# Sparsity Based Poisson Denoising

Raja Giryes and Michael Elad
Department of Computer Science, The Technion - Israel Institute of Technology
Haifa, 32000, Israel
{raja,elad}@cs.technion.ac.il

***Abstract*—Sparsity based techniques have been widely used for image denoising. In this work we focus on Poisson noise and propose initial stages for a new strategy for its removal. We start with a method that removes the noise by converting it into an additive Gaussian noise using the Anscombe transform, applying a variant of the OMP-denoising algorithm. Then, following the recent work by Salmon *et. al.*, we bypass the need for the Anscombe transform and rely directly on the noise statistics. The new strategy is shown to lead to near state-of-the-art results.**

## I. Introduction

In the last decade, the notion of sparsity has been widely used for image denoising [1]. The commonly used model is $\mathbf{y} = \mathbf{x} + \mathbf{e}$, where $\mathbf{y} \in \mathbb{R}^d$ is the noisy image; $\mathbf{x} \in \mathbb{R}^d$ is the original clean image, which we aim to recover, assumed to have a sparse representation $\boldsymbol{\alpha}$ under a given dictionary $\mathbf{D} \in \mathbb{R}^{d\times n}$; and $\mathbf{e}$ is an additive noise that can be either adversarial or with a known random distribution (usually Gaussian). Though covering many types, not all types of noise obey this formulation.

One such type is the Poisson distributed noise. Each sample $\mathbf{y}_i$ is Poisson distributed random variable with mean (and variance) $\mathbf{x}_i$. Such noise occur in many applications such as night vision, computed tomography (CT), fluorescence microscopy, astronomy, astrophysics and spectral imaging.

Many techniques [2], [3], [4] developed for this case use a transformation, such as the Anscombe [5] and Fisz [6] transforms, that approximately convert the Poisson noise into a Gaussian one for which many possible denoising algorithms are already available.

However, such approximations of the noise is efficient only when we have a high photon count [4], [7], [8], i.e., the noise level is low. In the case of low photon count, i.e., high noise level, the measurements are mostly either zero or one indicating whether a photon was recorded in the receiver or not. An example of such a noisy image is given in several figures in this paper.

One approach for dealing with this problem is improving the inverse Anscombe transform leading to better recovery results [4]. Another approach is to bypass the use of the transformation by directly relying on the Poisson noise statistics [8] . State-of-the-art results have been obtained using non-local PCA for patches of the noisy image, and reliance on the statistical properties of the noise [7], [8].

In this work we present the first stages for an alternative Poisson denoising strategy relying on the OMP-denoising technique [9]. We start with an Anscombe based algorithm that relies on a variant of simultaneous OMP (S-OMP) for joint sparsity [10] and the refined inversion scheme for the Anscombe transform [4]. Then we remove the need for the transformation by using a new model for the image that relies on the noise statistics as in [8]. In this model the original signal is assumed to have a sparse representation $\boldsymbol{\alpha}$ under a given dictionary $\mathbf{D}$ but instead of having $\mathbf{x} = \mathbf{D}\boldsymbol{\alpha}$ we have $\mathbf{x} = \exp(\mathbf{D}\boldsymbol{\alpha})$, where the exponent is calculated element-wise. This model implies a new $\ell_0$-minimization problem and we introduce a greedy technique for its approximation.

The organization of the paper is as follows. In Section II we describe the Poisson denoising problem with more details and present the previous work. In Section III we propose the two sparsity-based algorithms for Poisson denoising, and in Section IV we demonstrate their denoising performance with a comparison to other techniques. In Section V we conclude our work and discuss the next steps for it.

## II. Problem Setup

In the Poisson denoising setup the measurement $\mathbf{y}$ is a Poisson distributed noise with mean and variance $\mathbf{x}$, i.e,

$$P(\mathbf{y}[i]|\mathbf{x}[i]) = \begin{cases} \frac{(\mathbf{x}[i])^{\mathbf{y}_i}}{\mathbf{y}[i]!}\exp(-\mathbf{x}_i) & \mathbf{x}[i] > 0, \\ \delta_0(\mathbf{y}[i]) & \mathbf{x}[i] = 0, \end{cases} \tag{1}$$

where $\mathbf{x}[i]$ is the $i$-th element in $\mathbf{x}$ and $\delta_0$ is the Kronecker delta function. It is common to measure the noise power in terms of the peak of the signal. The peak value is defined as $\mathbf{x}_{\text{max}}$, where $\mathbf{x}_{\text{max}}$ is the maximal value in $\mathbf{x}$.

As mentioned before, a standard way for dealing with Poisson noise is using the Anscombe transform [5],

$$f_{\text{Anscombe}}(x) = 2\sqrt{x + \frac{3}{8}}. \tag{2}$$

This transformation converts a Poisson distributed data to an approximately Gaussian distributed data with variance 1. This property holds true whenever the mean of the Poisson data is greater than 4. After applying the transformation, any standard denoising technique for Gaussian noise can be applied on the data. Having the recovery result of the Gaussian denoising algorithm, the inverse Anscombe transform should be applied. However, this inverse transform introduces an undesired bias into the estimate. Because of this reason and the fact that in many cases the mean of the data is smaller then 4, more sophisticated techniques are desired for applying the inverse transform. One such technique is proposed in [4].

Another alternative, as posed in [7], [8], is to work directly with the Poisson data, removing the need for the transformation. By maximizing the log-likelihood of (1) and eliminating terms independent of $\mathbf{x}$ we get the following minimization problem

$$\underset{\mathbf{x}}{\operatorname{argmin}}\, \mathbf{1}^T\mathbf{x} - \mathbf{y}^T\log(\mathbf{x}), \tag{3}$$

where $\mathbf{1} \in \mathbb{R}^d$ is a vector composed of ones.

Using the standard sparsity model for $\mathbf{x}$ in (3), $\mathbf{x} = \mathbf{D}\boldsymbol{\alpha}$, leads to the need to add a non-negativity constraint on $\mathbf{x}$. In order to avoid this, we use a sparsity model for $\mathbf{x}$ that introduces non-negativity naturally. Following [8], we set $\mathbf{x} = \exp(\mathbf{D}\boldsymbol{\alpha})$ where $\boldsymbol{\alpha}$ is a $k$-sparse vector. With this setup we end up with the minimization problem

$$\underset{\boldsymbol{\alpha}}{\operatorname{argmin}}\, \mathbf{1}^T\exp(\mathbf{D}\boldsymbol{\alpha}) - \mathbf{y}^T\mathbf{D}\boldsymbol{\alpha} \quad s.t. \quad \|\boldsymbol{\alpha}\|_0 \le k \tag{4}$$

The work in [7], [8] applied a related model on the patches of the image, proposing the non-local PCA algorithm (NLPCA). In this method overlapping patches are extracted from the noisy image $\mathbf{y}$ and then clustered into a small number of disjoint groups $\{\mathbf{q}_{j,1}, \dots, \mathbf{q}_{j,p_j}\}$. Each group $j$ has its own dictionary $\mathbf{D}_j$ that contains only a few (much less then the patches dimension) columns. For each index $j$, the dictionary $\mathbf{D}_j$ and the representations $\{\boldsymbol{\alpha}_{j,1}, \dots, \boldsymbol{\alpha}_{j,p_j}\}$ of the patches $\{\mathbf{q}_{j,1}, \dots, \mathbf{q}_{j,p_j}\}$ are calculated by solving[1]

$$\begin{aligned}&[\mathbf{D}_j, \hat{\boldsymbol{\alpha}}_{j,1}, \dots, \hat{\boldsymbol{\alpha}}_{j,p_j}] = \\ &\underset{\mathbf{D}_j, \hat{\boldsymbol{\alpha}}_{j,1}, \dots, \hat{\boldsymbol{\alpha}}_{j,p_j}}{\operatorname{argmin}} \sum_{i=1}^{p_j} \mathbf{1}^T\exp(\mathbf{D}_j\boldsymbol{\alpha}_{j,i}) - \mathbf{y}^T\mathbf{D}_j\boldsymbol{\alpha}_{j,i}.\end{aligned} \tag{5}$$

This minimizing process gives an estimation $\hat{\mathbf{q}}_{j,i} = \exp(\mathbf{D}_j\boldsymbol{\alpha}_{j,i})$ for each patch. The recovered image is a result of a reprojection step that returns each reconstructed patch into its corresponding place in the image and averages. Since the clustering step is performed over the noisy image, the whole process is performed again with a clustering that relies on the reconstructed image.

The NLPCA for Poisson denoising is based on a variant that was developed for the Gaussian case [11]. This Gaussian technique can be used, of course, for the Poisson case by applying an Anscombe transform, but this approach was shown to be inferior for the case of low-photon count [7], [8].

Based on the above discussion we consider two techniques for Poisson denoising in the next section. Both rely on the OMP-denoising strategy proposed in [9] for the Gaussian case, and the NLPCA technique [7], [8].

## III. The Explored Algorithm

The first technique we consider uses the Anscombe transform and is presented in Fig 1. In the first step the Anscombe transform is performed on the noisy image $\mathbf{y}$. Then a set of overlapping patches $\mathbf{q}_i$[2] of size $\sqrt{d} \times \sqrt{d}$ is extracted from the transformed image. The set of patches is clustered into small groups of size $l$, using a Gaussian filter on $\mathbf{y}$ and grouping patches with small $\ell_2$ norm between them in the filtered image. A joint sparse coding step is performed for each group of patches with a variant of S-OMP presented in Algorithm 1. This step achieves a denoising effect for each patch. Having the group of denoised patches $\hat{\mathbf{q}}_i$, we return each of them to the place from which it was extracted in the image by averaging them. Having the reconstructed image we perform the refined inverse Anscombe transform proposed in [4].

In order to bypass the need for the Anscombe transform, the sparse coding step in the Anscombe based strategy should be replaced with a step that takes into account directly the Poisson statistics. In other words, we would like to have an approximation technique for solving (4). Such a method is proposed

**Algorithm 1** Gaussian Greedy Algorithm (an S-OMP variant)

**Require:** $k, \mathbf{D}, \{\mathbf{q}_1, \dots, \mathbf{q}_l\}$ where $\mathbf{q}_i \in \mathbb{R}^d$ is a Gaussian distributed vector with mean $\mathbf{p}_i = \mathbf{D}\boldsymbol{\alpha}_i$, and $k$ is the cardinality of $\boldsymbol{\alpha}_i$. All representation vectors $\boldsymbol{\alpha}_i$ are assumed to be jointly sparse, i.e., have the same support.

**Result:** $\hat{\mathbf{p}}_i = \mathbf{D}\hat{\boldsymbol{\alpha}}_i$ an estimate for $\mathbf{p}_i = \mathbf{D}\boldsymbol{\alpha}_i$.

Initialize the support $T^0 = \emptyset$ and set $t = 0$.
**while** $t < k$ **do**
  $t = t + 1$.
  Find new support element and representation estimate:
  $[\hat{\boldsymbol{\alpha}}_1^t, \dots, \hat{\boldsymbol{\alpha}}_l^t, j^t] = \operatorname{argmin}_{\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_l, j} \sum_{i=1}^l \left\|\mathbf{y} - \mathbf{D}_{T^{t-1}\cup\{j\}}\boldsymbol{\alpha}_i\right\|_2^2$.
  Update the support: $T^t = T^{t-1} \cup \{j^t\}$.
**end while**
Form the final estimate $\hat{\mathbf{p}}_i = \mathbf{D}\hat{\boldsymbol{\alpha}}_i^t, 1 \le i \le l$.

**Algorithm 2** Poisson Greedy Algorithm

**Require:** $k, \mathbf{D}, \{\mathbf{q}_1, \dots, \mathbf{q}_l\}$ where $\mathbf{q}_i \in \mathbb{R}^d$ is a Poisson distributed vector with mean and variance $\mathbf{p}_i = \exp(\mathbf{D}\boldsymbol{\alpha}_i)$, and $k$ is the cardinality of $\boldsymbol{\alpha}_i$. All representation vectors $\boldsymbol{\alpha}_i$ are assumed to be jointly sparse, i.e., have the same support.

**Result:** $\hat{\mathbf{p}}_i = \exp(\mathbf{D}\hat{\boldsymbol{\alpha}}_i)$ an estimate for $\mathbf{p}_i = \exp(\mathbf{D}\boldsymbol{\alpha}_i)$.

Initialize the support $T^0 = \emptyset$ and set $t = 0$.
**while** $t < k$ **do**
  $t = t + 1$.
  Find new support element and representation estimate:
  $[\hat{\boldsymbol{\alpha}}_1^t, \dots, \hat{\boldsymbol{\alpha}}_l^t, j^t] = \operatorname{argmin}_{\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_l, j} \sum_{i=1}^l \mathbf{1}^T\exp(\mathbf{D}_{T^{t-1}\cup\{j\}}\boldsymbol{\alpha}_i) - \mathbf{y}^T\mathbf{D}_{T^{t-1}\cup\{j\}}\boldsymbol{\alpha}_i$.
  Update the support: $T^t = T^{t-1} \cup \{j^t\}$.
**end while**
Form the final estimate $\hat{\mathbf{p}}_i = \exp(\mathbf{D}\hat{\boldsymbol{\alpha}}_i^t), 1 \le i \le l$.

[1] The non-local sparse PCA (NLSPCA) [8], a variation of NLPCA, adds an $\ell_1$ regularization term on $\alpha_{j,i}$.

[2] $1 \le i \le (m_1 - \sqrt{d} + 1)(m_2 - \sqrt{d} + 1)$ where $m_1$ and $m_2$ are the vertical and horizontal dimensions of the noisy image $\mathbf{y}$ respectively.

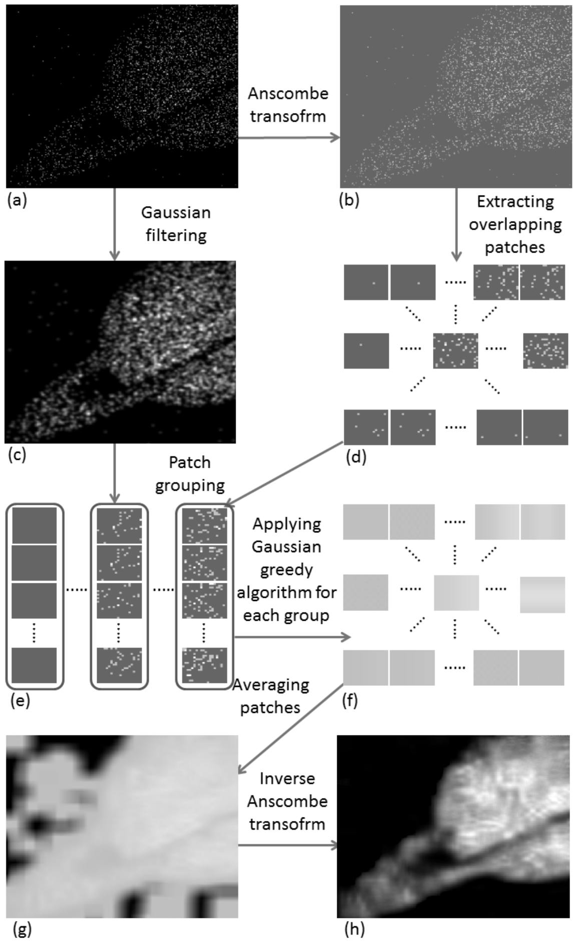



Fig. 1. The proposed Anscombe based Poisson denoising algorithm.

in Algorithm 2. The algorithm adds in a greedy way elements to the representation of the approximated patch in a similar way to what is done in the Gaussian algorithm. However, the minimization step for finding the new representation does not have a closed form solution like in the Gaussian case. We use the Matlab optimization toolbox for this task. As a dictionary, the two dimensional orthogonal DCT is used, with no training process involved. This leaves room for future improvement.

Having the Poisson greedy algorithm, we can now propose a new technique for Poisson image denoising. The proposed method is very similar to the Anscombe based one and is presented in Fig 2. It includes almost the same steps but without the Anscombe transformation and with the Poisson greedy algorithm instead of the Gaussian greedy algorithm.

## IV. Experiments

We test these algorithms in the case of low photon count. In all experiments we set $l = 50$. Figure 3 presents reconstruction results of a noisy text image (peak = 0.2). It can be observed that the Anscombe technique gives a sharper text while the model based technique recovers better the background.

Figures 4 and 5 present the reconstruction result for the noisy *Saturn* image with peak=0.2 and peak=0.1 respectively.

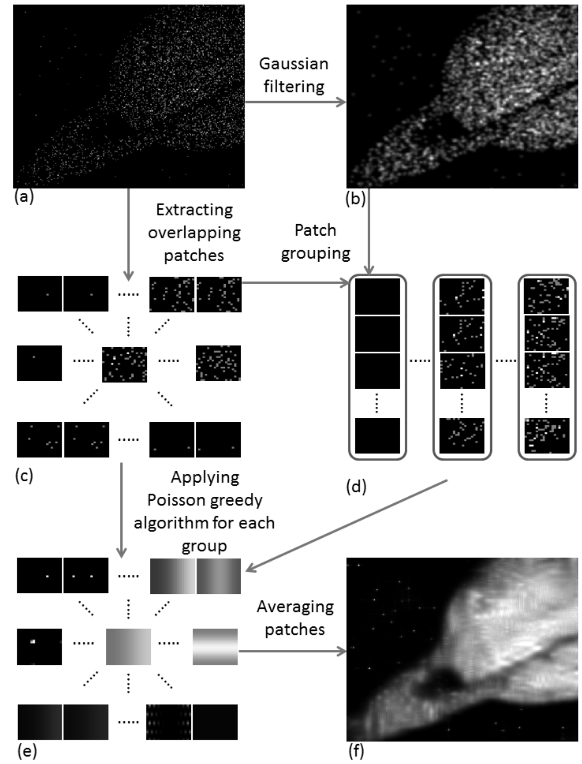



Fig. 2. The proposed model based Poisson denoising algorithm.

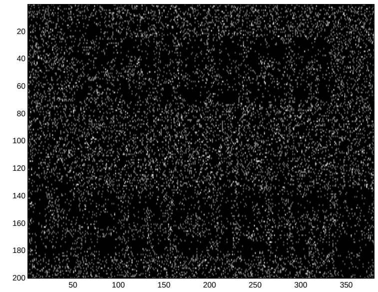
We compare in Fig. 4 our performance with the ones of NLPCA and NLSPCA [8] that achieve state-of-the-art results. Indeed, these methods achieve better reconstruction results than our techniques in terms of PSNR, but overall, the outcome looks very similar visually. We should add that our methods do not employ learned dictionary, which has the potential to boost further our outcome.

## V. Discussion and Conclusion

In this work we have presented two techniques for Poisson denoising. Both are patch based and rely on a sparse model for the patches of the original image. The first uses the Anscombe transform and then performs a variant of OMP-denoising using joint sparsity for the patches. The second works directly with the Poisson statistics using a new sparse model for the patches in the image. Both methods present comparable performance but do not achieve yet state-of-the-art results. Both presented methods suffer from two drawbacks.

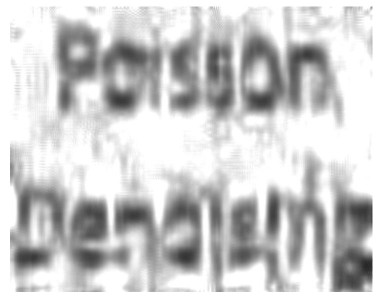
- In OMP-denoising [9] the stopping criterion for OMP is error based. However, in this work we used a cardinality based stopping criterion that forces a constant sparsity for all patches. Thus, an error based stopping criterion is needed. However, it is not clear what this criterion should be in our case. One approach that may be considered is using a generalization of the Stein unbiased risk estimator (SURE) for the Poisson noise case [12], [13].
- In [9] it was demonstrated that learning the dictionary $\mathbf{D}$ enhances the performance. Similarly, in [8] the small

(a) *Text* image.




(b) Poisson noisy image. Peak = 0.2.

(c) Anscombe based proposed method with $k = 3$. PSNR = 14.71dB.



(d) Model based proposed method with $k = 2$. PSNR = 14.79dB.

Fig. 3. Denoising of a text image with peak = 0.2.

dictionaries were learned. Thus, it should be natural to add a dictionary learning phase into the algorithms. As mentioned before, this change is more critical to the Poisson model based algorithm since less is known about the dictionaries that suites this model and the two dimensional DCT used in our work suites more the standard sparsity model than the proposed one.

In a future work we will explore both issues aiming at improving the performance of both methods. However, we believe that the second method has more potential and will profit more from the suggested improvements.

## Acknowledgment

R. Giryes is grateful to the Azrieli Foundation for the award of an Azrieli Fellowship.

## References

[1] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Review*, vol. 51, no. 1, pp. 34–81, 2009.

[2] B. Zhang, J. Fadili, and J. Starck, "Wavelets, ridgelets, and curvelets for poisson noise removal," *IEEE Transactions on Image Processing*, vol. 17, no. 7, pp. 1093–1108, July 2008.

[3] J. Boulanger, C. Kervrann, P. Bouthemy, P. Elbau, J.-B. Sibarita, and J. Salamero, "Patch-based nonlocal functional for denoising fluorescence microscopy image sequences," *IEEE Transactions on Medical Imaging*, vol. 29, no. 2, pp. 442–454, Feb. 2010.

[4] M. Makitalo and A. Foi, "Optimal inversion of the Anscombe transformation in low-count poisson image denoising," *IEEE Transactions on Image Processing*, vol. 20, no. 1, pp. 99–109, Jan. 2011.

[5] F. J. Anscombe, "The transformation of Poisson, Binomial and negative-Binomial data," *Biometrika*, vol. 35, no. 3-4, pp. 246–254, 1948.

[6] M. Fisz, "The limiting distribution of a function of two independent random variables and its statistical application," *Colloquium Mathematicum*, vol. 3, pp. 138–146, 1955.

[7] J. Salmon, C.-A. Deledalle, R. Willett, and Z. T. Harmany, "Poisson noise reduction with non-local PCA," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2012.*, March 2012, pp. 1109–1112.

[8] J. Salmon, Z. T. Harmany, C.-A. Deledalle, and R. Willett, "Poisson noise reduction with non-local PCA," *CoRR*, vol. abs/1206.0338, 2012.

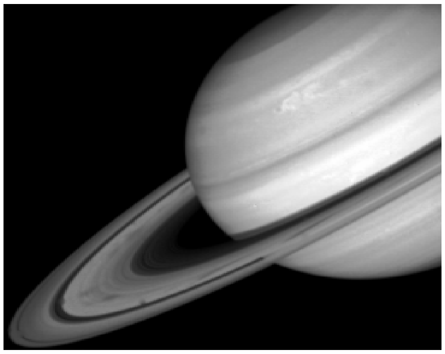
[9] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *Signal Processing, IEEE Transactions on*, vol. 54, no. 11, pp. 4311 –4322, nov. 2006.

[10] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. part i: Greedy pursuit," *Signal Processing*, vol. 86, no. 3, pp. 572 – 588, 2006, sparse Approximations in Signal and Image Processing.

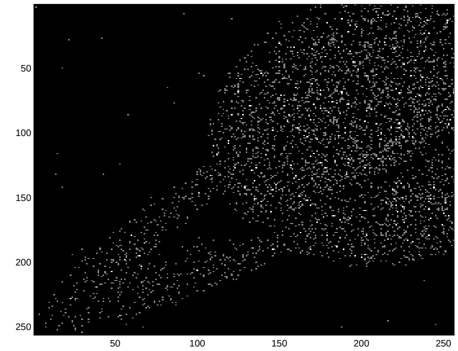
[11] G. Yu, G. Sapiro, and S. Mallat, "Solving inverse problems with piecewise linear estimators: From gaussian mixture models to structured sparsity," *IEEE Transactions on Image Processing*, vol. 21, no. 5, pp. 2481 –2499, may 2012.

[12] C.-A. Deledalle, F. Tupin, and L. Denis, "Poisson NL means: Unsupervised non local means for poisson noise," in *17th IEEE International Conference on Image Processing (ICIP), 2010*, sept. 2010, pp. 801 –804.

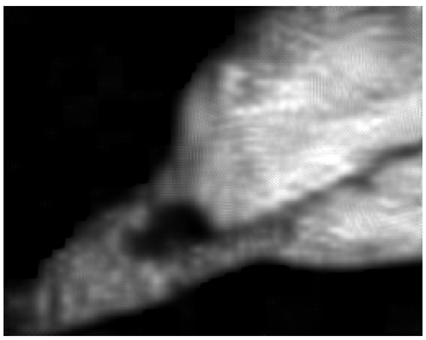
[13] R. Giryes, M. Elad, and Y. C. Eldar, "The projected GSURE for automatic parameter tuning in iterative shrinkage methods," *Applied and Computational Harmonic Analysis*, vol. 30, no. 3, pp. 407 – 422, 2011.



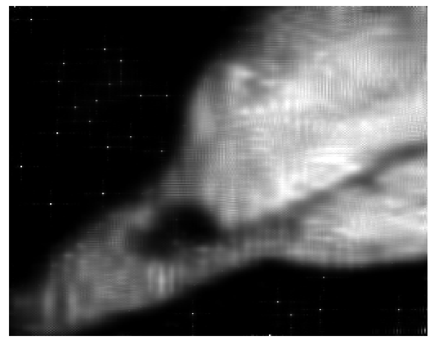
(a) *Saturn* image



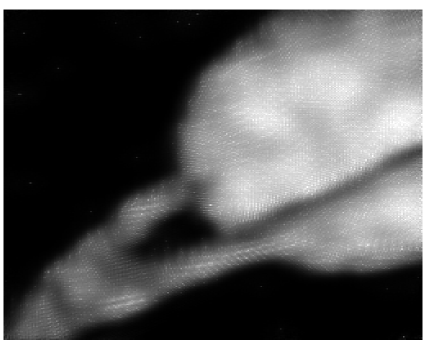

(b) Poisson noisy image. Peak =0.2.



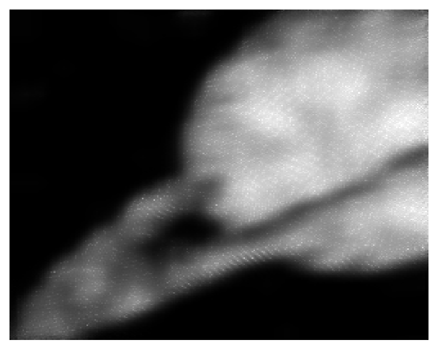
(c) Anscombe based proposed method with $k = 2$. PSNR = 22.44dB.



(d) Model based proposed method with $k = 2$. PSNR = 22.27dB.

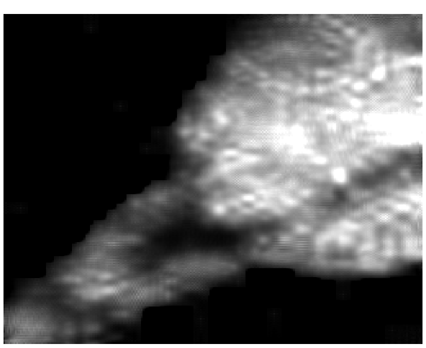


(e) NLPCA. PSNR = 22.97dB.

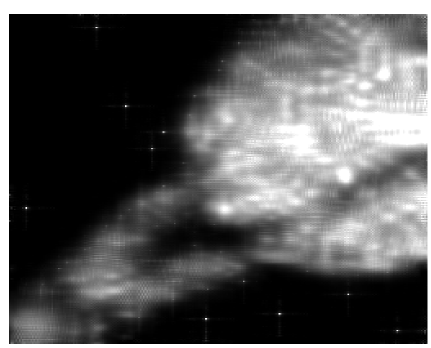


(f) NLSPCA. PSNR = 22.74dB.

Fig. 4. Denoising of *Saturn* image with peak = 0.2



(a) Anscombe based proposed method with $k = 2$. PSNR = 19.58dB.



(b) Model based proposed method with $k = 2$. PSNR = 19.51dB.

Fig. 5. Denoising of *Saturn* image with peak = 0.1