website[2]. For every image, 10 different

[1] We thank the anonymous reviewer for pointing out this paper

[2] http://perso.telecom-paristech.fr/~delon/Demos/Impulse/

Table 2. Results of denoising for mixed Gaussian-impulse noise in terms of the PSNR measure. For method [3], only the results reported in that paper are included (see text).

		[3]	PARIGI [9]	proposed
Lena	$p = 0.1$, $\sigma = 5$	34.98	34.72	**35.00**
	$p = 0.3$, $\sigma = 5$	32.04	**32.57**	31.03
	$p = 0.1$, $\sigma = 15$	**30.85**	30.31	30.76
	$p = 0.3$, $\sigma = 15$	29.11	**29.22**	29.15
Bridge	$p = 0.1$, $\sigma = 5$	-	26.96	**28.56**
	$p = 0.3$, $\sigma = 5$	-	**25.45**	25.26
	$p = 0.1$, $\sigma = 15$	-	25.34	**25.82**
	$p = 0.3$, $\sigma = 15$	-	23.38	**23.90**
Baboon	$p = 0.1$, $\sigma = 5$	-	24.81	**25.11**
	$p = 0.3$, $\sigma = 5$	-	**23.05**	22.55
	$p = 0.1$, $\sigma = 15$	-	**23.63**	23.15
	$p = 0.3$, $\sigma = 15$	-	**21.81**	21.58
Barbara	$p = 0.1$, $\sigma = 5$	30.48	**31.55**	28.90
	$p = 0.3$, $\sigma = 5$	25.92	**29.28**	25.42
	$p = 0.1$, $\sigma = 15$	27.31	**28.80**	25.78
	$p = 0.3$, $\sigma = 15$	24.55	**27.33**	24.00
Cameraman	$p = 0.1$, $\sigma = 5$	-	34.98	**35.63**
	$p = 0.3$, $\sigma = 5$	-	**31.40**	30.08
	$p = 0.1$, $\sigma = 15$	-	30.33	**30.54**
	$p = 0.3$, $\sigma = 15$	-	**28.59**	28.33
Boat	$p = 0.1$, $\sigma = 5$	-	**31.41**	31.13
	$p = 0.3$, $\sigma = 5$	-	**28.81**	27.59
	$p = 0.1$, $\sigma = 15$	-	**28.21**	28.18
	$p = 0.3$, $\sigma = 15$	-	**26.57**	26.30
Peppers	$p = 0.1$, $\sigma = 5$	-	**33.90**	30.11
	$p = 0.3$, $\sigma = 5$	-	**32.38**	28.68
	$p = 0.1$, $\sigma = 15$	-	**30.28**	28.39
	$p = 0.3$, $\sigma = 15$	-	**29.39**	27.63
Goldhill	$p = 0.1$, $\sigma = 5$	-	32.60	**33.10**
	$p = 0.3$, $\sigma = 5$	-	**30.64**	29.85
	$p = 0.1$, $\sigma = 15$	-	29.08	**29.36**
	$p = 0.3$, $\sigma = 15$	-	**27.99**	27.92

realizations of random noise were generated (as also used in [9]), and the results presented in Tables 1 and 2 are mean values over these 10 realizations (it should be said that the variations in the results for different noise realizations were small). Noise parameters (fraction of pure impulse noise and standard deviation of Gaussian noise) were selected as in [9] to enable fair comparison of methods.

Details about the parameters of the proposed method are as follows. Dictionary was learned on a set of training images using the FastICA algorithm [21] (dictionary was complete, which was enough for good results). Patch size was set to 8×8 (possibly better results could be obtained with other sizes, but in our experiments patch size was fixed). In the first phase of the image restoration algorithm, an image was processed 4 times for decreasing sequence of thresholds in the ROAD detector (in this way, we hope to detect noisy pixels possibly missed by ROAD; on the other hand, selecting a small threshold in ROAD would possibly also "detect" pixels that are not noisy). In every step in the first phase, pixels which were declared as noisy were discarded in computation (through $\mathbf{\Omega}$ in (1)), and after processing the whole image, only these pixels were replaced. In the second phase, overlapping image patches were processed. Depending on the noise level, no impulse detection was performed, or a high threshold for the ROAD statistic was selected. After the second phase, all image pixels were replaced by the approximations of their true values obtained with the algorithm. Parameters λ_1 and λ_2 were chosen by coarse cross-validation and possibly slightly better results could be obtained by tuning them more carefully. It should be said that we assumed that a rough estimate of the fraction of impulse noise is available (the same assumption was also used in [9]), and it was used for setting the value of parameter λ_2. L was set to 1 because this value resulted in faster convergence of the algorithm, despite the theoretical value being $L = 2$. The number of iterations was set to 3000. Average time elapsed for the algorithm was about 30 minutes per image; it should be noted that the algorithm could be parallelized which would make it much faster (the time mentioned above is the result of our "naive" implementation). Tables 1 and 2 summarize the results. For method [3] in the case of mixed Gaussian-impulse, only the results reported in that paper are included, since with our implementation we could not find the setting of the parameters that works well.

The method presented here obtained, for some images, comparable results with other two methods for pure impulse noise. For mixed Gaussian-impulse noise, the proposed method was generally inferior. However, the proposed method gave good results (in terms of SSIM) for images rich with fine details (Bridge, Baboon, Boat and Goldhill). Better results for specific images could possibly be obtained with more specialized dictionaries or by using several regularizations, as in [10, 22, 23]. It should be noted that none of the referenced papers presented comparisons with recent state-of-the-art method [9]. Also, in most papers (except [9]), only impulse noise fractions below 50 percent were used.

Images were not included in this paper due to lack of space, but all experiments can be reproduced using the code available at author's webpage[3].

4. CONCLUSIONS

We have presented a simple iterative mixed soft-hard thresholding algorithm for solving inverse problems with ℓ_0-ℓ_1 sparsity constraints, and its application to image restoration under impulse noise. Image regularization used in this paper is based on sparse representations in learned dictionary. Independent component analysis was used for dictionary learning purpose. Although other approaches discussed in the paper perform much better for some images and some problem settings, the approach proposed here performs at least slightly better for some images. This agrees with the statement from [9], namely that the performance of a denoising

[3] http://www.lair.irb.hr/ikopriva/marko-filipovi.html

method depends on the image (or texture) class. We also note that the algorithm presented here could possibly be applied to general inverse problems like robust regression.

Acknowledgements

The authors are very grateful to anonymous reviewers for their helpful comments.

REFERENCES

[1] A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian, "Practical poissonian-gaussian noise modeling and fitting for single-image raw-data," *Image Processing, IEEE Transactions on*, vol. 17, no. 10, pp. 1737–1754, 2008.

[2] A. Jezierska, E. Chouzenoux, J.-C. Pesquet, and H. Talbot, "A convex approach for image restoration with exact poisson-gaussian likelihood," Tech. Rep., 2013.

[3] Y. Xiao, T. Zeng, J. Yu, and M. K. Ng, "Restoration of images corrupted by mixed gaussian-impulse noise via l1-l0 minimization," *Pattern Recognition*, vol. 44, no. 8, pp. 1708 – 1720, 2011.

[4] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *Image Processing, IEEE Transactions on*, vol. 15, no. 12, pp. 3736–3745, 2006.

[5] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *Image Processing, IEEE Transactions on*, vol. 17, no. 1, pp. 53–69, 2008.

[6] Jian-Feng Cai, Raymond H. Chan, and Mila Nikolova, "Fast two-phase image deblurring under impulse noise," *Journal of Mathematical Imaging and Vision*, vol. 36, no. 1, pp. 46–53, 2010.

[7] Jian-Feng Cai, Raymond H. Chan, and Mila Nikolova, "Two-phase approach for deblurring images corrupted by impulse plus gaussian noise," *Inverse Problems and Imaging*, vol. 2, no. 2, pp. 187–204, 2008.

[8] Yan-Ran Li, Lixin Shen, Dao-Qing Dai, and B.W. Suter, "Framelet algorithms for de-blurring images corrupted by impulse plus gaussian noise," *Image Processing, IEEE Transactions on*, vol. 20, no. 7, pp. 1822–1837, 2011.

[9] J. Delon and A. Desolneux, "A patch-based approach for removing impulse or mixed gaussian-impulse noise," *SIAM Journal on Imaging Sciences*, vol. 6, no. 2, pp. 1140–1174, 2013.

[10] M. Yan, "Restoration of images corrupted by impulse noise and mixed gaussian impulse noise using blind inpainting," *SIAM Journal on Imaging Sciences*, vol. 6, no. 3, pp. 1227–1245, 2013.

[11] Y. Wang, A. Szlam, and G. Lerman, "Robust locally linear analysis with applications to image denoising and blind inpainting," *SIAM Journal on Imaging Sciences*, vol. 6, no. 1, pp. 526–562, 2013.

[12] Guoshen Yu, G. Sapiro, and S. Mallat, "Image modeling and enhancement via structured sparse model selection," in *Image Processing (ICIP), 2010 17th IEEE International Conference on*, 2010, pp. 1641–1644.

[13] Weihong Guo, Jing Qin, and Wotao Yin, "A new detail-preserving regularity scheme," Tech. Rep., 2013.

[14] Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro, "Online learning for matrix factorization and sparse coding," *J. Mach. Learn. Res.*, vol. 11, pp. 19–60, Mar. 2010.

[15] R. Garnett, T. Huegerich, C. Chui, and Wenjie He, "A universal noise removal algorithm with an impulse detector," *Image Processing, IEEE Transactions on*, vol. 14, no. 11, pp. 1747–1754, 2005.

[16] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[17] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, pp. 1413–1457, 2004.

[18] Thomas Blumensath and Mike E. Davies, "Iterative thresholding for sparse approximations," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5-6, pp. 629–654, 2008.

[19] Jérôme Bolte, Shoham Sabach, and Marc Teboulle, "Proximal alternating linearized minimization for non-convex and nonsmooth problems," *Mathematical Programming*, pp. 1–36, 2013.

[20] Zhou Wang and A.C. Bovik, "Mean squared error: Love it or leave it? a new look at signal fidelity measures," *Signal Processing Magazine, IEEE*, vol. 26, no. 1, pp. 98–117, 2009.

[21] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. on Neural Netw.*, vol. 10, no. 3, pp. 626 – 634, 1999.

[22] L. Ma, J. Yu, and T. Zeng, "Sparse representation prior and total variation–based image deblurring under impulse noise," *SIAM Journal on Imaging Sciences*, vol. 6, no. 4, pp. 2258–2284, 2013.

[23] Gabriel Peyré, Jalal Fadili, and Jean-Luc Starck, "Learning adapted dictionaries for geometry and texture separation," 2007.