

Automated Filter Parameter Selection using Measures of Noisiness

Ajit Rajwade, Anand Rangarajan and Arunava Banerjee,
Department of CISE, University of Florida, Gainesville, USA
avr, anand, arunava@cise.ufl.edu

Abstract

Despite the vast body of literature on image denoising, relatively little work has been done in the area of automatically choosing the filter parameters that yield optimal filter performance. The choice of these parameters is crucial for the performance of any filter. In the literature, some independence-based criteria have been proposed, which measure the degree of independence between the denoised image and the residual image (defined as the difference between the noisy image and the denoised one). We contribute to these criteria and point out an important deficiency inherent in all of them. We also propose a new criterion which quantifies the inherent ‘noisiness’ of the residual image without referring to the denoised image, starting with the assumption of an additive and i.i.d. noise model, with a loose lower bound on the noise variance. Several empirical results are demonstrated on two well-known algorithms: NL-means and total variation, on a database of 13 images at six different noise levels, and for three types of noise distributions.

1. Introduction

Contemporary literature on image denoising is certainly very vast. Some of the popular approaches include partial differential equations (PDEs) [14], wavelet thresholding [8], learning-based approaches like KSVD [10] and non-local approaches [4]. The PDE-based methods have undergone advancements such as the use of complex diffusion [18]. More recently, new approaches based on spatially-varying regressions [20] and spatially varying convolutions [19] have also been developed.

The typical denoising technique requires tuning parameters that are critical for its optimal performance. In denoising experiments reported in contemporary literature, the filter performance is usually measured by the mean-squared error (MSE) between the denoised and the original image. The parameters are picked so as to yield the least MSE value for a particular filter, but this requires knowledge of the

original image and is not extensible to real-world denoising situations. Hence we need criteria for parameter selection that do not refer to the original image. In this paper, we classify these criteria into two types: (1) independence-based criteria that measure the degree of independence between the denoised image and the residual, and (2) criteria that measure how noisy the residual image is, without referring to the denoised image. We contribute to and critique criteria of type (1), and proposed those of type (2). Our criteria make the assumption that the noise is i.i.d. and additive, and that a loose lower-bound on the noise variance is known.

This paper is organized as follows: Section 2 reviews existing literature for filter parameter selection, followed by a description of our criteria in Section 3, experimental results in Section 4 and discussion in Section 6

2 Literature Review

In PDE-based denoising, the choice of time-step for stopping the PDE evolution is a crucial parameter. Some researchers propose to stop the PDE when the variance of the residual equals the variance of the noise, which is assumed to be known [24], [16], [17]. This method ignores higher order statistics of the noise. Others use a hypothesis test between the empirical distribution of the residual and the true noise distribution [15] for polynomial order selection in regression-based smoothing. However the exact variance of the noise or its complete distribution is usually not known in practical situations. A decorrelation-based criterion independently proposed in [13] and [2] does not require any knowledge of the noise distribution except that the noise is independent of the original signal. As per this criterion, the optimal filter parameter is chosen to be one which minimizes the correlation coefficient between the denoised and the residual images, regardless of the noise variance. This criterion however has some problems: (1) in the limit of extreme oversmoothing or undersmoothing, the correlation coefficient is undefined as the denoised image could become a constant image, (2) it is too global a criterion (though using a sum of local measures is a ready alternative) and (3) it ignores higher-order dependencies. A solution to the third

issue is suggested by us in Section 3.

It should be noted that all the aforementioned criteria (as also the ones we suggest in this paper) are necessary but not sufficient for parameter selection. Gilboa et al [11] attempt to alleviate this by selecting a stopping time that seeks to maximize the signal-to-noise-ratio (SNR) directly. Their method however requires an estimate of the rate of change of the covariance between the residual and the noise w.r.t. the filtering parameter. This estimate in turn requires full knowledge of the noise distribution. Saddled with this method is the assumption that the covariance between the residual for any image and the actual noise, can be estimated from a single noise-image generated from the same noise distribution. This assumption is not justified theoretically though experimental results are impressive (see [11] for more details). Vanhamel et al [21] propose a criterion that maximizes an estimate of the correlation between the denoised image and the true, underlying image. This estimate, however, can be computed only by using some assumptions that have only experimental justification. In wavelet thresholding methods, risk based criteria have been proposed for the optimal choice of the threshold for the wavelet coefficients. Some of these (such as SURE - Stein’s unbiased risk estimator) [8] again require complete knowledge of the noise model, whereas others merely use the number of samples in the image (the size of the unknown signal) for computation of the threshold [7]. Although several asymptotic properties have been proved for these estimates, we are not aware of their small sample-size behavior (which is the practical setting). Recently, Brunet et al have developed no-reference quality estimates of the MSE between the denoised image and the true underlying image [3]. These estimates do not require knowledge of the original image, but they do require knowledge of the noise variance and obtain a rough, heuristic estimate of the covariance between the residual and the noise. Moreover the performance of these estimates has been tested only on Gaussian noise models.

3. Theory

3.1 independence measures

In what follows, we shall denote the denoised image obtained by filtering a noisy image I as D , its corresponding residual as R (note that $R = I - D$) and the true image underlying I as J . As mentioned earlier, independence-based criteria have been developed in image processing literature. In cases where a noisy signal is oversmoothed, the residual image clearly shows the distinct features from the image (referred to as method noise in [4]). This is true even in those cases where the noise is independent of the signal. Independence-based criteria are based on the assumption that when the noisy image is filtered optimally, the residual

would contain mostly noise and very little signal and hence it would be independent of the denoised image. It has been experimentally reported in [13] that the absolute correlation coefficient (denoted as CC) between D and R decreases almost monotonically as the filter smoothing parameter is increased (in discrete steps) from a lower bound to a certain ‘optimal’ value, after which its value increases steadily until an upper bound. However, CC ignores anything higher than second-order dependencies. To alleviate this problem, we propose to minimize the mutual information (MI) between D and R , as a criterion for parameter selection. This has been proposed as a (local) measure of noisiness earlier in [3], but it has been used in that paper only as an indicator of areas in the image where the residual is unfaithful to the noise model, rather than as an explicit parameter-selection criterion. In this paper, we also propose to use the following information-theoretic measures of correlation from [6] (see page 47) as independence criteria:

$$\eta_1(R, D) = 1 - \frac{H(R|D)}{H(D)} = \frac{MI(R, D)}{H(D)} \quad (1)$$

$$\eta_2(R, D) = 1 - \frac{H(D|R)}{H(R)} = \frac{MI(R, D)}{H(R)}. \quad (2)$$

Here $H(X)$ refers to the Shannon entropy of X , and $H(X|Y)$ refers to the conditional Shannon entropy of X given Y . η_1 and η_2 both have values bounded from 0 (full independence) to 1 (full dependence).

A problem with all these criteria (CC, MI, η_1, η_2) lies in the inherent probabilistic notion of independence itself. In the extreme case of oversmoothing, the ‘denoised’ image may turn out to have a constant intensity, whereas in the case of extreme undersmoothing (no smoothing or very little smoothing), the residual will be a constant (zero) signal. In such cases, CC, η_1, η_2 are ill-defined whereas MI turns out to be zero (its least possible value). What this indicates is that these criteria have the innate tendency to favor extreme cases of under- or over-smoothing. In practical applications, one may choose to get around this issue by choosing a local minimum of these measures within a heuristically chosen interval in the parameter landscape from 0 to ∞ , but we wish to drive home a more fundamental point about the inherent flaw in using independence measures. Moreover, it should be noted that localized versions of these measures (i.e. sum of local independence measures) may produce false optima if the filtering algorithm smoothes out local regions with fine textures.

3.2 Characterizing residual ‘noisiness’

Given the fact that the assumed noise model is i.i.d. and signal independent, we expect the residual produced by an ideal denoising algorithm to obey these characteristics. Most natural images (apart from homogenous textures)

show a considerable degree of statistical dissimilarity. To demonstrate this, we performed the following experiment on all 480 images from the Berkeley database [12]. Each image at four scales with successive downsampling factor of $\frac{2}{3}$ was tiled into non-overlapping patches of sizes $p \times p$ where $p \in \{16, 24, 32\}$. The rejection rate was on an average 81 percent which indicates that different regions from the images have clearly different distributions. A (two-sample) Kolmogorov-Smirnov (KS) test was run between the intensity values in all pairs of patches at the significance level of $\alpha = 0.05$. It should be noted that the tiling of the image into patches was very important: a KS test between sample subsets from random locations produced very low reject rates. Also, a similar experiment with the same scales and patch sizes run on pure Gaussian noise images resulted in a rejection rate of just about 7 percent. Next, Gaussian noise of $\sigma = 18$ (for intensity range [0,255]) was added to each image. Each image was filtered using the Perona-Malik filter [14] for 90 iterations with a step size of 0.05 and edginess criterion of $K = 40$ and the residual images were computed. At the end of the filtration, the aforementioned KS-test was performed again at $\alpha = 0.05$ between patch pairs from the residual image. The resulting rejection rate was 41 percent, indicating strong heterogeneity in the residual values. We therefore conjecture that statistical heterogeneity in the residuals is a strong indicator of the presence of structure in them. Moreover the percentage reject rate (denoted as h), the average value of the KS-statistic $\sup_x |F_1(x) - F_2(x)|$ (denoted as K) and the average negative logarithm of the p-value¹ from each test (denoted as P) are all indicators of the ‘noisiness’ of a residual (the lower the value, the noisier the residual). Hence these act as criteria for filter parameter selection. It should be noted that while computing P , there is the inherent assumption that the pairwise tests between individual patches are all independent. We prefer the criteria P and K to h because they do not require a significance level to be specified a priori.

The two-sample KS test-statistic is defined as

$$K = \sup_x |F_1(x) - F_2(x)| \quad (3)$$

where $F_1(x)$ and $F_2(x)$ are the respective empirical cumulative distribution functions (ECDF) of the two samples respectively. Under the null hypothesis, the distribution of K is known to be independent of the underlying true CDFs themselves. Therefore the K value has a special meaning in statistics. The advantage of the KS-based measure over MI or CC is that the P and K will keep increasing more and more when the image smoothing continues (as the residual will then contain more and more structure). This unlike MI or CC which will produce false minima. This is demonstrated in Figure 1 where the decreases in the values of MI

¹A p-value is defined as the probability of obtaining the test-statistic value less than or equal to the empirically computed value of the statistic.

or CC at high smoothing levels is quite evident. Just like MI or CC, the KS-based criteria do not require knowledge of the noise distribution or even the exact noise variance. However all these criteria could be fooled by the pathological case of zero or very low denoising. This is because in the very initial stages of denoising (obtained by, say, running a PDE with a very small stepsize for very few iterations), the residual is likely to be devoid of structure and independent of the (under)smoothed image. Consequently, all measures: MI, CC, K and P will acquire (falsely) low values. This problem can be avoided by making assumptions of the range of values for the noise variance (or a loose lower-bound), without requiring exact knowledge of the variance. This has been the strategy followed implicitly in parameter selection experiments in [13] where the PDE stepsizes are chosen to be 0.1 and 1. In all our experiments, we make similar assumptions. The exception is that KS-based measures do not require any upper bound on the variance to be known: just a lower bound suffices.

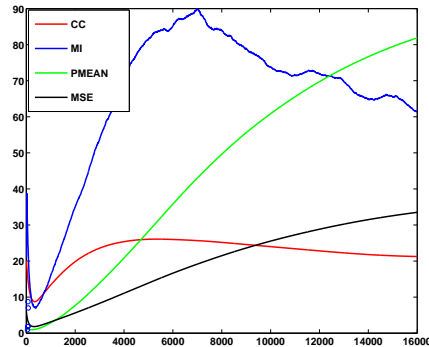


Figure 1. Plots of CC, MI, P , K and MSE on an image subjected to upto 16000 iterations of total variation denoising.

4 Experimental Results

To demonstrate the effectiveness of the proposed criteria, we performed experiments on all the 13 images from Lansel’s benchmark dataset [1]. All images from the dataset were down-sized from 512 x 512 to 256 x 256. We experimented with 6 noise levels $\sigma_n^2 \in \{10^{-4}, 5 \times 10^{-4}, 10^{-3}, 5 \times 10^{-3}, 0.01, 0.05\}$ on an intensity range of [0,1], and with three additive zero-mean noise distributions: Gaussian (the most commonly used assumption), bounded uniform (noise due to quantization) and negative exponential (noise model used for SAR imagery). The lower-bound assumed on the noise variance was 10^{-6} in all experiments. Two filtering algorithms were tested: the non-local means (NL-means)

filter [4] and total variation (TV) [17]. The equation for the NL-means filter is as follows:

$$\hat{I}(x) = \frac{\sum_{x_k \in N(x; SR)} w_k(x) I(x_k)}{\sum_{x_k \in N(x; SR)} w_k(x)}, \quad (4)$$

$$w_k(x) = \exp\left(-\frac{\|q(x; QR) - q(x_k; QR)\|^2}{\sigma}\right) \quad (5)$$

where $\hat{I}(x)$ is the estimated smoothed intensity, $N(x; SR)$ is a search window of diameter SR around point x , $w_k(x)$ is a weight factor, $q(x; QR)$ is a patch of diameter QR centered at x and σ is a smoothing parameter². In our experiments, a patch size of 12 x 12 was used, with a search window of 50 x 50. σ was chosen by running the NL-means algorithm for 55 different σ values for smoothing, from the set $\{1 : 1 : 10, 20 : 20 : 640, 650 : 50 : 1200\}$. The optimal σ values were computed using the following criteria: $CC(D, R)$, $MI(D, R)$, $\eta_1(D, R)$, $\eta_2(D, R)$; sum of localized versions of all above measures on a 12 x 12 window; h , P and K using two-sample KS tests on non-overlapping 12 x 12 patches; and h_n , P_n and K_n values computed using KS-test between the residual and the true noise samples (which we know here as these are synthetic experiments). All information theoretic quantities were computed using 40 bins as the image size was 256 x 256 (the optimal number of bins for n samples is $n^{1/3}$).

The total variation(TV) filter is obtained by minimizing the following energy:

$$E(I) = \int_{\Omega} |\nabla I(x)| dx \quad (6)$$

for an image defined on the domain Ω , which gives a geometric heat equation PDE that is iterated some T times starting from the given noisy image as an initial condition. The stopping time T is the equivalent of the smoothing parameter here. For the TV filter, in addition to all the criteria mentioned before, we also tested Gilboa's method [11] (assuming knowledge of the noise distribution) and Vanhamel's method [21].

4.1 Validation method

In order to validate the σ or T estimates produced by these criteria, it is important to see how well they are in tune with those filter parameter values that are optimal with regard to different well-known quality measures between the denoised image and the original image. The most commonly used quality measure is the MSE. However as documented beautifully in [23], MSE has several limitations.

²Note that we use σ to denote the smoothing parameter of the filtering algorithm and σ_n^2 to denote the variance of the noise.

Hence we also experimented with structured similarity index (SSIM) developed in [22]; with the CC, MI, η_1 and η_2 between the denoised and the original image (as well as with the sum of their local versions); and with the L1 difference between the denoised and the original image.

4.2 Results on NL-Means

Results on NL-Means for Gaussian noise of six different variances are shown in Tables 1 through to 6. In all these tables, Δ_X = absolute difference between 'X' values as predicted by the criterion, and the optimal 'X' value. The 'X' value is defined to be the quality measure 'X' between the denoised and the noisy image. $d\sigma_X$ is the absolute difference between the σ value for NL-means predicted by the algorithm and the σ value when the quality measure X was optimal. Here X stands for the L1 norm or SSIM. The other quality measures are not shown here to save space. The last two rows of the tables indicate the minimum and maximum of the optimal quality measure values across all the 13 images on which the experiments were run (which gives an idea about the scale of the Δ_X values). Some results on the 'stream' and 'mandrill' images are shown in Figure 2 and 3 with the corresponding residuals. Experiments were also performed on images degraded with bounded uniform noise of total width $2 \times 5 \times 10^{-4}$ and $2 \times 5 \times 10^{-3}$ (on an intensity range of [0,1]) with results shown in Tables 7 and 8. Similar experiments were run with negative exponential noise where these criteria performed well. These results are shown in the supplemental material in Tables 7 and 8.

For low and moderate noise levels, it was observed that the criteria P or K produced errors an order of magnitude better than MI, η_1 and η_2 (which were the closest competitors) and even two orders of magnitude better than CC. Our observation was that CC and information theoretic criteria tend to cause undersmoothing for NL-means. At high noise levels, we saw that all criteria produced a high error in prediction of the optimal parameter. An explanation for this is that the NL-means algorithm by itself does not produce very good results at high noise levels, and requires high σ values which produce highly structured residuals. For low σ values, it produces residuals that resemble noise in the sense of various criteria, but this leads to hugely undersmoothed estimates.

An interesting phenomenon we observed was that the same KS-test based measures (i.e. P_n and K_n) between the residuals and the actual noise samples (which we know, as these are synthetic experiments) did not perform as well as the KS-test measures (i.e. P and K) between pairs of patches from the residual. We conjecture that this is owing to biases inherent in the NL-means algorithm (as in many others - see [5]) due to which the residuals have slightly different means and variances as compared to the actual noise,

Table 1. (NLMeans)Gauss. noise $\sigma_n^2 = 0.0001$.

-	Δ_{L1}	$d\sigma_{L1}$	Δ_{SSIM}	$d\sigma_{SSIM}$
h	0.088	10.462	0.002	7.692
P	0.031	7.538	0.005	12.462
K	0.040	7.846	0.004	10.308
CC	0.085	9.846	0.010	17.846
MI	0.189	17.077	0.011	18.615
η_1	0.189	17.077	0.011	18.615
η_2	0.176	18	0.011	20.769
Local MI	0.055	8.769	0.007	16.154
h_{NM}	0.851	31.462	0.009	16.385
P_{NM}	0.087	13.231	0.002	4.923
K_{NM}	0.215	19.385	0.001	5.538
Min	2.225	-	0.888	-
Max	9.147	-	0.986	-

Table 2. (NLMeans)Gauss. noise $\sigma_n^2 = 0.0005$.

-	Δ_{L1}	$d\sigma_{L1}$	Δ_{SSIM}	$d\sigma_{SSIM}$
h	0.029	7.692	0.003	12.769
P	0.010	4.615	0.006	17.385
K	0.014	5.385	0.004	15.077
CC	0.068	8.462	0.007	15.077
MI	0.087	14.615	0.009	19.692
η_1	0.087	14.615	0.009	19.692
η_2	0.232	19.923	0.015	26.538
Local MI	0.155	14.615	0.005	13.538
h_n	0.436	28.154	0.003	8
P_n	0.047	10.769	0.003	9.692
K_n	0.128	16.154	0.001	5.846
Min	2.683	-	0.884	-
Max	9.383	-	0.981	-

even though the residuals may be homogenous.

4.3 Effect of patch size on the KS test

The KS test employed here operates on image patches. The patch-size can be a crucial parameter: too low a patch size (say 2×2) will lead to reduction in the power of the KS test and cause (false) rejection at all smoothing levels, whereas too high a patch size will lead to (false) acceptance at all smoothing levels. We chose a patch size so that the number of samples for estimation of the cumulative was sufficient. This was determined in such a way that the maximum absolute error between the estimated and true underlying CDFs was no more than 0.1 with a probability of 0.9. Then, using the Dvortzky-Kiefer-Wolfowitz inequality, it follows that there should be at least 149 samples [9]. Hence we chose a patchsize of 12×12 . However, we also performed an experiment with NL-means where the KS-test was performed across multiple scales from 12 to 60 in steps of 8 (for an image of size 256×256), and average h , P and K values were calculated. However for the several experiments described in the previous sections, we just used the patchsize of 12×12 , as the multiscale measure did not produce significantly better results.

5 Results on Total Variation

Results for total variation diffusion with Gaussian noise of variance 5×10^{-4} and 5×10^{-3} are shown in Tables 9 and 10. For this method, the KS-based measures performed well in terms of errors in predicting the correct number of iterations and the correct quality measures, but not as well as MI within the restricted stopping time range. The results were also compared to those obtained from Gilboa's method [11] which performed the best, though we would like to remind

Table 3. (NLMeans)Gauss. noise $\sigma_n^2 = 0.001$.

-	Δ_{L1}	$d\sigma_{L1}$	Δ_{SSIM}	$d\sigma_{SSIM}$
h	0.041	9.231	0.004	16.154
P	0.024	5.385	0.005	16.923
K	0.035	7.692	0.004	14.615
CC	0.151	16.154	0.004	12.308
MI	0.126	19.231	0.008	18.462
η_1	0.126	19.231	0.008	18.462
η_2	0.157	26.923	0.013	29.231
Local MI	0.191	20.308	0.003	10.308
h_n	0.218	22.308	0.001	6.154
P_n	0.041	10	0.002	12.308
K_n	0.100	15.385	0.001	6.923
Min	3.069	-	0.879	-
Max	9.601	-	0.976	-

Table 4. (NLMeans)Gauss. noise $\sigma_n^2 = 0.005$.

-	Δ_{L1}	$d\sigma_{L1}$	Δ_{SSIM}	$d\sigma_{SSIM}$
h	0.206	33.846	0.003	24.615
P	0.207	33.846	0.003	24.615
K	0.488	43	0.005	33.769
CC	2.253	92.308	0.034	55.385
MI	2.677	79.538	0.054	67.231
η_1	2.720	81.077	0.054	68.769
η_2	2.119	105.231	0.053	105.231
Local MI	3.889	107.846	0.069	74
h_n	1.337	38	0.032	38
P_n	1.335	33.538	0.033	36.615
K_n	1.336	33.538	0.034	42.769
Min	4.838	-	0.791	-
Max	10.695	-	0.955	-

Table 5. (NLMeans)Gauss. noise $\sigma_n^2 = 0.01$.

-	Δ_{L1}	$d\sigma_{L1}$	Δ_{SSIM}	$d\sigma_{SSIM}$
<i>h</i>	12.121	226.692	0.202	177.462
<i>P</i>	12.121	226.692	0.202	177.462
<i>K</i>	12.121	226.692	0.202	177.462
<i>CC</i>	8.886	207.154	0.149	157.923
<i>MI</i>	11.701	224.462	0.195	175.231
η_1	11.701	224.462	0.195	175.231
η_2	6.218	200	0.119	163.077
Local MI	12.121	226.692	0.202	177.462
h_n	1.891	60.923	0.045	64
P_n	3.642	86.462	0.081	108
K_n	3.649	88	0.082	115.692
Min	6.285	-	0.735	-
Max	11.661	-	0.933	-

Table 6. (NLMeans)Gauss. noise $\sigma_n^2 = 0.05$.

-	Δ_{L1}	$d\sigma_{L1}$	Δ_{SSIM}	$d\sigma_{SSIM}$
<i>h</i>	14.704	906.154	0.183	643.846
<i>P</i>	11.290	838.462	0.140	576.154
<i>K</i>	9.597	805.385	0.118	543.077
<i>CC</i>	18.959	1020	0.249	757.692
<i>MI</i>	19.435	1026.154	0.253	763.846
η_1	19.550	1027.692	0.255	765.385
η_2	19.435	1026.154	0.253	763.846
Local MI	19.783	1030.769	0.258	768.462
h_n	24.516	1028.462	0.305	796.923
P_n	26.721	1120.769	0.325	858.462
K_n	26.721	1120.769	0.325	858.462
Min	11.748	-	0.555	-
Max	17.478	-	0.806	-

Table 7. (NLMeans)Unif. noise width = 0.001.

-	Δ_{L1}	$d\sigma_{L1}$	Δ_{SSIM}	$d\sigma_{SSIM}$
<i>h</i>	0.055	9.692	0.003	10.923
<i>P</i>	0.013	5.692	0.006	14.923
<i>K</i>	0.021	6.615	0.005	14.000
<i>CC</i>	0.087	11.077	0.009	17.231
<i>MI</i>	0.188	13.385	0.007	17.077
η_1	0.188	13.385	0.007	17.077
η_2	0.267	16.846	0.009	19.000
Local MI	0.244	17.077	0.011	21.692
h_n	0.770	30.231	0.008	14.538
P_n	0.054	10.308	0.003	7.231
K_n	0.114	14.462	0.001	4.615
Min	2.339	- 0.887	- 1.195	-
Max	9.176	- 0.985	- 2.011	-

Table 8. (NLMeans)Unif. noise width = 0.01.

-	Δ_{L1}	$d\sigma_{L1}$	Δ_{SSIM}	$d\sigma_{SSIM}$
<i>h</i>	0.034	10.769	0.005	18.462
<i>P</i>	0.027	9.231	0.005	20.000
<i>K</i>	0.042	12.308	0.005	16.923
<i>CC</i>	0.430	34.615	0.004	14.615
<i>MI</i>	0.137	16.923	0.007	27.692
η_1	0.137	16.923	0.007	27.692
η_2	0.125	21.538	0.011	35.385
Local MI	0.477	36.308	0.004	16.308
h_n	0.025	9.231	0.004	20.000
P_n	0.020	9.231	0.006	26.154
K_n	0.025	10.769	0.006	27.692
Min	3.522	- 0.860	- 1.148	-
Max	9.835	- 0.970	- 1.834	-

Table 9. (TV)Gauss. noise $\sigma_n^2 = 0.0005$.

-	Δ_{L1}	dt_{L1}	Δ_{SSIM}	dt_{SSIM}
<i>h</i>	0.558	53.462	0.006	56.538
<i>P</i>	0.522	48.462	0.006	52.308
<i>K</i>	0.513	46.538	0.006	50.385
<i>CC</i>	3.487	365.000	0.088	371.923
<i>MI</i>	0.103	20.769	0.001	23.077
η_1	0.103	20.769	0.001	23.077
η_2	2.478	267.692	0.062	274.615
Local MI	0.479	36.923	0.005	32.308
h_n	0.538	69.615	0.007	76.538
P_n	0.523	68.846	0.007	75.769
K_n	0.528	69.615	0.007	76.538
Gilboa [11]	0.050	10.231	0.001	16.385
VanHamel [21]	0.818	51.954	0.009	45.031
Min	2.622	-	0.975	-
Max	4.426	-	0.995	-

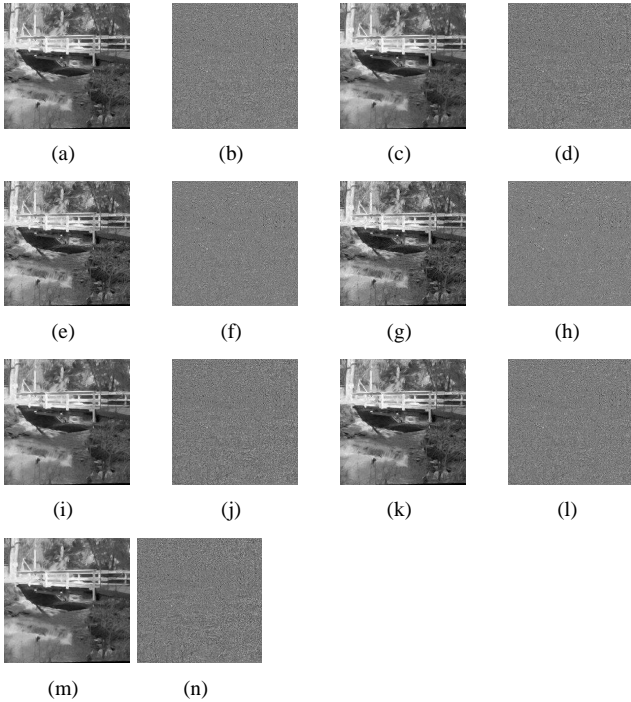


Figure 2. Images with Gaussian noise with $\sigma_n^2 = 5 \times 10^{-3}$ denoised by NLMeans. Parameter selected for optimal noisiness measures: (a) P , (c) K , (e) CC , (g) MI ; and optimal quality measures: (i) $L1$, (k) $SSIM$, (m) MI between denoised image and residual. Corresponding residuals in (b),(d),(f),(h),(j),(l),(n).

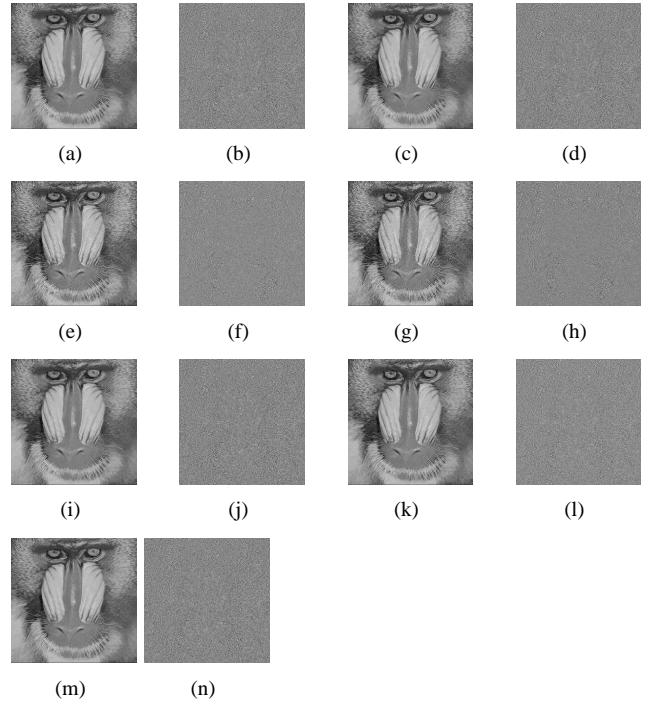


Figure 3. Images with Gaussian noise with $\sigma_n^2 = 5 \times 10^{-3}$ denoised by NLMeans. Parameter selected for optimal noisiness measures: (a) P , (c) K , (e) CC , (g) MI ; and optimal quality measures: (i) $L1$, (k) $SSIM$, (m) MI between denoised image and residual. Corresponding residuals in (b),(d),(f),(h),(j),(l),(n).

the reader that Gilboa’s method requires knowledge of the full noise distribution. Also, in the case of total variation, the KS-based measures did not outperform MI. An explanation for this is that the total variation method is unable to produce homogenous residuals for its optimal parameter set, as it is specifically tuned for piecewise constant images. This assumption does not hold good for commonly occurring natural images. As against this, NL-means is a filter expressly derived from the assumption that patches in ‘clean’ natural images (and those with low or moderate noise) have several similar patches in distant parts of the image.

6 Discussion and Avenues for Future Work

In this paper, we have contributed to and critiqued independence-based criteria for filter parameter selection and presented a criterion that measures the homogeneity of the residual statistics. On the whole we have contributed to the paradigm of exploiting statistical properties of the residual images for driving the denoising algorithm. The

proposed noisiness measures require no other assumptions except that the noise should be i.i.d. and additive (and that a loose lower bound on the noise variance be known, to prevent the measures from attaining falsely low values in cases of extreme under-smoothing). Unlike CC or MI , our noisiness measure is guaranteed not to yield false minima in case of over-smoothing.

The homogeneity based noisiness criterion requires averaging of the P or K values from different patches. For future work, this can be replaced by performing k -sample versions of the Kolmogorov-Smirnov tests (or related tests such as Cramer von-Mises and Anderson Darling) between individual patches versus a large pooled sample containing the entire residual image.

The assumption of i.i.d. noise may not hold in some denoising scenarios. In case of a zero-mean Gaussian noise model with intensity dependent variances, a heuristic approach would be to normalize the residuals suitably using feedback from the denoised intensity values (regarding them as the ‘true’ image values) and then running the KS-tests. The efficacy of this approach needs to be tested on

Table 10. (TV)Gauss. noise $\sigma_n^2 = 0.005$.

-	Δ_{L1}	dt_{L1}	Δ_{SSIM}	dt_{SSIM}
h	0.665	129.615	0.008	102.692
P	0.493	109.615	0.006	80.385
K	0.430	102.308	0.006	73.077
CC	2.156	350.769	0.073	376.923
MI	0.422	88.846	0.012	118.077
η_1	0.422	88.846	0.012	118.077
η_2	1.849	296.923	0.063	331.538
Local MI	5.475	270.769	0.084	240.769
h_n	0.194	59.615	0.008	96.538
P_n	0.221	74.231	0.008	115.000
K_n	0.216	75.769	0.008	116.538
Gilboa [11]	0.094	60.000	0.002	42.000
VanHamel [21]	4.658	303.400	0.068	262.631
Min	4.995	-	0.892	-
Max	11.284	-	0.980	-

denoising algorithms that are capable of handling intensity dependent noise. In case the noise obeys a Poisson distribution (for which the variance is equal to the mean, and hence which is neither fully additive nor multiplicative), there are two ways to proceed: either apply a variance stabilizer transformation which converts the data into that corrupted by Gaussian noise with variance of one, or else suitably change the definition of the residual itself.

Moreover, the existence of a universally optimal parameter selector is not yet established. In other words, different criteria may perform better or worse for different classes of denoising algorithms or with different noise distributions or assumptions on the noise model. This is, as per our survey of the literature, an open problem in image processing. Lastly, despite encouraging experimental results, there is no established theoretical relationship between the performance of noisiness criteria for filter parameter selection and the ‘ideal’ parameters (which themselves are dependent upon criteria like MSE that are not universally agreed upon). For this it is important to study risk-based criteria such those in [7] or else no-reference measures of image quality driven by noise models.

References

- [1] <http://www.stanford.edu/~slansel/DenoiseLab/download.htm>.
- [2] J.-F. Aujol, G. Gilboa, T. Chan, and S. Osher. Structure-texture image decomposition—modeling, algorithms, and parameter selection. *Int. J. Comput. Vision*, 67(1):111–136, 2006.
- [3] D. Brunet, E. Vrscaj, and Z. Wang. The use of residuals in image denoising. In *ICIAR*, pages 1–12, 2009.
- [4] A. Buades, B. Coll, and J.-M. Morel. A review of image denoising algorithms, with a new one. *Multiscale modelling and simulation*, 4(2):490–530, 2005.
- [5] P. Chatterjee and P. Milanfar. Is denoising dead? *Accepted to IEEE Trans. on Image Processing*, 2010.
- [6] T. Cover and J. Thomas. *Elements of Information Theory*. Wiley-Interscience, New York, USA, 1991.
- [7] D. Donoho and I. Johnstone. Ideal adaptation via wavelet shrinkage. *Biometrika*, 81:425–455, 1994.
- [8] D. Donoho and I. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 90(432):1200–1224, 1995.
- [9] A. Dvoretzky, J. Kiefer, and J. Wolfowitz. Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Annals of Mathematical Statistics*, 27(3):642–669, 1956.
- [10] M. Elad and M. Aharon. Image denoising via learned dictionaries and sparse representation. In *CVPR*, pages 1–6, 2006.
- [11] G. Gilboa, N. Sochen, and Y. Zeevi. Estimation of optimal pde-based denoising in the snr sense. *Image Processing, IEEE Transactions on*, 15(8):2269–2280, Aug. 2006.
- [12] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *ICCV*, volume 2, pages 416–423, 2001.
- [13] P. Mrázek and M. Navara. Selection of optimal stopping time for nonlinear diffusion filtering. *Int. J. Comput. Vision*, 52(2-3):189–203, 2003.
- [14] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639, 1990.
- [15] J. Polzehl and V. Spokoiny. Image denoising: a pointwise adaptive approach. *Annals of Statistics*, 31(1):30–57, 2003.
- [16] L. Rudin and S. Osher. Total variation based image resoration with free local constraints. In *International Conference on Image Processing*, pages 31–35, 1994.
- [17] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.
- [18] D. Seo and B. Vemuri. Complex diffusion on scalar and vector valued image graphs. In *EMMCVPR*, pages 98–111, 2009.
- [19] Ö. Subakan, B. Jian, B. Vemuri, and E. Vallejos. Feature preserving image smoothing using a continuous mixture of tensors. In *ICCV*, pages 1–6, 2007.
- [20] H. Takeda, S. Farsiu, and P. Milanfar. Kernel regression for image processing and reconstruction. *IEEE Transactions on Image Processing*, 16(2):349–366, 2007.
- [21] I. Vanhamel, C. Mihai, H. Sahli, A. Katartzis, and I. Pratikakis. Scale selection for compact scale-space representation of vector-valued images. *Int. J. Comput. Vision*, 84(2):194–204, 2009.
- [22] Z. Wang and A. Bovik. A universal image quality index. *IEEE Signal Processing Letters*, 9(3):81–84, 2002.
- [23] Z. Wang and A. Bovik. Mean squared error: love it or leave it? - a new look at signal fidelity measures. *IEEE Signal Processing Magazine*, 26(1):98–117, 2009.
- [24] J. Weickert. Coherence-enhancing diffusion of colour images. *Image and Vision Computing*, 17:201212, 1999.