

# The Use of Residuals in Image Denoising

Dominique Brunet<sup>1</sup>, Edward R. Vrscay<sup>1</sup> and Zhou Wang<sup>2</sup>

<sup>1</sup> Department of Applied Mathematics, Faculty of Mathematics,

<sup>2</sup> Department of Electrical and Computer Engineering, Faculty of Engineering,  
University of Waterloo, Waterloo, Ontario, Canada N2L 3G1  
dbrunet@uwaterloo.ca, ervrscay@uwaterloo.ca, zhouwang@ieee.org

**Abstract.** State-of-the-art image denoising algorithms attempt to recover natural image signals from their noisy observations, such that the statistics of the denoised image follow the statistical regularities of natural images. One aspect generally missing in these approaches is that the properties of the residual image (defined as the difference between the noisy observation and the denoised image) have not been well exploited. Here we demonstrate the usefulness of residual images in image denoising. In particular, we show that well-known full-reference image quality measures such as the mean-squared-error and the structural similarity index can be estimated from the residual image without the reference image. We also propose a procedure that has the potential to enhance the image quality of given image denoising algorithms.

**Key words:** image denoising, residual image, no-reference image quality assessment, structural similarity, spatial statistics

## 1 Introduction

The problem of image denoising may be conveniently summarized as follows: Given a noisy version  $\mathbf{y}$  of an image  $\mathbf{x}$  – for example, the image that might be obtained when image  $\mathbf{x}$  is sent over a noisy transmission channel – construct a denoising operator  $\mathcal{D}$  so that  $\mathcal{D}(\mathbf{y})$  is “close” to  $\mathbf{x}$ . There are, of course, some concepts in the above summary that must be addressed more concretely in any such denoising study, including: (1) “noise,” i.e., a model for the corruption by noise is needed, and (2) “closeness”, i.e., a measure for image quality is needed.

In this paper, we are interested in denoising natural images contaminated with additive white Gaussian noise. Let  $\mathbf{x}$  be a greyscale image represented by a matrix defined on a rectangular lattice  $I = \{i = (i_1, i_2)\}_{1 \leq i_1 \leq m_1, 1 \leq i_2 \leq m_2}$ . Also assume that  $\mathbf{x}$  belongs to the set of natural images  $X$ , i.e., images that are likely to be seen by the human visual system. Let  $N$  be independent white Gaussian noise on  $I$  of zero-mean and variance  $\sigma^2$ . More precisely, the following assumptions are made:

1.  $X$  and  $N$  are independent random fields;
2.  $N(i)$  and  $N(j)$  are independent for  $i \neq j$  ( $i, j \in I$ );
3. The noise is Gaussian, i.e.  $N(i) \sim \mathcal{N}(0, \sigma^2)$  for all  $i \in I$ .

We can then write the noise distortion model as  $Y = X + N$ ; a particular outcome will be represented by  $\mathbf{y} = \mathbf{x} + \mathbf{n}$ . For any noisy image  $Y$ , the action of a denoiser may be viewed as separating  $Y$  into the sum of a denoised image  $D = \mathcal{D}(Y)$  and a *residual image* defined by  $R = Y - D$ . To illustrate a particular outcome, in Fig. 1 we show the original noiseless *Lena* image  $\mathbf{x}$  ( $512 \times 512$  pixels) along with a noise image  $\mathbf{n}$  and the resulting noisy image  $\mathbf{y} = \mathbf{x} + \mathbf{n}$ . A denoised image  $\mathbf{d} = \mathcal{D}(\mathbf{y})$  obtained by Gaussian filtering is then shown, along with the residual image  $\mathbf{r} = \mathbf{y} - \mathbf{d}$ . For the moment, we mention that in most, if not all, image denoising studies, the noiseless image  $\mathbf{x}$  is available for testing, enabling a direct comparison of  $\mathbf{d}$  and  $\mathbf{x}$  for the evaluation of denoising operators. But what if  $\mathbf{x}$  is not available, as is the case in practical situations? Can one estimate the efficacy of the denoising operator from knowledge of  $\mathbf{d}$  and  $\mathbf{y}$  alone? This is the essence of *no-reference* quality assessment for image denoising.

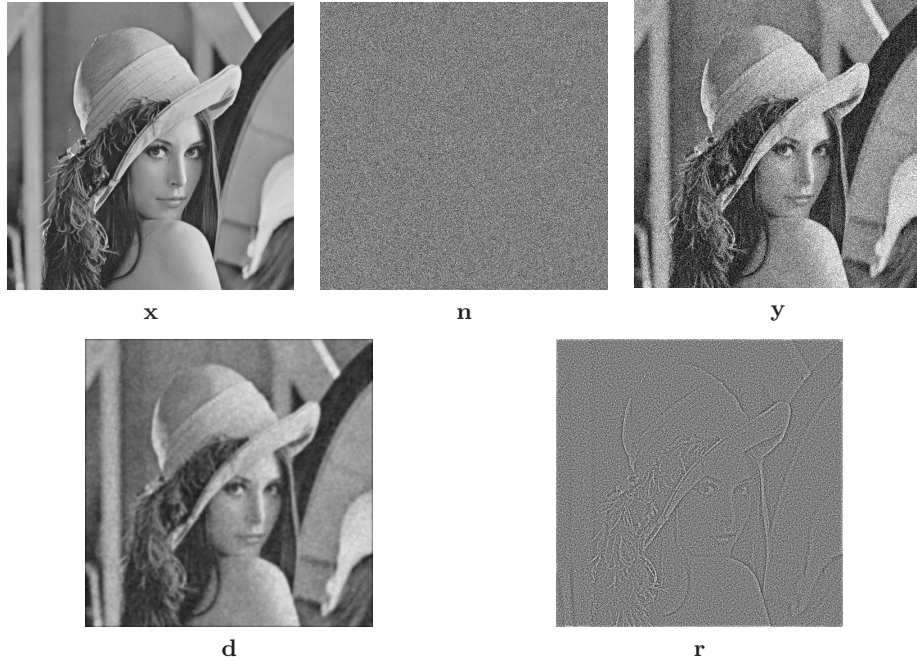
Separating the noisy image  $\mathbf{y}$  into the sum of  $\mathbf{d}$  and  $\mathbf{r}$  is an ill-posed problem. In order to provide a meaningful solution, additional knowledge about the target images  $\mathbf{d}$  and  $\mathbf{r}$  must be imposed. Statistical models of natural images have attracted great interest, providing useful statistical prior knowledge about the clean natural image, and thus helping the denoiser to distinguish the image signal from noise. One useful aspect generally missing from state-of-the-art image denoising algorithms is that the properties of the residual image  $\mathbf{r}$  are not well explored. More precisely, these algorithms attempt to approximate  $\mathbf{x}$  with  $\mathbf{d}$  (by using the prior knowledge about  $\mathbf{x}$ ), but ignore the desirable property that  $\mathbf{r}$  should approximate  $\mathbf{n}$ . This motivates us to study the use of  $\mathbf{r}$  in image denoising by looking at how well  $\mathbf{r}$  approximates  $\mathbf{n}$ . In the literature, most examinations of the residual image have involved only a visual inspection of  $\mathbf{r}$  [1]. Several authors have employed  $\mathbf{r}$  in the design of image denoising algorithms [7, 3, 10, 9, 11, 15]. Nevertheless, to the best of our knowledge, statistical analyses of  $\mathbf{r}$  (i.e., whether  $\mathbf{r}$  is a valid sample from a particular noise distribution) and how such statistics can be used for no-reference image quality assessment and for improving image denoising results, have not been deeply investigated.

## 2 Statistical Tests on the Image Residual

Here, the idea is to apply a statistical test locally on the residual image to determine whether or not it behaves like pure noise. Indeed, some assumptions on the nature of the noise were made, and we can test if they are observed in the residual image. We employ two kinds of hypothesis tests about the residual image: (1) independence and (2) goodness-to-fit. The first test will help determine whether or not  $\mathbf{d}$  and  $\mathbf{r}$  are independent and if there is any autocorrelation in  $\mathbf{r}$ . The second test will hopefully show if  $\mathbf{r}$  follows a normal distribution and if the intensity values of  $\mathbf{r}$  are identically distributed with zero-mean and variance  $\sigma^2$ .

### 2.1 Testing the Independence between $\mathbf{d}$ and $\mathbf{r}$

We examined the use of two well-known statistical tests: Pearson's correlation coefficient test and the maximum-likelihood ratio test (or G-test).



**Fig. 1.** **x** : Original noiseless *Lena* image. **n** : a noise image ( $\sigma = 35$ ). **y** := **x** + **n**, the noisy image. **d** :=  $\mathcal{D}(\mathbf{y})$ , the denoised image (Gaussian filtering). **r** := **y** - **d**, the residual image. Notice that **y** = **x** + **n** and **y** = **d** + **r** are two different decompositions of the noisy image **y**.

**Pearson's correlation coefficient test:** Given  $n$  data points  $(x, y)$ , the (biased) correlation coefficient is defined by

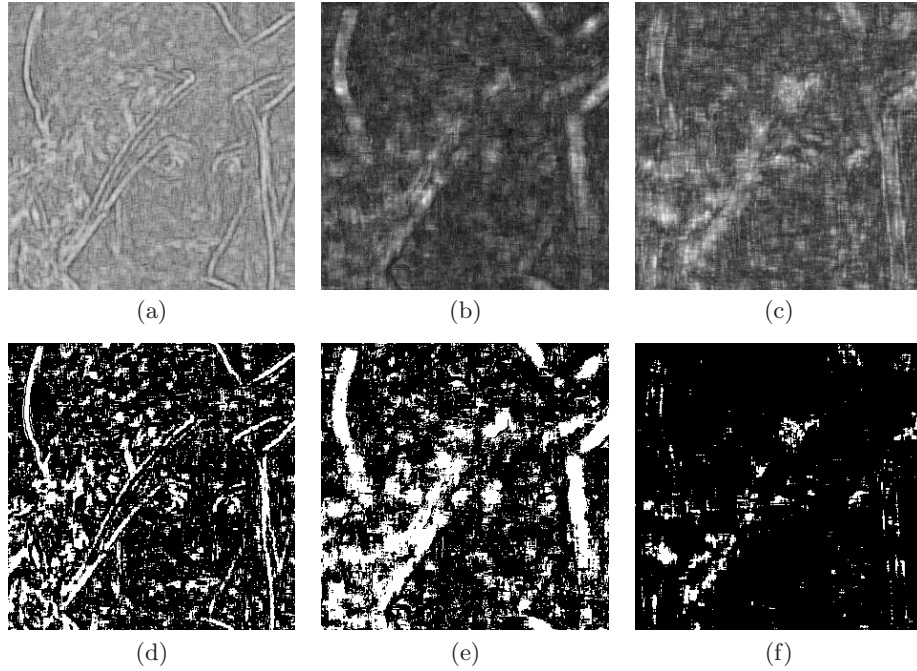
$$r = \frac{s_{xy}}{s_x s_y}, \quad (1)$$

where  $s_x$  and  $s_y$  are the (biased) sample standard deviations of, respectively,  $x$  and  $y$  and  $s_{xy}$  is the sample covariance of  $x$  and  $y$ . We use a criterion based on the value

$$t = r \sqrt{\frac{n-2}{1-r^2}}. \quad (2)$$

The criterion  $t$  follows a Student-t distribution with  $n-2$  degrees of freedom. The test is only valid if the data is randomly sampled from a normal distribution, but we will later test for normality as well. The main advantage of this test is that the joint distribution does not need to be computed. As such, validity is achieved with smaller sample sizes. In addition, the data does not need to be binned. That being said, the correlation coefficient describes only linear dependency and lack of robustness.

Using the same images as in Fig. 1, an example of the correlation coefficient between a denoised image  $\mathbf{d}$  and its residual  $\mathbf{r}$  is shown in Fig. 2(a). In Fig. 2(d) we show the result of a correlation coefficient test between  $\mathbf{d}$  and  $\mathbf{r}$ . A  $7 \times 7$  sliding window was used to compute the local correlation.



**Fig. 2.** (a): Pearson's correlation coefficient between  $\mathbf{d}$  and  $\mathbf{r}$  (black for negative correlation and white for positive correlation). (b): The local mutual information map between  $\mathbf{d}$  and  $\mathbf{r}$ . (c): The local Kolmogorov-Smirnov statistics of  $\mathbf{r}$ . (d): The result of the Pearson's correlation coefficient test (white: reject independence hypothesis, black: do not reject) (e): The result of the G-test (white: reject independence hypothesis, black: do not reject) (f): The result of the K-S test (white: reject normality hypothesis, black: do not reject). All images were cropped between (100, 100) and (400, 400).

**G-test:** Given  $n$  data points  $(x, y)$  and a  $I \times J$  binning for these points, the G-test is based on the mutual information

$$MI(x; y) = \sum_{i=1}^I \sum_{j=1}^J p_{i,j} \ln \left( \frac{p_{i,j}}{p_i p_j} \right), \quad (3)$$

where  $p_{i,j}$  is the sample joint probability that  $x$  is in the  $i$ -th bin and  $y$  is in the  $j$ -th bin and  $p_i$  and  $p_j$  are the sample marginal probabilities that,  $x$  is in

the  $i$ -th bin and  $y$  is in the  $j$ -th bin, respectively. The criterion  $G = 2nMI(x; y)$  follows a chi-square distribution with  $r = (I - 1)(J - 1)$  degrees of freedom. The advantage of this method over Pearson’s correlation coefficient is that it takes account of all types of dependencies. The drawback, however, is that a large number of samples is required to estimate the joint distribution.

Even though it is desired to compute local statistics, a sufficient amount of data is required to obtain good estimates of the joint distribution. Following the rule-of-thumb,  $\#bins = n^{1/3}$ , a  $15 \times 15$  window was used along with 6 bins as a tradeoff. (The binning can be adapted according to the local range of the intensity values.) Using the images in Fig. 1, we present an example of the mutual information computed between a denoised image  $\mathbf{d}$  and its residual  $\mathbf{r}$  in Fig. 2(b). In Fig. 2(e), a G-test has been performed between  $\mathbf{d}$  and  $\mathbf{r}$ .

## 2.2 Testing the Autocorrelation of $\mathbf{r}$

Let  $\mathbf{r}|_{W_i}$  denote the restriction of  $\mathbf{r}$  over the window  $W_i$  centered at  $i \in I$ . For any  $i, j \in I$  ( $i \neq j$ ), we wish to test the autocorrelation in  $\mathbf{r}|_{W_i}$  and  $\mathbf{r}|_{W_j}$ . The test is restricted to pairs of windows that are close to each other, i.e., ( $\|i - j\|_1 < b$ ), where  $b$  is the radius of the search. The local autocorrelation coefficient  $r_{i,j}$  is computed as the correlation coefficient (see Eq. 1) between  $\mathbf{r}|_{W_i}$  and  $\mathbf{r}|_{W_j}$ . We can then perform a series of hypothesis tests as in the Pearson’s correlation coefficient test. A sample is considered to be somewhat autocorrelated if we reject at least once the hypothesis that the local autocorrelation of  $\mathbf{r}$  is zero.

## 2.3 Testing the Normality of $\mathbf{r}$

Several goodness-of-fit tests are available to test the normality of a given sample. Here we consider the Kolmogorov-Smirnov (K-S) test, applicable to any continuous distribution with the requirement, however, that all parameters are known. The K-S statistic is defined by

$$k = \sup_{x \in [0,1]} |E(x) - F(x)|, \quad (4)$$

where  $F$  is the theoretical cumulative distribution function of the distribution being tested and where  $E$  is the empirical cumulative distribution function of the sample. The hypothesis is rejected if the test statistic,  $k$ , is greater than a computed critical value.

Using again the residual image  $\mathbf{r}$  from Fig. 1, the K-S test was performed locally, using a  $15 \times 15$  sliding window. The test can reveal at the same time if there is a departure from normality and if the mean and variance are, respectively, 0 and  $\sigma^2$  over the entire residual image. Figure 2(c) shows the resulting local K-S statistics. In Fig. 2(f) are presented the results of the hypothesis test.

### 3 No-reference estimation of full-reference quality measures

#### 3.1 Estimating the covariance of the noise and the residual

The covariance between the random fields  $R$  and  $N$ ,

$$\sigma_{R,N} := \mathbf{E}[(R - \mathbf{E}[R]) \cdot (N - \mathbf{E}[N])] = \mathbf{E}[R \cdot N], \quad (5)$$

must be estimated from a knowledge of  $\mathbf{r}$  and the statistics of the noise. Here,  $\cdot$  denotes the Hadamard (term-by-term) matrix product and  $\mathbf{E}[R]$  denotes the matrix of the expectations of each entries.

When a noisy image  $Y$  is denoised, some or nearly all the noise may be removed, but some blurred or distorted structures of the original image may still remain. As such, we model the denoised image and the residual as follows,

$$Y = X + N = (X_{(d)} + N_{(d)}) + (X_{(r)} + N_{(r)}) = D + R \quad (6)$$

where  $X_{(d)}$  and  $X_{(r)}$  are the structure parts in, respectively,  $D$  and  $R$  and  $N_{(d)}$  and  $N_{(r)}$  are the noise parts in, respectively,  $D$  and  $R$ . We assume that  $X_{(d)}$  and  $X_{(r)}$  are independent from both  $N_{(d)}$  and  $N_{(r)}$ . For purposes of estimation, we also assume that  $\sigma_{N_{(d)},N_{(r)}} \geq 0$ . This seems reasonable since any noise in  $D$  will be an attenuation of the noise already present in  $Y = D + R$ . Under all of these assumptions, one can then show that

$$\sigma_{R,N} = \sigma_{X_{(r)},N_{(d)}} + \sigma_{X_{(r)},N_{(r)}} + \sigma_{N_{(r)},N_{(d)}} + \sigma_{N_{(r)}}^2 = \sigma_{N,N_{(r)}}, \quad (7)$$

$$\sigma_{R,N} = \sigma_{R,D+R-X} = \mathbf{E}[D \cdot R] + \mathbf{E}[R \cdot R] - \mathbf{E}[R \cdot X] = \mathbf{E}[R^2] - \mathbf{E}[X_{(r)}^2], \quad (8)$$

$$\sigma_{R,N} = \sigma_{R,Y-X} = \sigma_{R,Y} - \sigma_{R,X}, \quad (9)$$

$$\sigma^2 = \sigma_N^2 = \sigma_{N_{(d)}}^2 + \sigma_{N_{(d)},N_{(r)}} + \sigma_{N,N_{(r)}}. \quad (10)$$

With the extra assumption that  $\sigma_{X,R} \geq 0$ , i.e. structure in  $X$  may be present to some extent in  $R$  as well, we have the result

$$0 \leq \sigma_{R,N} \leq \min(\mathbf{E}[R^2], \sigma_{Y,R}, \sigma^2). \quad (11)$$

In fact, we observe experimentally that  $\sigma_{R,N} \approx \text{R.H.S.}$  for almost all the denoising algorithms. For this reason, we have chosen

$$\hat{\sigma}_{R,N} = \min(\mathbf{E}[R^2], \sigma_{Y,R}, \sigma^2). \quad (12)$$

In addition, an estimate of  $\sigma_{D,N}$  may also be obtained as well, since

$$\sigma_{D,N} + \sigma_{R,N} = \sigma^2. \quad (13)$$

### 3.2 No-reference Peak Signal-to-Noise Ratio estimate

The mean-square error (MSE) between the original image  $\mathbf{x}$  and the denoised image  $\mathbf{d}$  is

$$\text{MSE}(\mathbf{x}, \mathbf{d}) = \frac{1}{m_1 m_2} \sum_{i_1=1}^{m_1} \sum_{i_2=1}^{m_2} (\mathbf{x}_i - \mathbf{d}_i)^2, \quad (14)$$

where  $i = (i_1, i_2) \in I$  are the indices of the images. The peak signal-to-noise ratio (PSNR) is computed from the mean-square error by the following formula:

$$\text{PSNR}(\mathbf{x}, \mathbf{d}) = 10 \log_{10} \left( \frac{L^2}{\text{MSE}(\mathbf{x}, \mathbf{d})} \right), \quad (15)$$

where  $L$  is the dynamic range of the images.

Observing that

$$\text{MSE}(\mathbf{x}, \mathbf{d}) = \text{MSE}(\mathbf{r}, \mathbf{n}) = \frac{1}{m_1 m_2} \sum_{i_1=1}^{m_1} \sum_{i_2=1}^{m_2} (\mathbf{r}_i^2 + \mathbf{n}_i^2 - 2\mathbf{r}_i \mathbf{n}_i), \quad (16)$$

our no-reference estimate is derived as follows (using Eq. 12):

$$\widehat{\text{MSE}}(\mathbf{x}, \mathbf{d}) = \overline{\mathbf{r}^2} + \sigma^2 - 2 \min \left( \overline{\mathbf{r}^2}, s_{\mathbf{y}\mathbf{r}}, \sigma^2 \right), \quad (17)$$

where  $\overline{\mathbf{r}^2}$  is the sample mean of  $\mathbf{r}^2$  and  $s_{\mathbf{y}\mathbf{r}}$  the sample covariance of  $\mathbf{y}$  and  $\mathbf{r}$ . The no-reference estimate of the PSNR follows directly from the formula (15) applied to our estimate (17).

The results of a number of experiments comparing PSNR to its no-reference estimate are shown in Fig. 3(a). We used three images (*Lena*, *Boat* and *Barbara*), artificially adding white Gaussian noise with seven different standard deviation values ( $\sigma = 10, 15, 25, 35, 50, 70$  and  $100$ ). Ten denoising methods (Gaussian filtering, wavelet soft-thresholding [6], stationary wavelet hard-thresholding [4], anisotropic filter [12], total variation minimization [14], Wiener adaptive filter [8], curvelet denoising [2], scaled mixtures of Gaussian [13], non-local means [1] and sparse 3D transform-domain collaborative filtering [5]) were then applied to the resulting noisy images. A last “denoising” procedure is a pathological one, where the denoised image is the noisy image added with a Gaussian white noise image of variance  $\sigma^2$  (i.e.  $\mathbf{d} = \mathbf{y} + \mathbf{m}$  where  $\sigma_M^2 = \sigma_N^2$  and  $\sigma_{M,N} = \sigma_{X,M} = 0$ ).

### 3.3 No-reference Structural Similarity Map

The Structural Similarity (SSIM) index [16] combines information on the local luminance, the local contrast and the local correlation between two greyscale images. It assumes values between 1 for a perfect match between the two images and  $-1$  for very poor quality. The SSIM map between the images  $\mathbf{x}$  and  $\mathbf{d}$  is

$$\text{SSIM}(\mathbf{x}, \mathbf{d}) = \left( \frac{2\overline{\mathbf{x}} \cdot \overline{\mathbf{d}} + c_1}{\overline{\mathbf{x}^2} + \overline{\mathbf{d}^2} + c_1} \right) \cdot \left( \frac{2s_{\mathbf{x}} \cdot s_{\mathbf{d}} + c_2}{s_{\mathbf{x}}^2 + s_{\mathbf{d}}^2 + c_2} \right) \cdot \left( \frac{s_{\mathbf{x}\mathbf{d}} + c_3}{s_{\mathbf{x}} \cdot s_{\mathbf{d}} + c_3} \right), \quad (18)$$

where  $\bar{\mathbf{x}}$  and  $\bar{\mathbf{d}}$  are the local mean map of  $\mathbf{x}$  and  $\mathbf{d}$  respectively,  $s_{\mathbf{x}}$  and  $s_{\mathbf{d}}$  are the local standard deviation map of  $\mathbf{x}$  and  $\mathbf{d}$  respectively,  $s_{\mathbf{xd}}$  is the local covariance map between  $\mathbf{x}$  and  $\mathbf{d}$  and  $\cdot$  is the Hadamard matrix product. The mean is taken over all the entries of the resulting quality map (18) in order to produce a single SSIM index. A local map is obtained by convolving the image with a circular Gaussian filter  $g$ :

$$\bar{\mathbf{x}} = g * \mathbf{x}, \quad (19)$$

$$s_{\mathbf{x}}^2 = g * (\mathbf{x} \cdot \mathbf{x}) - (g * \mathbf{x}) \cdot (g * \mathbf{x}), \quad (20)$$

$$s_{\mathbf{xd}} = g * (\mathbf{x} \cdot \mathbf{d}) - (g * \mathbf{x}) \cdot (g * \mathbf{d}). \quad (21)$$

The constants  $c_1$ ,  $c_2$  and  $c_3$  are small parameters that ensure numerical stability of the division. They may also be chosen to model the human vision system [16].

From the knowledge of the noisy image  $\mathbf{y}$  and a model of the noise image  $\mathbf{n}$ , we can obtain some information about  $\mathbf{x}$ . Indeed, we have

$$\bar{\mathbf{y}} = \bar{\mathbf{x}} + \bar{\mathbf{n}} \quad \text{and} \quad s_{\mathbf{y}}^2 = s_{\mathbf{x}}^2 + 2s_{\mathbf{xn}} + s_{\mathbf{n}}^2, \quad (22)$$

as well as (see Sec. 1)  $\mathbf{E}[N] = 0$ ,  $\sigma_{X,N} = 0$  and  $\sigma_N^2 = \sigma^2$ . Therefore

$$\hat{\bar{\mathbf{x}}} = \bar{\mathbf{y}} \quad \text{and} \quad \hat{s}_{\mathbf{x}} = \sqrt{\max(0, s_{\mathbf{y}}^2 - \sigma^2)}. \quad (23)$$

The maximum in the standard deviation approximation ensures that the square root of a non-negative number is taken. Indeed, it can happen that  $s_{\mathbf{y}}^2 - \sigma^2 < 0$  in flat regions of  $\mathbf{x}$  where the noise dominates. It remains to approximate  $s_{\mathbf{xd}}$ , the sample covariance between  $\mathbf{x}$  and  $\mathbf{d}$ . By assumption,  $s_{\mathbf{xd}} = s_{\mathbf{yd}} - s_{\mathbf{dn}}$ . Using our estimate  $\hat{s}_{\mathbf{dn}}$  (Eq. 12 and 13), we finally obtain

$$\hat{s}_{\mathbf{xd}} = \max\left(0, s_{\mathbf{yd}} - \sigma^2 + \min[\bar{\mathbf{r}}^2, s_{\mathbf{yr}}, \sigma^2]\right). \quad (24)$$

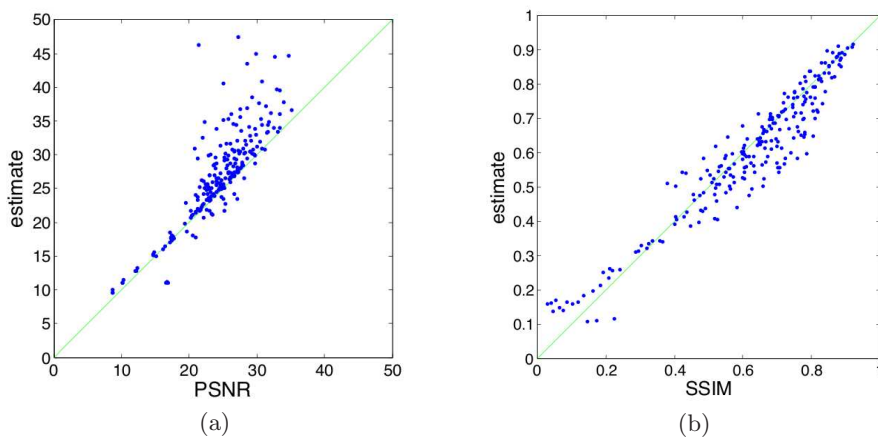
The positivity of  $\hat{s}_{\mathbf{xd}}$  is imposed because a positive correlation between  $\mathbf{x}$  and  $\mathbf{d}$  is expected. Gathering the estimates (24) and (23) in the SSIM equation (18), we obtain the no-reference SSIM map

$$\widehat{\text{SSIM}}(\mathbf{x}, \mathbf{d}) = \left(\frac{2\hat{\bar{\mathbf{x}}} \cdot \bar{\mathbf{d}} + c_1}{\hat{\bar{\mathbf{x}}}^2 + \bar{\mathbf{d}}^2 + c_1}\right) \cdot \left(\frac{2\hat{s}_{\mathbf{x}} \cdot s_{\mathbf{d}} + c_2}{\hat{s}_{\mathbf{x}}^2 + s_{\mathbf{d}}^2 + c_2}\right) \cdot \left(\frac{\hat{s}_{\mathbf{xd}} + c_3}{\hat{s}_{\mathbf{x}} \cdot s_{\mathbf{d}} + c_3}\right). \quad (25)$$

Proceeding as in the experiment for the no-reference PNSR estimate, we present in Fig. 3(b) a scatter plot of quality indices against their estimates for different images, noise levels and denoising algorithms. We observe a good estimate of the SSIM index when the noise level is reasonable, but for high levels of noise ( $\sigma > 50$ ), the estimate is less accurate.

When comparing the full-reference structural similarity index map (Eq. 18) with its no-reference estimate (Eq. 25), we found that  $\hat{s}_{\mathbf{x}}$  slightly overestimates  $s_{\mathbf{x}}$  in flat regions of  $\mathbf{x}$ , creating instabilities in the correlation term. Nevertheless, on average we obtain very promising results.





**Fig. 3.** (a) Scatter plot of the PNSR and its no-reference estimate. (b) Scatter plot of the SSIM index and its no-reference estimate. In both cases, the estimates are generally good for all denoising algorithms examined.

## 4 Quality Enhancement of Denoised Images

In this section we present the results of some preliminary experiments which show that denoising a residual image and adding it to a denoised image can potentially enhance the quality of the latter. We then show how one could implement this scheme iteratively. Finally, we shall take advantage of the local statistical tests designed in the previous section to suggest a better way to improve the image quality.

### 4.1 Denoising the Residual

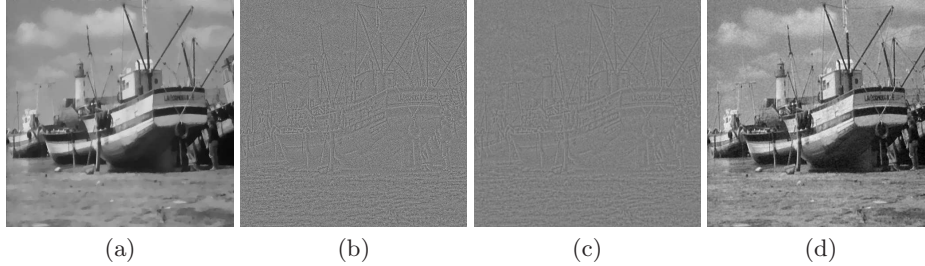
We apply a different denoiser on the residual image in attempt to separate the remaining structure from the dominating noise. In section 4.3 we will show a non-traditional way to perform this task based on the particular nature of the residual. The justification of adding back this “denoised” residual to the denoised image comes from the fact that the cleaned residual contains the structure of the original image removed by the denoiser. Mathematically, the procedure may be expressed as follows,

$$\mathcal{E}(\mathbf{d}) = \mathbf{d} + \mathcal{D}_2(\mathbf{r}) = \mathcal{D}_1(\mathbf{y}) + \mathcal{D}_2(\mathbf{y} - \mathcal{D}_1(\mathbf{y})), \quad (26)$$

where  $\mathcal{D}_1$  and  $\mathcal{D}_2$  are two denoising algorithms.

We consider image enhancement to be achieved if the image quality obtained by adding the denoised residual is greater than the image quality obtained by simply denoising the image with either of the two methods. Here we show an example where Eq. 26 is used successively. Gaussian white noise of standard deviation  $\sigma = 15$  was first added to create a noisy image. This image was then

denoised with the Total Variation Minimization algorithm ( $\mathcal{D}_1$ ). The residual image was then denoised with the Adaptive Wiener Filter ( $\mathcal{D}_2$ ). The results are shown in Fig. 4. The PSNR and the SSIM indices for  $\mathcal{E}(\mathbf{d})$  are greater than the quality measurements obtained by processing the image with either  $\mathcal{D}_1(\mathbf{y})$  or  $\mathcal{D}_2(\mathbf{y})$  alone.



**Fig. 4.** (a): *Boat* denoised by Total Variation Minimization (PSNR = 28.6, SSIM = 0.75). (b): Residual image. (c): Residual “denoised” by Adaptive Wiener Filter. (d): Enhanced denoised image (PSNR = 29.8, SSIM = 0.77). The images were cropped between (100, 100) and (400, 400).

In general, simply taking a traditional denoising algorithm to denoise the residual and then adding back the result to the denoised image does not guarantee an improvement in the quality of the denoised image. The point, however, is that with a good no-reference quality measure (cf. Section 3), we can at least perform this step and then verify whether or not an improved image is obtained.

## 4.2 Iterative Schemes

We now present an iterative scheme of the above denoising procedure. Let  $\mathcal{Q}$  denote a no-reference quality measure. Then the larger that  $\mathcal{Q}(\mathbf{d})$  is, the higher the quality of  $\mathbf{d}$ . We iterate the algorithm  $J$  times and then choose the image with the best quality. As such, the algorithm will always produce an image of at least the same quality as that of the denoised image.

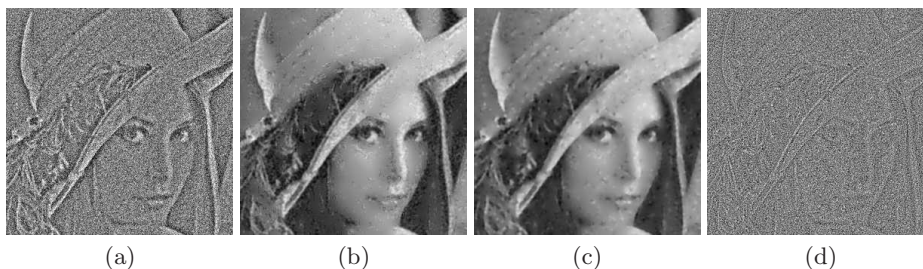
1. Set  $j \leftarrow 1$ ,  $\tilde{\mathbf{d}}_{(0)} \leftarrow \mathbf{y}$  and  $\mathbf{d}_{(0)} \leftarrow 0$ ;
2. Denoise the image:  $\mathbf{d}_{(j)} \leftarrow \mathcal{D}_1(\tilde{\mathbf{d}}_{(j)})$ ;
3. Compute the residual:  $\mathbf{r}_{(j)} \leftarrow \mathbf{y} - \mathbf{d}_{(j)}$ ;
4. Denoise the residual:  $\tilde{\mathbf{r}}_{(j)} \leftarrow \mathcal{D}_2(\mathbf{r}_{(j)})$ ;
5. Add it back to the denoised image:  $\tilde{\mathbf{d}}_{(j)} \leftarrow \mathbf{d}_{(j)} + \tilde{\mathbf{r}}_{(j)}$ ;
6. While  $j < J$ , increment  $j \leftarrow j + 1$  and go to step 2.
7. Find  $j$  that maximizes  $\mathcal{Q}(\mathbf{d}_{(j)})$  or  $\mathcal{Q}(\tilde{\mathbf{d}}_{(j)})$ ;
8. If  $\mathcal{Q}(\mathbf{d}_{(j)}) > \mathcal{Q}(\tilde{\mathbf{d}}_{(j)})$  return  $\mathbf{d}_{(j)}$ , else return  $\tilde{\mathbf{d}}_{(j)}$ .

In the next section, we employ this algorithm, combined with a new way to denoise the residual.

### 4.3 Using Statistical Tests to Improve the Denoising

Denoising could be further improved by using a statistical test to determine which parts of the residual need to be added back to the denoised image. If one of the hypothesis tests is rejected, it is likely that the residual is not Gaussian white noise, hence we may consider that the denoising was insufficient. In this case we perform an enhancement such as Eq. 26.

The experiment is as follows. A Gaussian filter with large amplitude was applied to the noisy *Lena* image ( $\sigma = 35$ ) of Fig. 1. A Pearson’s coefficient correlation test and a Kolmogorov-Smirnov test were then performed on the residual. The residual was denoised with an adaptive Wiener filter. Only the portions of the cleaned residual where one of the hypothesis tests was rejected were added back to the denoised image. This procedure was then iterated with decreasing amplitudes of the Gaussian filter. The idea is to first smoothen the flat regions of the image, and then to work on the details. This iterative scheme produced gains in both PSNR and SSIM. Visually, as seen in Fig. 5, the results are good, although a salt-and-pepper-like noise appears in some regions of the image. This comes from the dichotomic nature of hypothesis tests: a pixel is considered as either noisy or clean.



**Fig. 5.** (a)  $\mathbf{r}_{(1)}$ : Residual image of  $\mathbf{d}_{(1)}$ . (b)  $\tilde{\mathbf{d}}_{(1)}$ : First iteration of the denoised image enhancing algorithm (PSNR = 25.9, SSIM = 0.63). (c)  $\mathbf{d}_{(7)}$ : Best image obtained by the algorithm (PSNR = 26.3, SSIM = 0.70). (d)  $\mathbf{r}_{(7)}$ : Residual image of  $\mathbf{d}_{(7)}$ .

## 5 Concluding remarks

We have demonstrated the potential of using image residuals in denoising applications. There are several interesting avenues for future research. For example, one could try to improve the no-reference estimates of the PSNR and of the SSIM index. In addition, there are two very useful applications of these no-reference measures: (1) to perform image quality assessment; (2) to design better denoising algorithms by optimizing the denoising parameters *a posteriori* or by directly solving the problem of finding  $\mathbf{d}$  that maximizes  $\mathcal{Q}(\mathbf{d})$ . Finally, the use of the

residual and statistical tests to improve denoising algorithms requires much more work. The goal is to design a general method for enhancing *any* given denoiser.

**Acknowledgements.** This research was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) in the form of a Postgraduate Scholarship (DB), Discovery Grants (ERV, ZW) and a Strategic Grant (ZW), and in part by Ontario Ministry of Research & Innovation in the form of a Early Researcher Award (ZW), which are gratefully acknowledged.

## References

1. A. Buades, B. Coll, and J.-M. Morel. A review of image denoising algorithms, with a new one. *Multiscale Modeling & Simulation*, 4(2):490–530, 2005.
2. E. Candès, L. Demanet, and L. Ying. New tight frames of curvelets and optimal representations of objects with piecewise  $C^2$  singularities. *Comm. Pure Appl. Math.*, 57(2):219–266, 2004.
3. K.-S. Chuang and H.K. Huang. Assessment of noise in a digital image using the join-count statistic and the moran test. *Phys. Med. Bid.*, 37(2):357–369, 1992.
4. R. R. Coifman and D. L. Donoho. *Wavelets and Statistics*, chapter Translation-Invariant Denoising, pages 125–150. Springer-Verlag, New York, NY, 1995.
5. K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3D transform-domain collaborative filtering. *IEEE Trans. Image Process.*, 16(8):2080–2095, 2007.
6. D.L. Donoho and I.M. Johnstone. Ideal spatial adaption via wavelet shrinkage. *Biometrika*, 81:425–455, 1994.
7. F.-C. Jeng and J. W. Woods. Inhomogeneous gaussian image models for estimation and restoration. *IEEE Trans. Acoust., Speech, Signal Proc.*, 36(8):1305–1312, 1988.
8. J.-S. Lee. Digital image enhancement and noise filtering by use of local statistics. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2(2):165–168, 1980.
9. F. Malgouyres. A noise selection approach of image restoration. In *SPIE, International Conference on Wavelets IX*, volume 4478, pages 34–41, 2001.
10. F. Murtagh, J.-L. Starck, and M. Louys. Very-high-quality image compression based on noise modeling. *Int. J. Imag. Syst. Tech.*, 9(11):38–45, 1998.
11. S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. An iterative regularization method for total variation-based image restoration. *SIAM Multiscale Model. and Simu.*, 4:460–489, 2005.
12. P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.*, 12(7):629–639, 1990.
13. J. Portilla, V. Strela, M. Wainwright, and E.P. Simoncelli. Image denoising using scale mixtures of gaussians in the wavelet domain. *IEEE Trans. on Image Proc.*, 12(11):1338–1351, 2003.
14. R. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. D*, 60(1-4):259–268, 1992.
15. Q. Shan, J. Jia, and A. Agarwala. High-quality motion deblurring from a single image. *ACM Transactions on Graphics (SIGGRAPH)*, 27(3):1–10, 2008.
16. Z. Wang, A.C. Bovik, H.R. Skeikh, and E.P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. on Image Proc.*, 13(4):600–612, 2004.