



COMPRESSED SENSING BASED DENOISING FOR BIOMEDICAL IMAGES

Final report

Diana Mandache

January 2017

Contents

1 Introduction

The topic of image denoising is one of high interest in the image processing area, notably in the domains of biology and medicine where the acquisition techniques inflict many ambiguities on the images. Several authors published their research results on this subject. As an example, *IEEEExplore* provides about 9000 results for the search term “image denoising”.

This paper is aimed to be a bibliographic study on biological image denoising, based on recent journal and conference articles and books. It is not an overview of all the denoising algorithms reported by the literature, but rather an inquiry on a few more modern techniques with fresh expansion potential. Based on this paper, a practical project that implements and extends the algorithms presented will be conducted. The report is structured as follows: firstly, a short introductory summary on the techniques of denoising and compressed sensing, followed by a review on different optimization methods used in image reconstruction, then is presented the way the Fourier transform is employed as a measurement basis and in the following section some interesting flavors are given on the applicability of multiple CS reconstructions and fusions, lastly it is presented the implementation of the algorithm.

2 Denoising

Noise is caused by an abnormal fluctuation of the signal resulting in unwanted variations of color, brightness and/or contrast in the image. There are different causes for the presence of noise. As a result there are also many types of noise. To give a few examples, there is shot (or photon) noise that can be modeled with a Poisson function, there is thermal noise which respects a Gaussian distribution and is independent of the signal intensity. Noise is caused by the acquisition device and conditions; in the case of biological images obtained from a microscope there is present a mixed Poisson-Gaussian noise [1].

The concept behind image denoising is, being given an image y damaged by noise, to restore it to the highest possible fidelity to the original signal x , so that $y = \Phi x$, where Φ is the measurement basis, x is called the ground truth image, while \hat{x} is the predicted image obtained after denoising. Ultimately, image denoising can be viewed as recovering a signal from inaccurate or incomplete measurements, which is exactly what compressed sensing does.

3 Compressed Sensing

It is well known from the theory of digital signal processing, according to Nyquist-Shannon sampling theorem, that a sequence of discrete values extracted from the original analog signal with a certain sampling rate f_s (number of samples), allows reproducing the original signal with the bandwidth bounded to the half of the sampling rate ($\frac{f_s}{2}$) with no errors [12].

If the Nyquist-Shannon condition cannot be achieved because of the limitations caused by constructive parameters of the optical sensors, i.e. the samples are taken at a slower rate than twice the band limit (it is called sub-sampling), then the newer theory of compressive sampling demonstrates that the original signal can be reconstructed with acceptable errors.

Compressive sampling, also known as compressed sensing (CS) is a ten year old method proposed by [10], this theory provides a new approach for signal acquisition stating that a signal can be exactly reconstructed using a significantly small number of random linear measurements, under certain sparsity conditions. Since most signals are indeed compressible in some transform domains, CS has attracted a lot of attention in many applications, including medical and biological imaging due to its potential of reducing the sampling rates, power consumption and computation complexity in the image acquisition.

Given the signal $x \in \Re^N$ and M samples, such as $M \ll N$, we obtain the observation $y \in \Re^M$, with respect to $y = \Phi x$, where Φ is a random $M \times N$ orthonormal matrix called measurement basis and the sub-sampling rate is $\tau = M/N$, this ranges between 0 and 1 and intuitively implies the percentage

of the signal that was actually observed. In practice a sub-sampling rate between 10% and 50% allows good signal reconstruction.

Although the number of unknowns is larger than the number of measurements, the signal can still be exactly recovered if it is sparse enough [10]; this is the fundamental concept of CS theory. Sparsity is the key aspect that allows recovery x from y , it can be achieved by applying a sparsity transform to x , this consists in assuming that x has a sparse representation in some known basis Ψ , called sparsity basis or dictionary, such that Ψx has most of its coefficients zero.

4 Convex Optimization

4.1 l_0, l_1 and l_2 optimization

According to the definition given in the previous section the reconstruction of the signal x is the sparsest signal (one with most null values) that can be found with respect to the constraints imposed by acquisition. Therefore the following optimization problem arises:

$$\hat{x} = \arg \min \|\Psi x\|_0 \quad \text{subject to} \quad y = \Phi x \quad (1)$$

This approach has major drawbacks in terms of computational complexity, due to the lack of l_0 -norm's mathematical representation l_0 -minimization is solved by an NP-hard greedy algorithm called Matching Pursuit.

In order to try and resolve the computational obstacle, l_0 -norm, one may intuit to replace it by l_2 -norm, since the l_2 -optimization is easy to solve using the famous algorithm of Least Square Regression. However, even though its solution is easy to compute, it will not necessarily be the best solution. Because of the smooth nature of l_2 -norm, it is hard to find a single, best solution for the problem. See Fig. 1 for a visual comparison between the three norms.

As a tradeoff between l_0 and l_2 norms, l_1 optimization is widely used in the more recent years. Its computational complexity is also high (the solution

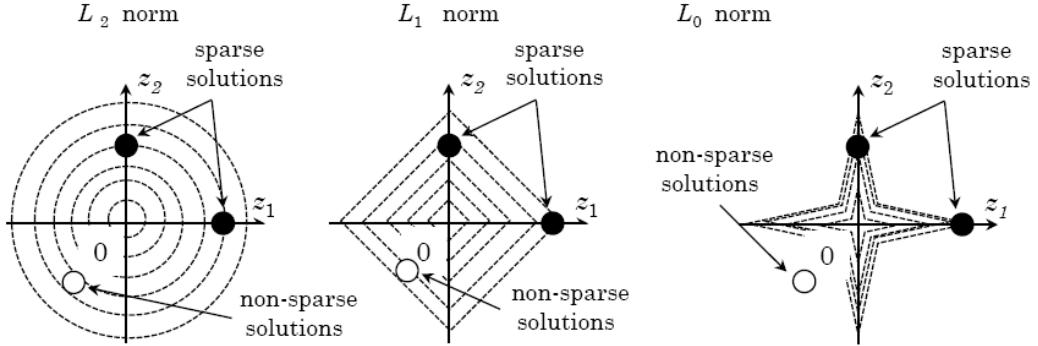


Figure 1: **Solution space of l_2, l_1 and l_0 for a 2D vector (z_1, z_2)** [11]

has to be found among an infinitely large space of solutions), but under the right conditions it can find the sparsest solution. One classical algorithm for solving this problem is Basis Pursuit. Hence, it has come to this optimization problem:

$$\hat{x} = \arg \min \|\Psi x\|_1 \quad \text{subject to} \quad y = \Phi x \quad (2)$$

Further we try to relax the constraints imposed on the observation on the hypothesis that the observation may not be perfectly accurate, for instance, it can be exposed to noise. Under this hypothesis the observation vector becomes $y = \Phi x + b$, where b is an additive noise such that $\|b\|_2 \leq \epsilon$. Given this new constraint, the optimization problem from Eq. 2 becomes:

$$\hat{x} = \arg \min \|\Psi x\|_1 \quad \text{subject to} \quad \|\Phi x - y\|_2 \leq \epsilon \quad (3)$$

4.2 Total variation optimization

Another further relaxation was introduced inspired by the fact that most of image functions are sparse or compressible in a certain domain, like the time domain (gradient) [8] or frequency domain (wavelet, Fourier). The property of gradient sparsity can be enforced on signal x , hence removing Ψ from the problem formulation. This idea has been exploited by replacing the l_1 norm

with the total variation (TV) norm defined as:

$$\|x\|_{TV} = \sum_{p,q} \sqrt{[x(p+1, q) - x(p, q)]^2 + [x(p, q+1) - x(p, q)]^2} \quad (4)$$

which is equivalent to:

$$\|x\|_{TV} = \sum_{p,q} \sqrt{\partial_h x(p, q)^2 + \partial_v x(p, q)^2} \quad (5)$$

where $\partial_h x$ and $\partial_v x$ are the partial horizontal and vertical derivatives of the image x and p, q are pixel coordinates.

A new approach to the optimization problem emerged, reconstructing the original signal by minimizing the $\|x\|_{TV}$ operator:

$$\hat{x} = \arg \min \|x\|_{TV} \quad \text{subject to} \quad \|\Phi x - y\|_2 \leq \epsilon \quad (6)$$

This approach is best suited for piecewise smooth functions [7], since this is the a priori assumption made on the signal. As for biological imaging, it works best for simple cells and it is fairly inconvenient for complex tissues [1].

It gained popularity due to its relatively easy computation (especially compared to computing $\|\Psi x\|_1$) and the sharpness of the reconstruction, property that can easily be deduced from the operator's definition.

4.3 NESTA algorithm

The optimization algorithm that was used in the paper [1] that represents the focal point of this bibliographic research report is the NESTA algorithm. It was introduced in [5] and it is based on Nesterov's work on minimizing non-smooth functions. It can solve both l_1 and TV problems without specific requirements on the sparsity matrix Ψ . Based on [7], we observe that NESTA is a fast, robust and adaptable algorithm that consists in an accelerated gradient descent with back-projection. In NESTA the parameter ϵ from Eq. 3 is defined as:

$$\epsilon = \sigma \sqrt{\tau N + 2\sqrt{2\tau N}} \quad (7)$$

where σ is the standard deviation of the Gaussian noise and τ is the sub-sampling rate.

4.4 FISTA algorithm

Fast Iterative Shrinkage/Thresholding Algorithm (FISTA) is a gradient-based method proposed by [6] which minimizes a sum of two convex functions: $\hat{x} = \arg \min f(x) + g(x)$, where f is smooth and g nonsmooth. For our problem formulation FISTA solves the optimization problem defined in Eq.8 where $\lambda > 0$ is the regularization parameter which provides a tradeoff between fidelity to measurements and noise sensitivity.

$$\hat{x} = \arg \min \|\Phi x - y\|_2^2 + 2\lambda\|x\|_{TV} \quad (8)$$

The authors name the case where Φ is the identity matrix the denoising problem and the opposite case is called the deblurring problem. In practice, solving the deblurring problem requires intermediary solutions of the denoising problems, this aspect will be further detailed in the Implementation section of this paper.

For this project the FISTA algorithm was coded in Java, with Φ being the Fast Fourier Transform (FFT) and Φ^T the Inverse Fast Fourier Transform (IFFT), respectively.

5 Fourier Transform

5.1 FT as observation

A common choice in the literature for the projection matrix Φ is the Fourier transform (FT) because of its properties and convenience in multiple applications [12]. Moreover, MRI images are acquired directly in the Fourier space. Noticeably CS theory has improved the field of medical imaging, where it has enabled speedups by a factor of seven in pediatric MRI while preserving diagnostic quality. As for biological images, the Fourier coefficients from images acquired with a microscope can be measured via digital holography or optical Fourier Transform.

5.2 Sampling patterns in FT

As illustrated in the simple yet eloquent experiment in [4], randomly uniform sampling in the Fourier domain is far from being optimal. On the other hand, the most of high spatial energy seems to be concentrated in the low frequency area of the Fourier domain as illustrated in Fig. 2.

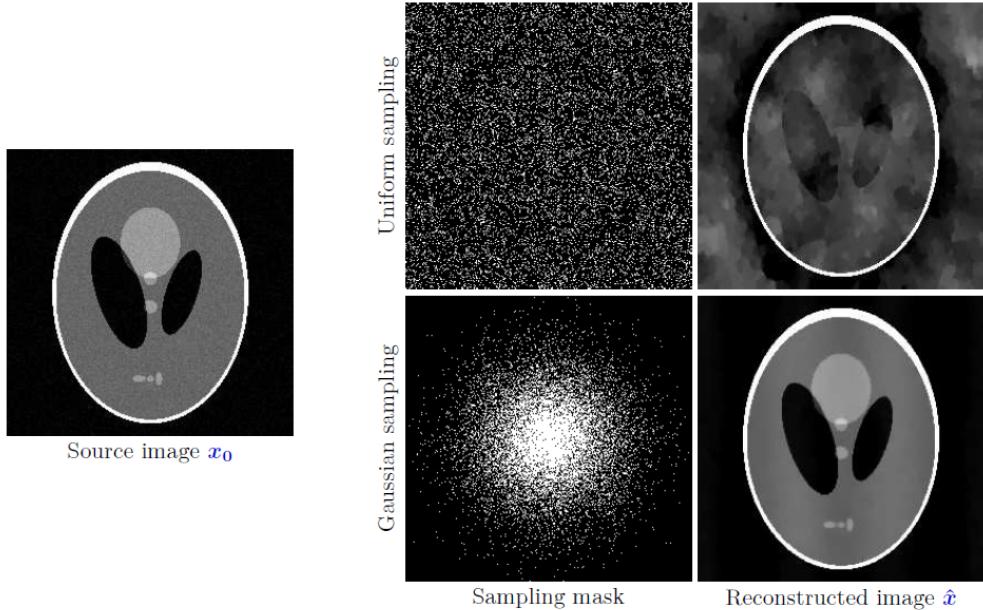


Figure 2: Two reconstructions of the same noisy version of the Shepp-Logan phantom image, using two different sampling masks in the Fourier domain (sub-sampling rate $\tau = 0.15$) [4]

Even though the Gaussian and polynomial sampling masks are widely used, in [4] is presented a new approach especially suitable for biological images. By fixing a cut-off frequency v_c , all the coefficients below this frequency threshold are kept and coefficients in the higher frequencies are randomly picked with respect to the global sub-sampling rate τ . As stated in [1] the parameter v_c represents the radius of the fully sampled area, which is aimed to contain 90% of the total energy in the spatial domain; however, this corresponds to a much smaller number of coefficients (from 3% up to 20%) in the Fourier domain.

Allegedly, depending of the type of application developed and the a priori information on the type of input, as well as the expected output, these sampling patterns can be customized in different ways. For instance, in [3], the problem tackled is the simplification (in the purpose of compression) of Spectral Domain Optical Coherence Tomography (SD-OCT) images of cardiac tissue. Consequently, knowing the particularities of these types of medical images, namely that cardiac tissue layers are horizontal and somewhat parallel, a suitable sampling pattern can be chosen. By applying a Hough transform on the edges, the most probable direction is chosen. The Hough transform takes a binary image and detects the directions of the straight edges. For every point in the spatial domain there is a curve in the Hough domain that represents all the possible lines that can pass through that point [9], an intersection of two curves from the Hough space represents a line. In the situation presented, the direction chosen corresponds to the point in the Hough space where the most curves join. From this direction a star pattern is obtained in the Fourier domain, by adding equally dispersed branches up to a total number given. Subsequently, random coefficients are sampled up to reaching the sub-sampling rate τ .

Hence, partially random patterns in the Fourier transform matrix with higher sampling density in the low frequency area of the domain are advised and will be employed in this project.

6 Multiple CS Reconstructions

6.1 Multiple measurements and reconstructions

Given a certain problem formulation, one may desire to acquire high reconstruction quality while simultaneously maintaining a very low sub-sampling rate. Inspired by video reconstruction, [1] and [3] combine multiple suboptimal reconstructions to obtain a high quality denoised image.

Given a noisy image y and its Fourier transform Φ , a number of R observations y_k can be generated from zeroing coefficients of Φ with respect to the

chosen sub-sampling rate τ and the preferred sampling pattern (as presented in 5.2):

$$y_k = \Phi_k y \quad \text{with} \quad k \in \{1, \dots, R\} \quad (9)$$

From each partial measurement y_k is computed a partial prediction \hat{x}_k of the ground truth image x using TV regularization method presented in 4.2. All the local predictions are assembled into a final prediction \hat{x} by means of a custom fusion function f .

6.2 Fusion functions

A straightforward approach for the fusion function is averaging the partial estimators, nevertheless, as acknowledged in [1], this method does not perform as desired in preserving the quality of the edges in the image.

$$\hat{x}_{mean} = \frac{1}{R} \sum_{k=1}^R \hat{x}_k \quad (10)$$

Other intuitive fusion operators are Median, Min, Max, Mode in neighborhood.

A more effective fusion function is proposed in [1], it amplifies the efficiency of the mean function by adding a new operator called *variance map* which acts as the standard deviation of the partial measures. This operator emphasizes the incoherence between the partial reconstructions which concentrate around the edges of the image, it is computed as follows:

$$\sigma_x = \sqrt{\frac{1}{R-1} \sum_{k=1}^R (\hat{x}_k - \hat{x}_{mean})^2} \quad (11)$$

The proposed method is called spatially-adaptive fusion and combines the two previously introduced operators:

$$\hat{x} = \sigma_x \circ y + (1 - \sigma_x) \circ \hat{x}_{mean} \quad (12)$$

where \circ is the Hadamard product (element-wise multiplication of matrices) and, in this case, σ_x represents the normalized variance map which acts as a

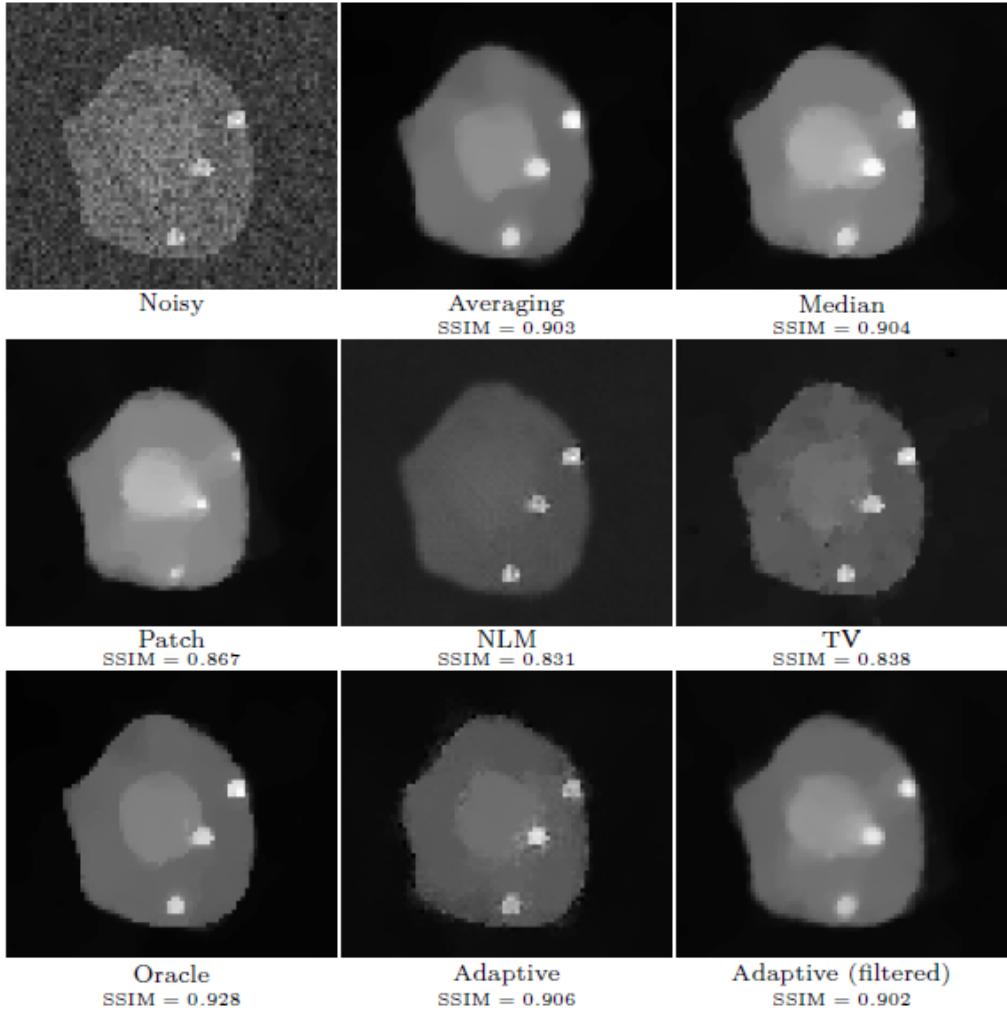


Figure 3: Denoising results on the synthetic cell image (perturbed with a mixed Poisson-Gaussian noise with parameters $\sigma = 0.1$ and $\lambda = 0.02$) obtained with different methods. [1]

weight matrix with values between 0 and 1; the higher weights being assigned to edge pixels.

Observing the good results obtained with this approach, which seems to be comparable with the more classical methods, it will be exploited and extended in the project that follows this paper. See results in Fig.3.

7 Implementation

After studying state of the art methods, there has been implemented a denoising plugin for the Icy platform.

7.1 ICY platform

Icy is a free and open source platform that comes with a wide range of functionalities presented in the form of different plug-ins. They serve for segmentation, feature detection, edge detection, tracking, optical flow computation or even machine learning. It also has an intuitive user interface that makes it very easy to use for its intended target user base that consist of biologist with little to no knowledge in computing.

The programming language that Icy is written in, i.e. Java, makes it available on all operating systems due to the Java Virtual Machine. Therefore it has a high degree of both usability and availability.

7.2 The algorithm

The proposed denoising algorithm, presented in the pseudocode below, takes as input arguments y the noisy image, R the number of reconstructions, τ the subsampling rate for every reconstruction, v the cutoff frequency and λ the smoothness parameter of FISTA.

The input arguments are constrained as follows: for best results y has to be a grayscale image, for colored images the output will be a greyscale image; R varies between 1 and 10 reconstructions; τ obviously ranges between 10% and 100%; v between 0 and 1 and λ can take values in the range 1 - 10.

It returns as output the denoised image \hat{x} . There are also visualized the sampling in the Fourier space and the variation map. Those can help the curious understand how the algorithm works and, on the other hand, they

can serve as a guide for the user on how to fine tune the parameters in order to get the best results.

```

input :  $y, R, \tau, v, \lambda$ 
output:  $\hat{x}$ 
begin
  for  $r \leftarrow 1$  to  $R$  do
     $mask[r] = generateSamplingMask(\tau, v)$ 
     $y_i[r] = FFTsampling(y, mask[r])$ 
     $x_i[r] = optimization(y_i[r], \lambda)$ 
  end
   $\hat{x} = fusion(x_i)$ 
  return  $\hat{x}$ 
end

```

Algorithm 1: Proposed algorithm

The sampling and fusioning functions respect the methods presented before in the sections 5.2 and 6.2, respectively.

The algorithm presented below solves the optimization in Eq.8 by calling at every iteration the algorithm of fast gradient projection $FGP(b, \lambda, N)$ whose implementation is described in detail in [6] and whose purpose is to return \hat{x} that minimizes $\|x - b\|_2^2 + 2\lambda\|x\|_{TV}$.

```

input :  $b, \lambda, N$ 
output:  $\hat{x}$ 
step 0:  $y^{(1)} = x^{(0)} = 0, L = 8\lambda$ 
step k: ( $1 \leq k < N$ )
   $x^{(k)} = FGP(y^{(k)} - \frac{2}{L}\Phi^T(\Phi y^{(k)} - b), \frac{2\lambda}{L}, N)$ 
   $t^{(k+1)} = \frac{1+\sqrt{1+4t^{(k)2}}}{2}$ 
   $y^{(k+1)} = x^{(k)} + \frac{t^{(k)}-1}{t^{(k+1)}}(x^{(k)} - x^{(k-1)})$ 

```

Algorithm 2: Optimization

Please note that for readability purposes the notation y from Eq.8 representing the noisy image has been changed to b in the pseudocode. We remind that Φ and Φ^T represent FFT and IFFT, respectively, while L is the step-

size. The stopping criterion of the algorithm is either reaching the maximum number of iterations N or getting an error smaller than a value $\epsilon = 0.0001$, the error is computed by subtracting the Frobenius norms of the results of the last two iterations.

7.3 Results

In this section there are shown the results of the method using two instances of the Shepp-Logan phantom image contaminated with different levels of noise, as well as on a HeLa cell microscopic image acquired in the laboratories of Pasteur Institute. For quality evaluation purposes the plugin was provided also with the ground truth image. For the synthetic cell, producing the ground truth was a straightforward task, but in the case of the real cell, the clean microscopic image, uncontaminated by noise, was obtained as a result of special acquisition conditions, like an abnormally long acquisition time. This way it can be computed the Peak Signal-to-Noise Ratio (PSNR) between the uncontaminated image x and the prediction \hat{x} obtained by the presented method. Both the ground truth and the noisy images can be seen in Fig.4 and Fig.5.

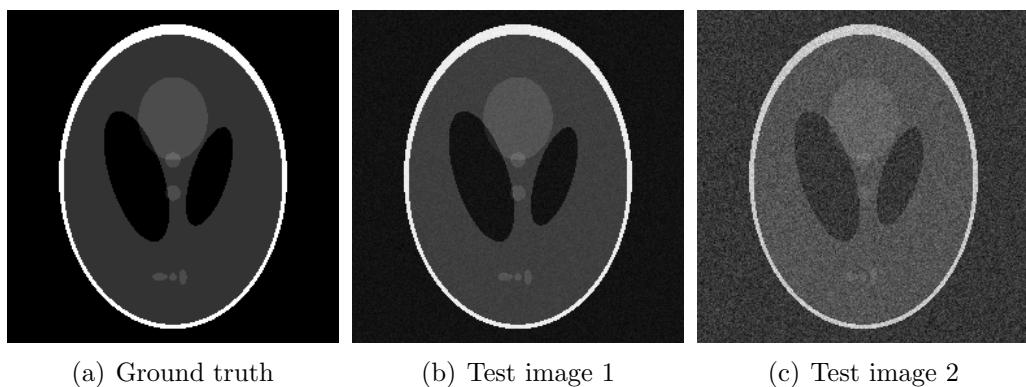


Figure 4: **Synthetic test data - *Shepp-Logan phantom cell***

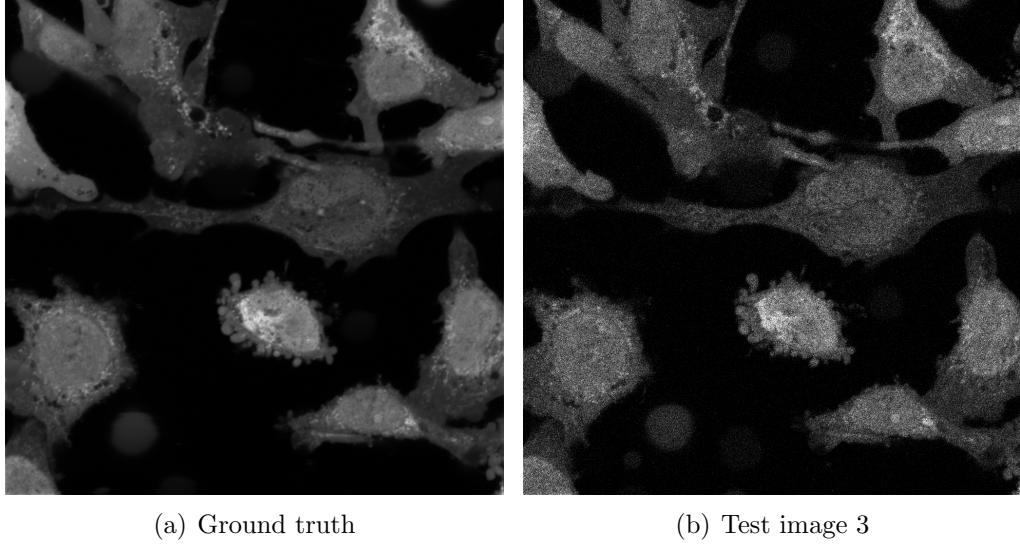
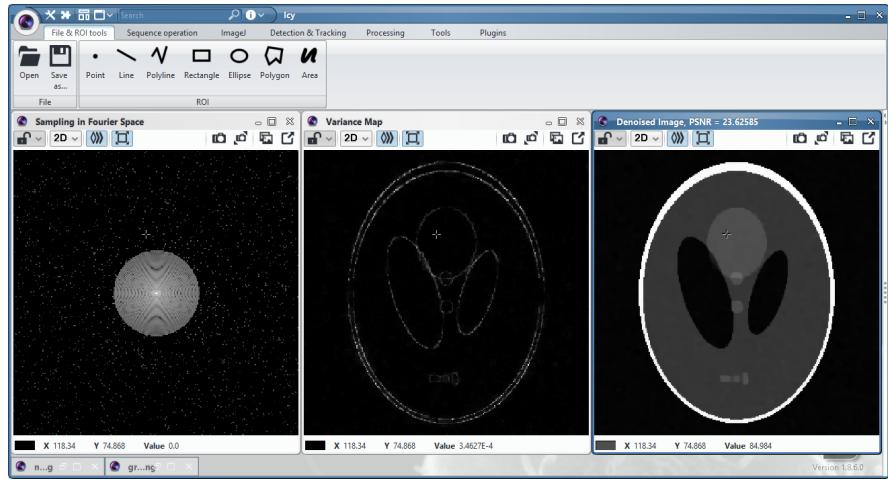


Figure 5: **Real test data - *HeLa cell***

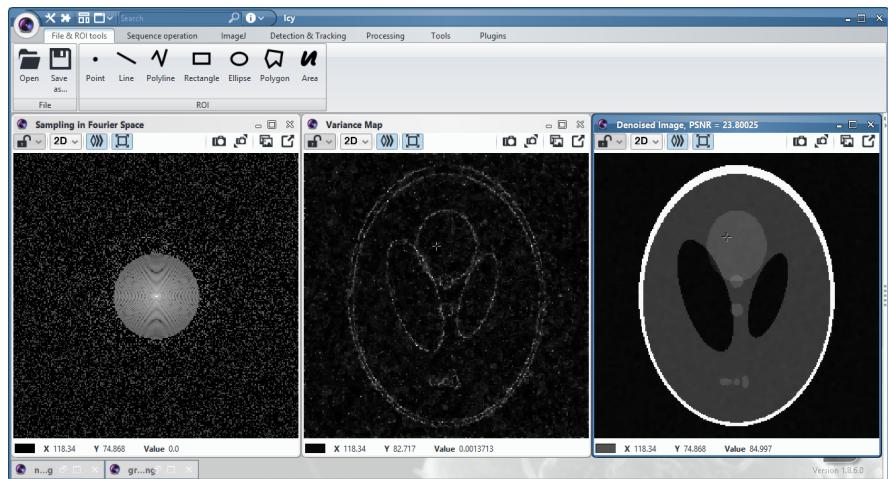
Based on the results obtained, it can be shown that this a robust method, mainly because there are obtained similar qualitative results for lower (10%) and higher (40%) subsampling rates, nevertheless, only a small portion of the signal is acquired by virtue of compressed sensing. What is more, it is actually advised to sample a sufficiency small number of coefficients in order not to get in the 'noisy area' of frequencies.

The randomly sampled Fourier coefficients (high frequency coefficients) can bring incoherence to the variance map, so, in some cases, better results are obtained from a low subsampling rate and a low cutoff frequency.

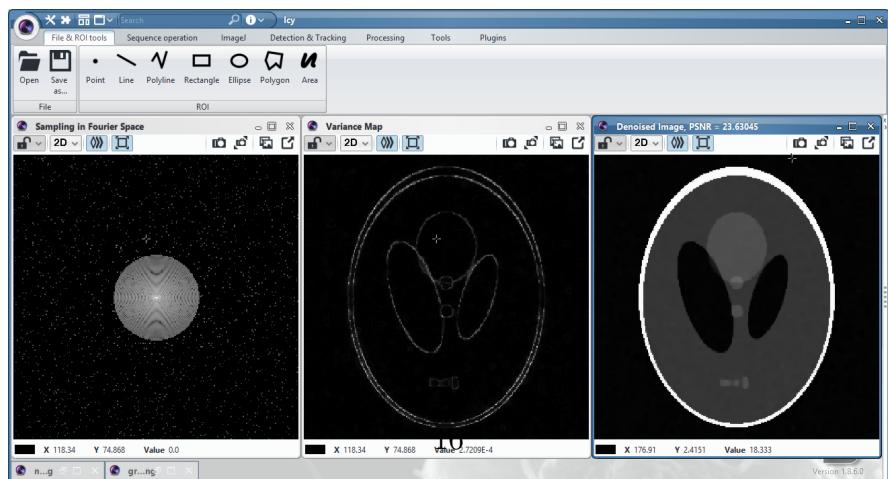
The smoothing parameter is particularly important in the case of very noisy images in order to reduce the artifacts. On the other hand, we notice that its influence in the quality of the reconstruction is not really backed up by the PSNR score, but the visual quality is noticeably improved.



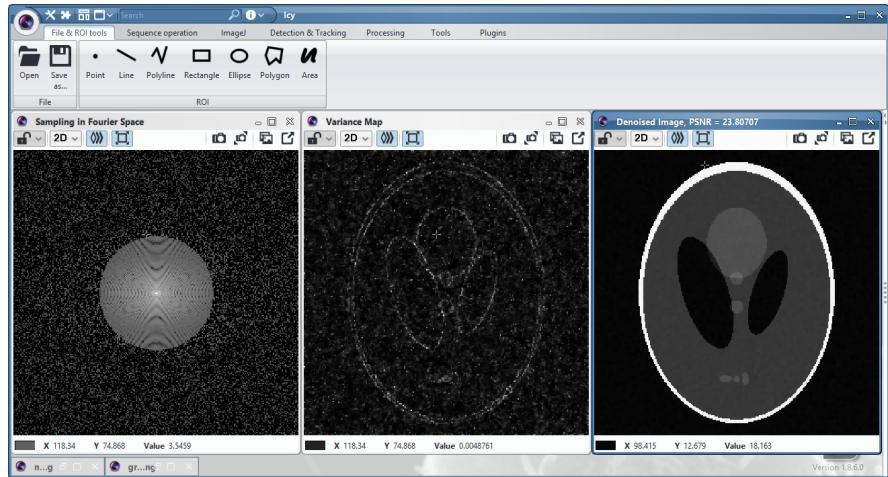
(a) $R = 4, \tau = 10\%, v = 0.3, \lambda = 1$



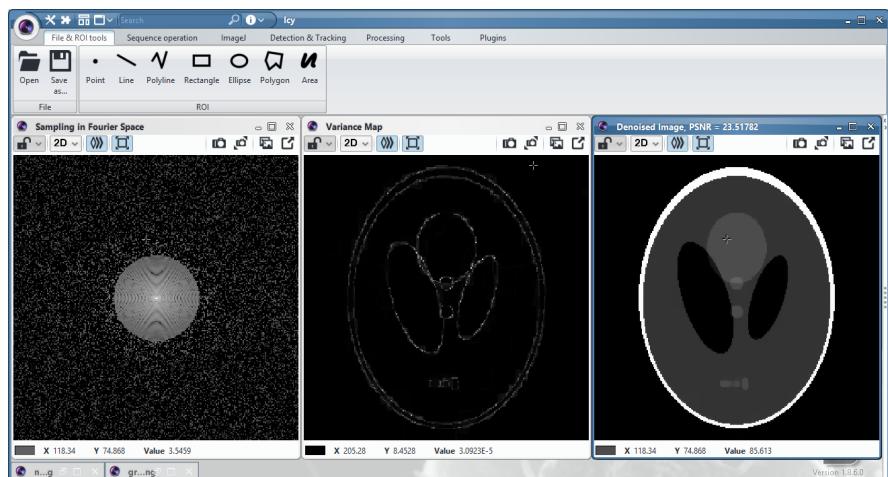
(b) $R = 4, \tau = 30\%, v = 0.3, \lambda = 1$



(c) $R = 10, \tau = 10\%, v = 0.3, \lambda = 1$

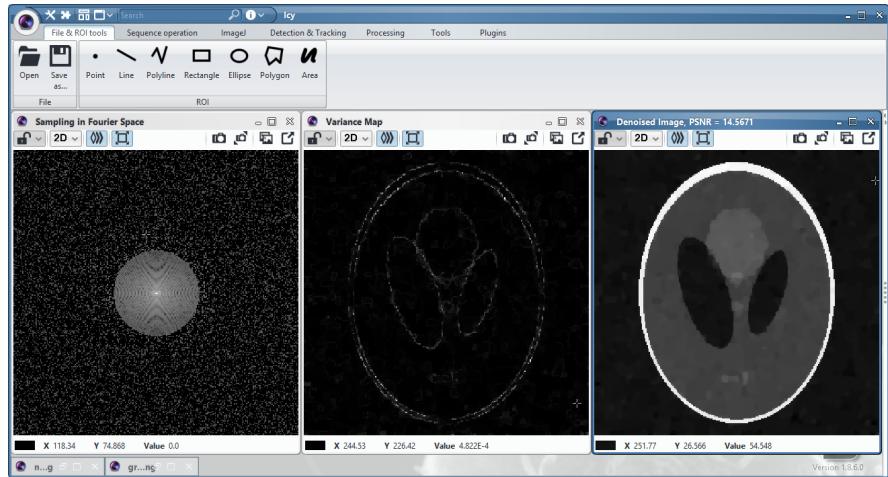


(d) $R = 4, \tau = 40\%, v = 0.4, \lambda = 1$

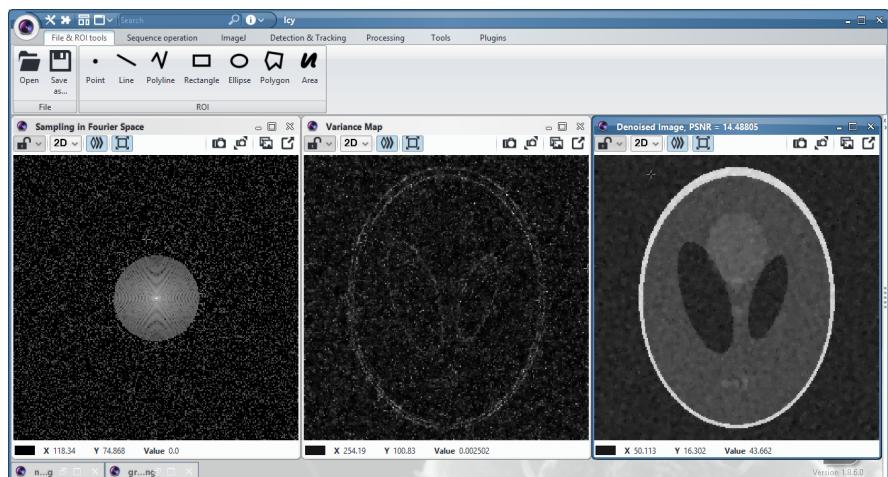


(e) $R = 4, \tau = 30\%, v = 0.3, \lambda = 5$

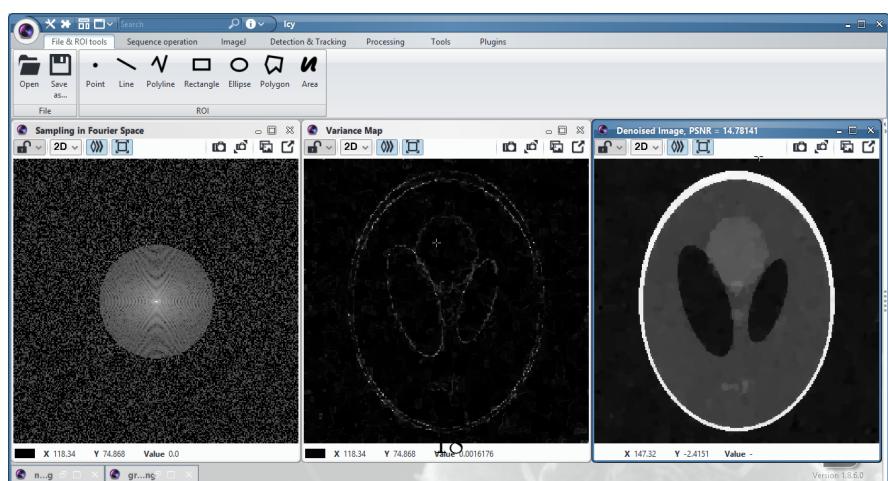
Figure 6: Test results on image 1
(sampling in Fourier domain, variance map, denoised image)



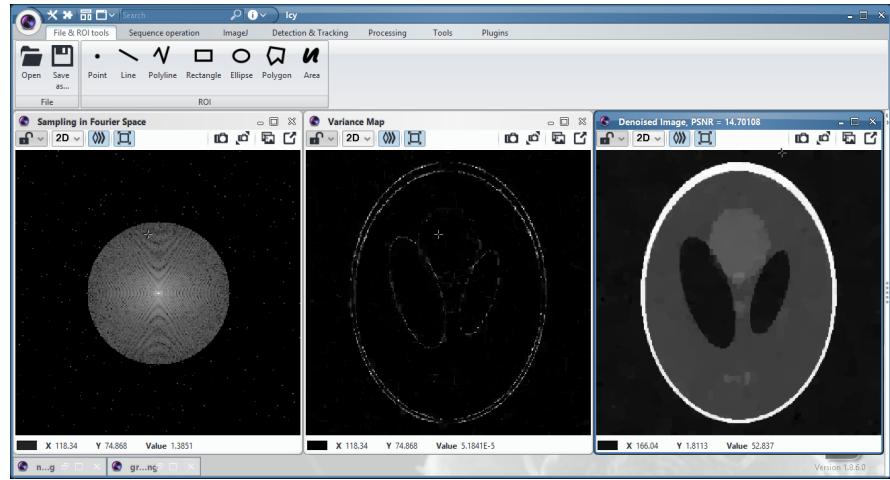
(a) $R = 4, \tau = 30\%, v = 0.3, \lambda = 7$



(b) $R = 4, \tau = 30\%, v = 0.3, \lambda = 2$

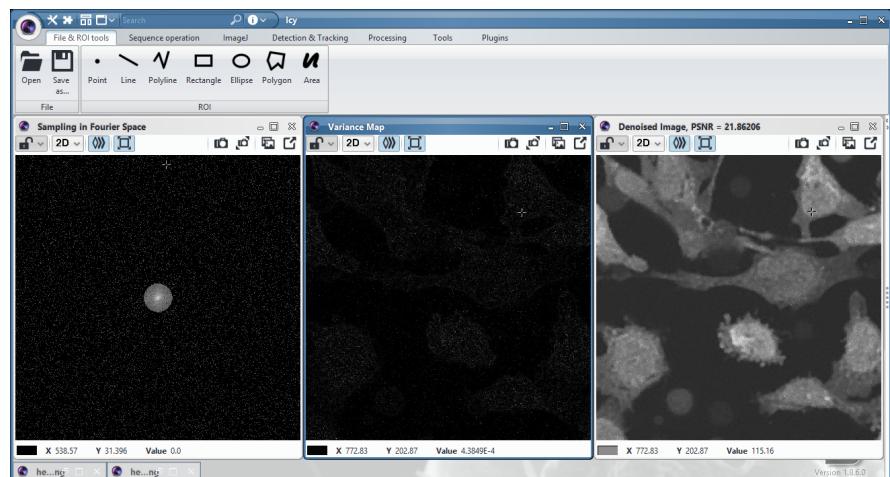


(c) $R = 4, \tau = 50\%, v = 0.4, \lambda = 9$

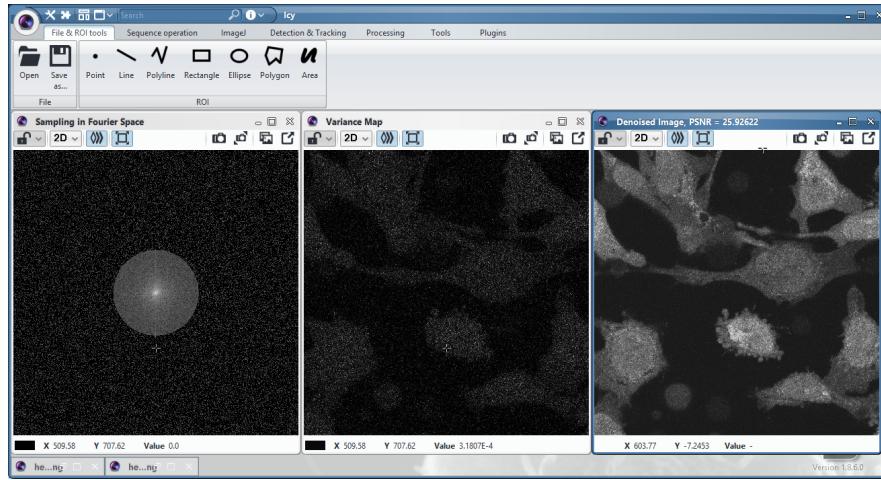


(d) $R = 4, \tau = 20\%, v = 0.5, \lambda = 10$

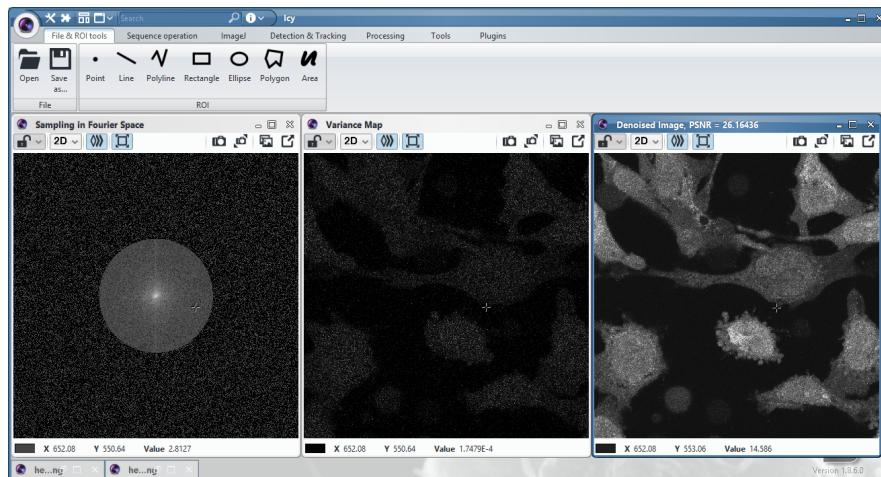
Figure 7: Test results on image 2
(sampling in Fourier domain, variance map, denoised image)



(a) $R = 2, \tau = 10\%, v = 0.1, \lambda = 2$



(b) $R = 2, \tau = 30\%, v = 0.3, \lambda = 1$



(c) $R = 2, \tau = 40\%, v = 0.4, \lambda = 2$

Figure 8: Test results on image 3
(sampling in Fourier domain, variance map, denoised image)

Table 1: **Test results from parameter tuning**

	R	$\tau(\%)$	$v(Hz)$	λ	PSNR(dB)
Test image 1	4	10	0.3	1	23.62585
	4	10	0.3	1	23.80025
	10	10	0.3	1	23.63045
	4	40	0.4	1	23.80707
	4	30	0.3	5	23.51782
Test image 2	4	30	0.3	7	14.5671
	4	30	0.3	2	14.48805
	4	50	0.4	9	14.78141
	4	20	0.5	10	14.70108
Test image 3	2	10	0.1	2	21.86206
	2	30	0.3	1	25.92622
	2	40	0.4	2	26.1646

8 Conclusion

This paper documented my efforts in both literature review and successful implementation of a researched method for biological image denoising, that makes use of the acclaimed theory of compressed sensing with TV regularization. The final result is a ready-to-use plugin that is highly accessible and user friendly due to its integration with the Icy platform. It is also very flexible due to the multiple parameters the user can tune depending on the image that he wishes to process.

For future development of the plugin, I intend to add another degree of freedom for the algorithm, i.e. multiple sampling patterns for the user to choose from. Those might be Gaussian sampling and fully random sampling.

References

- [1] W. Meinier, E. Angelini and J.-C. Olivo-Marin. *Image denoising by adaptive compressed sensing reconstructions and fusions*, in Proc. SPIE 9597, September 2015.
- [2] W. Meinier, Y. Le Montagner, E. Angelini and J. C. Olivo-Marin. *Image denoising by multiple compressed sensing reconstructions*, 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), New York, NY, 2015, pp. 1232-1235.
- [3] W. Meinier, Y. Can, C. P. Hendon, J. C. Olivo-Mariri, A. Laine and E. D. Angelini. *Sparsity-based simplification of spectral-domain optical coherence tomography images of cardiac samples*, 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), Prague, 2016, pp. 373-376.
- [4] Y. LeMontagner. *Algorithmic solutions toward applications of compressed sensing for optical imaging*, Ph.D. thesis, Télécom ParisTech, 2013.
- [5] Stephen Becker, Jérôme Bobin, and Emmanuel J. Candès. *NESTA: A Fast and Accurate First-Order Method for Sparse Recovery*, SIAM Journal on Imaging Sciences, 2011.
- [6] A. Beck and M. Teboulle. *Fast Gradient-Based Algorithms for Constrained Total Variation Image Denoising and Deblurring Problems*, IEEE Transactions on Image Processing, vol. 18, no. 11, pp. 2419-2434, November 2009.
- [7] Y. Le Montagner, E. Angelini and J. C. Olivo-Marin. *Comparison of reconstruction algorithms in compressed sensing applied to biological imaging*, 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Chicago, IL, 2011, pp. 105-108.
- [8] Leonid I. Rudin, Stanley Osher, Emad Fatemi. *Nonlinear total variation based noise removal algorithms*, Physica D Nonlinear Phenomena 60, November 1992.
- [9] C. Galambos, J. Kittler, and J. Matas. *Progressive probabilistic hough transform for line detection*, IEEE Computer Society Conference, Los Alamitos, CA, 1999.

- [10] E. J. Candès. *Compressive sampling*, Proceedings of the International Congress of Mathematicians, Madrid, Spain, 2006.
- [11] Hiroyuki Kudo, Taizo Suzuki, Essam A. Rashed. *Image reconstruction for sparse-view CT and interior CT—introduction to compressed sensing and differentiated backprojection*, Quant Imaging Med Surg 2013, pp. 147-161.
- [12] Alan V. Oppenheim, Ronald W. Schafer, and John R. Buck. *Discrete-Time Signal Processing (2nd Ed.)*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1999.