

# GAUSSIAN-CAUCHY MIXTURE MODELING FOR ROBUST SIGNAL-DEPENDENT NOISE ESTIMATION

*Lucio Azzari and Alessandro Foi*

Department of Signal Processing, Tampere University of Technology  
P.O. Box 553, FIN-33101 Tampere, Finland

## ABSTRACT

We introduce an adaptive Gaussian-Cauchy mixture modeling for the likelihood of pairwise mean/standard-deviation scatter points found when estimating signal-dependent noise. The maximization of the likelihood is used to identify the noise-model parameters, following an adaptive mixture parameter that controls the balance between the Gaussian and the heavy-tailed Cauchy. This renders the estimation robust with respect to outliers, typically present in large quantities among the scatter points from images dominated by texture. The modeling is directly suited to describing also observations subject to clipping, i.e. under- or over-exposure. Experiments on a dataset of badly exposed and highly textured images demonstrate the effectiveness of the adaptive Gaussian-Cauchy mixture likelihood for the accurate estimation of the noise standard-deviation curve.

**Index Terms**— Signal-dependent noise, robust estimation, mixture modeling, clipping

## 1. INTRODUCTION

With signal-dependent noise we refer to a model of noise with variable standard deviation that depends on the particular intensity value of the affected signal. This model has been widely adopted by the scientific and engineering community, because it well approximates the noise affecting data acquired by many sensing devices, including imaging sensors such as CCD and CMOS cameras.

The common procedure [1–8] for signal-dependent noise estimation consists in dividing the data or image of interest into uniform or homogeneous regions. Each group of samples is then used for the estimation of a mean-standard deviation pairs. The collection of such pairs yields a cloud of points scattered around the curve that describes the dependency of the noise standard deviation from the mean of the signal. Finally, in order to estimate the noise standard-deviation curve, a global parametric model is fitted to the points. The global fitting can be hampered by outliers among the mean and standard deviation pairs. This problem becomes

particularly relevant when dealing with images with numerous highly textured regions.

The estimation of the standard-deviation curve is further complicated by the inevitably limited range of the sensing device, which causes clipping of data whenever it falls outside an admissible interval. In imaging, this phenomenon may occur when the scene is under- or over-exposed.

To provide a unique and reliable solution to these two distinct problems, we introduce a novel likelihood function based on an adaptive Gaussian-Cauchy mixture modeling of the mean and standard-deviation pairs. The presented work generalizes the method [1], significantly improving its robustness against outliers. The estimation algorithm based on the introduced Gaussian-Cauchy mixture likelihood provides accurate estimates of the noise standard-deviation curve from images that can be badly exposed as well as dominated by texture.

The remainder of the paper is organized as follows. In Section 2, we introduce the observation model for signal-dependent noise, also considering the case of clipped data. The important affine variance model is presented too. Section 3 briefly describes the first part of the algorithm, which deals with the estimation of the mean-standard deviation pairs that constitute the scatterplot. The core of our contribution is given in Section 4, with a complete description of the proposed robust Gaussian-Cauchy mixture likelihood. The maximization of the likelihood with an adaptive mixture parameter is also explained. Section 5 provides experimental validation of our approach over a dataset of real raw images affected by clipping and high-frequency texture, confirming the effectiveness of the proposed adaptive mixture modeling.

## 2. PROBLEM STATEMENT

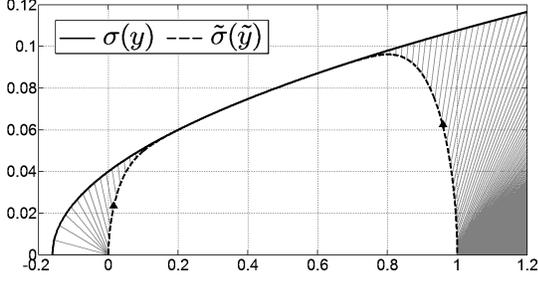
Let us consider a noisy image  $z$  as the sum of a noise-free image  $y$  and noise with signal-dependent standard deviation  $\sigma(y)$ ,

$$z(\mathbf{x}) = y(\mathbf{x}) + \sigma(y(\mathbf{x}))\xi(\mathbf{x}), \quad (1)$$

where  $\mathbf{x} \in \mathbf{X} \subset \mathbb{N}^2$  denotes the spatial coordinate of a pixel, and  $\xi(\mathbf{x})$  is a standardized random variable with zero-mean and unitary variance. We model the noise as spatially independent (i.e. diagonal covariance matrix). For the sake of simplicity and in agreement with the analysis in [1, 9], we

Contact info: firstname.lastname@tut.fi

♥ This work was supported by the Academy of Finland (project no. 252547).



**Fig. 1:** Effect of clipping on the standard deviation function in case of affine variance (2) [9]. Due to clipping, the standard deviation  $\tilde{\sigma}(\tilde{y})$  goes to zero as it approaches the bounds (0, 1).

treat  $\xi(\mathbf{x})$  as a standard normal, i.e.  $\xi(\mathbf{x}) \sim \mathcal{N}(0, 1)$ , so that only the standard deviation of  $\sigma(y(\mathbf{x}))\xi(\mathbf{x})$  depends on the particular value of the noise-free signal  $y(\mathbf{x})$ .

Our goal is to robustly estimate the function  $\sigma: \mathbb{R} \rightarrow \mathbb{R}^+$  that links the noise standard deviation  $\sigma(y)$  to the signal  $y$ .

A peculiar and recurrent signal-dependent noise model is the affine variance model that well approximates the noise affecting digital image sensors:

$$\sigma^2(y(\mathbf{x})) = ay(\mathbf{x}) + b. \quad (2)$$

According to this model, the noise  $\eta(y(\mathbf{x}))$  has variance that affinely depends on the signal intensity. In this case, the estimation of the curve  $\sigma(y)$  can be reduced to the estimation of  $a$  and  $b$ .

The output of any imaging device has a limited range, which, without loss of generality, we assume rescaled to the  $[0, 1]$  interval. Here, 0 and 1 are the lowest (i.e. darkest) and highest (i.e. brightest) values, respectively. Values of  $z$  that exceed these bounds are inevitably *clipped*, i.e. they are replaced by the bounds themselves:

$$\tilde{z}(\mathbf{x}) = \max(0, \min(z(\mathbf{x}), 1)). \quad (3)$$

Throughout the text, the *tilde* decoration  $\tilde{\cdot}$  is used to indicate variables and quantities subject to clipping.

The noise model for the observed clipped image  $\tilde{z}$  can be expressed as

$$\tilde{z}(\mathbf{x}) = \tilde{y}(\mathbf{x}) + \tilde{\sigma}(\tilde{y}(\mathbf{x}))\tilde{\xi}(\mathbf{x}), \quad (4)$$

where  $\tilde{y}$  is the expectation of the clipped data  $\tilde{z}$ , and  $\tilde{\sigma}(\tilde{y}(\mathbf{x}))$  is the standard deviation of the clipped noisy data. Clipping causes an apparent distortion of the standard-deviation curve that describes the signal-dependent noise model, as illustrated in Fig. 1. In particular, the noise standard deviation approaches zero whenever the intensity approaches the range bounds 0 and 1. A comprehensive analysis of the interplay between clipping and signal-dependent noise models can be found in [9].

The difference between  $\tilde{\sigma}(\tilde{y}(\mathbf{x}))$  and  $\sigma(y(\mathbf{x}))$  constitutes a dramatic deviation from the traditional affine-variance model (2). This makes the estimation of the noise parameters particularly difficult when portions of the image are not correctly exposed.

### 3. LOCAL ESTIMATION OF MEAN-STANDARD DEVIATION PAIRS

Overall, the proposed approach is characterized by two main stages: first, an estimation of a collection of expectation/standard deviation pairs (i.e. the construction of a scatterplot); and, second, the maximization of a likelihood model that explains these estimates (i.e. the fitting of a global parametric curve to the scatterplot). Our present contribution concerns the latter stage only, as described in detail in the next section. For the former stage, we leverage, without modification, the first stage of the original algorithm [1], which we briefly summarize here below.

The pairwise estimates of standard deviation and mean are computed from a high-pass and from low-pass version of the image, respectively denoted as  $\tilde{z}^{\text{wdet}}$  and  $\tilde{z}^{\text{wapp}}$ . These are obtained by convolving  $\tilde{z}$  with a 2-D wavelet function  $\psi$  and its scaling function  $\varphi$ , respectively:

$$\tilde{z}^{\text{wdet}} = \downarrow_2(\tilde{z} \otimes \psi) \quad \text{and} \quad \tilde{z}^{\text{wapp}} = \downarrow_2(\tilde{z} \otimes \varphi), \\ \sum \psi = 0, \quad \|\psi\|_2 = 1, \quad \sum \varphi = 1,$$

where  $\downarrow_2$  is the decimation operator, here used in order to discard every second row and second column of an image.

It can be shown [1] that the standard deviation (std) of the detail coefficients is comparable to plugging the expectation of the approximation coefficients into the standard deviation function  $\tilde{\sigma}$ :

$$\text{std}\{\tilde{z}^{\text{wdet}}\} \simeq \tilde{\sigma}(\mathbb{E}\{\tilde{z}^{\text{wapp}}\}). \quad (5)$$

The approximation (5) becomes locally accurate in uniform regions of  $\tilde{z}$ . Therefore, in order to reliably compute the local mean-variance pairs excluding texture areas (non-uniform regions), it is necessary to segment  $\tilde{z}$ .

For the sake of brevity, we refer the reader to [1] for the technical details about the segmentation. The obtained segments  $S_i$ ,  $i = 1, \dots, N$ , include pixels whose expected intensity values belong to a narrow intensity interval, and can thus be treated as level sets.

Finally, for each level set  $S_i$ , we estimate its mean  $\hat{y}_i$  and standard deviation  $\hat{\sigma}_i$  as the sample mean and sample median of  $\tilde{z}^{\text{wapp}}$  and  $|\tilde{z}^{\text{wdet}}|/0.6745$  restricted on  $S_i$ , respectively.

### 4. ROBUST ML FITTING OF A GLOBAL PARAMETRIC MODEL

The global optimization step requires a model of the distributions of  $\hat{y}_i$  and  $\hat{\sigma}_i$ . In the ideal case without outliers, the estimates follow normal distributions of the forms, respectively,

$$\hat{y}_i \sim \mathcal{N}(\tilde{y}_i, \tilde{\sigma}^2(\tilde{y}_i)c_i), \quad \hat{\sigma}_i \sim \mathcal{N}(\tilde{\sigma}(\tilde{y}_i), \tilde{\sigma}^2(\tilde{y}_i)d_i) \quad (6)$$

where the coefficients  $c_i$  and  $d_i$  are scaling parameters that depend on the number of samples  $n_i$  in the level set  $S_i$  [1]. However, in case of images dominated by texture, the high-frequency components of  $y$  can introduce severe outliers in

the local estimates  $\hat{\sigma}_i$ , and consequently mislead the estimation of the noise model parameters, as illustrated in Fig. 3.

The Cauchy PDF is characterized by heavy tails, which makes it suitable for estimation in presence of outliers.

The PDFs of the estimates are therefore expressed as

$$\begin{aligned} \wp(\hat{y}_i | \tilde{y}_i = \tilde{y}) &= g_{\mathcal{N}}(\hat{y}_i; \tilde{y}, \tilde{\sigma}^2(\tilde{y}) c_i), \\ \wp(\hat{\sigma}_i | \tilde{y}_i = \tilde{y}) &= (1 - \lambda) g_{\mathcal{N}}(\hat{\sigma}_i; \tilde{\sigma}(\tilde{y}), \tilde{\sigma}^2(\tilde{y}) d_i) \\ &\quad + \lambda g_{\mathcal{C}}(\hat{\sigma}_i; \tilde{\sigma}(\tilde{y}), \tilde{\sigma}^2(\tilde{y}) d_i), \end{aligned} \quad (7)$$

where  $g_{\mathcal{N}}$  and  $g_{\mathcal{C}}$  are the normal and Cauchy PDFs with location parameter  $\mu$  and scaling parameter  $\nu$ :

$$\begin{aligned} g_{\mathcal{N}}(x; \mu, \nu^2) &= \frac{1}{\sqrt{2\pi\nu^2}} e^{-\frac{(x-\mu)^2}{2\nu^2}}, \\ g_{\mathcal{C}}(x; \mu, \nu^2) &= \frac{\nu}{\pi(\nu^2 + (x-\mu)^2)}, \end{aligned}$$

and  $\lambda \in [0, 1]$  is the mixture coefficient between these distributions.

The joint probability density can be expressed as the product

$$\wp((\hat{y}_i, \hat{\sigma}_i) | \tilde{y}_i = \tilde{y}) = \wp(\hat{y}_i | \tilde{y}_i = \tilde{y}) \wp(\hat{\sigma}_i | \tilde{y}_i = \tilde{y}).$$

Given the distributions of all the pairs  $\{\hat{y}_i, \hat{\sigma}_i\}_{i=1}^N$ , we finally obtain the posterior likelihood function  $\tilde{L}$  integrating all the densities  $\wp((\hat{y}_i, \hat{\sigma}_i) | \tilde{y}_i = \tilde{y})$  with respect to the prior density  $\wp_0(y)$  of  $y$  as

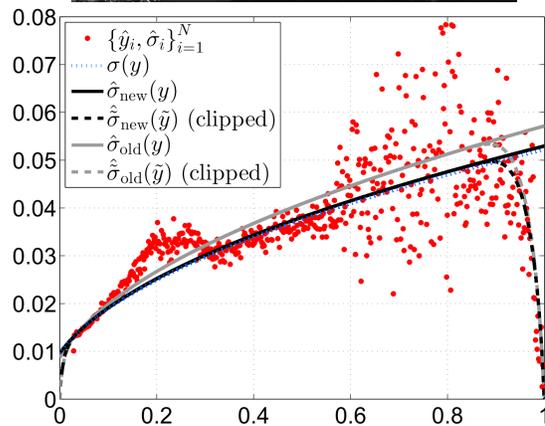
$$\tilde{L}(\theta, \lambda) = \prod_{i=1}^N \int_{-\infty}^{\infty} \wp((\hat{y}_i, \hat{\sigma}_i) | \tilde{y}_i = \tilde{y}) \wp_0(y) dy, \quad (8)$$

where  $\theta$  is an  $m$ -dimensional vector composed by the model parameters to be estimated. The vector  $\theta$  determines univocally both the clipped standard-deviation curve  $\tilde{\sigma}(\tilde{y})$  and the (unclipped) standard-deviation curve  $\sigma(y)$ . E.g.,  $\theta = [a, b]$ , i.e.  $m = 2$ , in case of the affine mean-variance relation (2). The integration in (8) makes the distributions independent from the unknown value of  $y$ . Similar to [1], in our experiments we assume a uniform prior density  $\wp_0$ .

A principal feature of the proposed approach consists in treating the mixture coefficient  $\lambda$  as an unknown variable. Therefore, our optimization is performed with respect to  $m + 1$  parameters, i.e. the vector  $\theta$  and the mixture coefficient  $\lambda$ :

$$\hat{\theta} = \arg \max_{\theta \in \mathbb{R}^m, \lambda \in [0, 1]} \tilde{L}(\theta, \lambda). \quad (9)$$

Note that the maximization of  $\tilde{L}$  corresponds to optimizing the fit of the clipped standard-deviation curve  $\tilde{\sigma}(\tilde{y})$  to the scatterplot, because the probabilities of the points  $(\hat{y}_i, \hat{\sigma}_i)$  are determined by this curve through the PDFs (7). The value of  $\lambda$  upon maximization of (9), denoted as  $\hat{\lambda}$ , provides an indication of the amount of outliers in the scatterplot. When estimating noise in an image dominated by texture, one can expect to obtain  $\hat{\lambda} \approx 1$ , i.e. the mixture reduces to a Cauchy distribution; conversely, if the image is mostly smooth or piecewise smooth,  $\hat{\lambda} \approx 0$ , i.e. the mixture reduces to a Gaussian.



**Fig. 2:** Top: Raw image (blue channel) with extensive portions of high-frequency texture and overexposed areas. The image was captured with a Fujifilm FinePix S9600 camera at ISO 800. Bottom: Mean vs. standard deviation diagram. The red scatterplot points correspond to individual mean and standard-deviation estimates  $(\hat{y}_i, \hat{\sigma}_i)$ . Note the presence of outliers in the scatterplot. The standard-deviation curves  $\hat{\sigma}_{\text{old}}(\tilde{y})$  and  $\hat{\sigma}_{\text{new}}(\tilde{y})$  fitted through maximization of the original and of the proposed robust likelihood functions are shown as dashed lines. The respective *unclipped* curves  $\hat{\sigma}_{\text{old}}(y)$  and  $\hat{\sigma}_{\text{new}}(y)$  are drawn as solid lines and are compared against the ground truth  $\sigma(y)$  (dotted line).

#### 4.1. Numerical Solution

We solve (9) numerically, using the Nelder-Mead iterative downhill simplex method [10], under the constraint  $\lambda \in [0, 1]$ . Similar to [1], we initialize the optimization from a very rough least-square (LS) solution. For what concerns  $\lambda$ , in our experiments we use 1 as initial value, in order to account for the worst-case scenario of a scatterplot corrupted by several outliers.

We remark that the need for solving (9) numerically does not follow from introducing the Cauchy term in the mixture, but was already necessary with the Gaussian-only model [1] due to the heteroskedasticity.

## 5. EXPERIMENTS ON HIGHLY TEXTURED IMAGES

To validate the effectiveness of the robust likelihood with adaptive mixture of Gaussian-Cauchy distributions, we com-

pare the proposed approach against the original algorithm [1], which is based on a Gaussian likelihood (i.e. fixed  $\lambda = 0$ ). The corresponding estimated standard-deviation curves are denoted as  $\hat{\sigma}_{\text{new}}(\tilde{y})$  and  $\hat{\sigma}_{\text{old}}(\tilde{y})$  for the clipped model (4), and as  $\hat{\sigma}_{\text{new}}(y)$  and  $\hat{\sigma}_{\text{old}}(y)$  for the unclipped model (1), respectively. A typical example of a clipped highly textured image is shown in Fig. 2. The image has been taken with a Fujifilm S9600 camera at ISO 800. The effect of clipping can be easily seen at the right in the scatterplot, where the red dots drop towards the limiting coordinate  $(0, 1)$ .

The textures in the image cause overestimation of the standard deviations, producing outliers, which can be seen in the scatterplot particularly for  $\tilde{y} \in [0.1, 0.3] \cup [0.6, 0.9]$ . The original algorithm [1] results in the *unclipped* standard-deviation curve  $\hat{\sigma}_{\text{old}}(y)$  (solid gray line), which misestimates the ground-truth curve  $\sigma(y)$ <sup>1</sup> (dotted line). On the contrary, the proposed mixture model is by design robust to outliers in the scatterplot, and  $\hat{\sigma}_{\text{new}}(y)$  is in good agreement with the ground-truth  $\sigma(y)$ . The estimated  $\hat{\lambda}$  is 1; which indicates that the amount of outliers is significant and justifies the use of a heavy-tailed distribution for modeling the likelihood.

We have also verified the advantage of the proposed robust algorithm on the NED dataset [11]. This challenging dataset consists of 25 heavily textured raw images captured with a Nikon D80 camera, at ISO values from 100 to 320. In Table 1 we report the average normalized integral error (ANIE) over all images in the dataset, each channel separately. The normalized integral errors are computed as

$$\text{NIE} = \int_{\Omega} \frac{|\hat{\sigma}(\tilde{y}) - \tilde{\sigma}(\tilde{y})|}{\tilde{\sigma}(\tilde{y})} d\tilde{y} \Big/ \int_{\Omega} 1 d\tilde{y},$$

where  $\hat{\sigma}(\tilde{y})$  is the estimated clipped standard deviation curve,  $\tilde{\sigma}(\tilde{y})$  is the ground-truth curve, and  $\Omega = \{\tilde{y} | \tilde{\sigma}(\tilde{y}) > 0\}$ .

The proposed robust algorithm systematically improves over the previous one. While the numerical values in the table may be difficult to appreciate, Fig. 3 gives a clear illustration of the kind of misestimation problems resolved through the proposed adaptive Gaussian-Cauchy mixture modeling. As can be seen in the figure, large regions of texture yield severe outliers in the estimation of mean-standard deviation pairs and a considerable amount of points in the scatterplot are far from the correct parametric trend.

Matlab codes implementing the proposed algorithm are available [12].

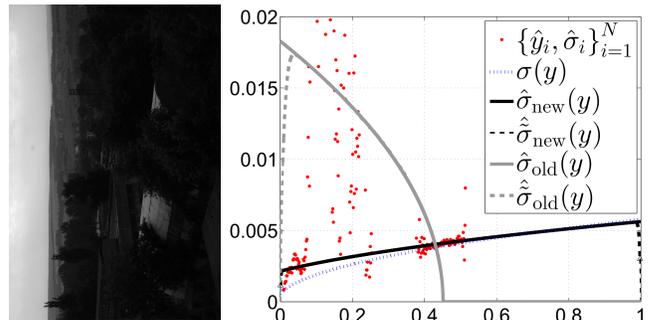
## 6. DISCUSSION AND CONCLUSIONS

The experiments and results described in Section 5 confirm the advantage of adopting a robust global fitting model when dealing with highly textured images. The presented algorithm, with its adaptive Gaussian-Cauchy mixture, is robust to outliers and ensures more reliable results than the previous

<sup>1</sup>The ground-truth curve has been estimated from an ideal texture-free image obtained with out-of-focus optics, as described in [1].

ANIE	Red	Green	Blue
Proposed	0.3088	0.6036	0.3157
Old	0.4399	0.8387	0.3708

**Table 1:** Average normalized integral errors (ANIE) from the estimates on the NED dataset. The results for each channel is shown separately.



**Fig. 3:** Underexposed raw image (red channel) with intensity range normalized to  $[0, 1]$  from the NED dataset [11], and the relative scatterplot of mean-standard deviation pairs, severely corrupted by outliers. The noise curve  $\hat{\sigma}_{\text{old}}$  estimated by the method [1] is directly impacted by the outliers and results in a failure. The proposed method provides instead a robust fit to the scatterplot and the estimated curve  $\hat{\sigma}_{\text{new}}$ , although not perfect, achieves a reasonable approximation of the underlying noise model.

algorithm based on a Gaussian-only modeling. The benefit of the presented approach is particularly significant in challenging cases with severe outliers. Through the automatic optimization of the mixture parameter, we conveniently encompass also less problematic images attaining high precision.

As can be seen in the figures, the outliers  $\hat{\sigma}_i$  do typically lie in the right heavy tail (i.e. overestimation) of the conditional PDF (7) given the true  $\tilde{\sigma}$ . Therefore, one may question the role of the left heavy tail. However, when maximizing the likelihood (8) starting from an inaccurate guess  $\hat{\theta}$ , it can happen that the candidate function  $\hat{\sigma}$  fits the overestimated  $\hat{\sigma}_i$ , while the non-overestimated values appear instead as outliers in the left tail of the PDF (7). Thus, the left heavy tail effectively prevents (9) from getting stuck at such local maxima.

The proposed likelihood model is relatively independent of the procedure utilized to produce the scatterplot pairs; thus, the proposed optimization is relevant to a wide class of algorithms based on similar fitting of a standard-deviation curve. A special feature of our model consists in the explicit treatment of the clipping due to under- or over-exposure (3)(4), commonly encountered with imagers having limited dynamic range. Moreover, while in this paper we have considered only the affine variance model (2), we note that the procedure is independent of the specific parametric model between  $\theta$  and  $\sigma^2$  (1). In particular, the algorithm [12] can handle any polynomial order.

## 7. 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] P. Gravel, G. Beaudoin, and J.A. De Guise, "A method for modeling noise in medical images," *Medical Imaging, IEEE Transactions on*, vol. 23, no. 10, pp. 1221–1232, 2004.
- [3] S.I. Olsen, "Estimation of Noise in Images: An Evaluation," *CVGIP: Graphical Models and Image Processing*, vol. 55, no. 4, pp. 319 – 323, 1993.
- [4] S.K. Abramov, B. Vozel, J.T. Astola, K. Chehdi, and V.V. Lukin, "Segmentation-based method for blind evaluation of noise variance in images," *Journal of Applied Remote Sensing*, vol. 2, no. 1, pp. 023533–023533–16, 2008.
- [5] M. Lebrun, M. Colom, A. Buades, and J. M. Morel, "Secrets of image denoising cuisine," *Acta Numerica*, vol. 21, pp. 475–576, 5 2012.
- [6] J. Immerkaer, "Fast Noise Variance Estimation," *Computer Vision and Image Understanding*, vol. 64, no. 2, pp. 300–302, 1996.
- [7] B. Aiazzi, L. Alparone, S. Baronti, M. Selva, and L. Stefani, "Unsupervised estimation of signal-dependent CCD camera noise," *EURASIP Journal on Advances in Signal Processing*, vol. 2012, no. 1, pp. 231, 2012.
- [8] N. Acito, M. Diani, and G. Corsini, "Signal-Dependent Noise Modeling and Model Parameter Estimation in Hyperspectral Images," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, no. 8, pp. 2957–2971, 2011.
- [9] A. Foi, "Clipped noisy images: Heteroskedastic modeling and practical denoising," *Signal Processing*, vol. 89, no. 12, pp. 2609 – 2629, 2009.
- [10] J. A. Nelder and R. Mead, "A Simplex Method for Function Minimization," *The Computer Journal*, vol. 7, no. 4, pp. 308–313, 1965.
- [11] M.L. Uss, B. Vozel, V.V. Lukin, and K. Chehdi, "NED2012 Database," <http://rsd.khai.edu/ned2012/ned2012.php>.
- [12] A. Foi and L. Azzari, "ClipPoisGaus\_stdEst2D," <http://www.cs.tut.fi/~foi/sensornoise.html/>, Matlab software.