Proc. of the IEEE Int. Conf. on Image Processing (ICIP), vol. 1, pp. 14-17, Sep. 2003.

EFFICIENT REMOVING OF IMPULSIVE NOISE BASED ON AN ℓ_1-ℓ_2 COST-FUNCTION

Mila NIKOLOVA

CNRS URA820–ENST Dpt. TSI, ENST, 46 rue Barrault, 75013 Paris, France (nikolova@enst.fr)
CMLA UMR 8536–ENS de Cachan, 61 av. President Wilson, 94235 Cachan Cedex, France

ABSTRACT

We consider images corrupted by impulsive noise (outliers). A large variety of methods are based on order statistic filters. Other methods use state-conditioned filters. We propose a different approach, consisting of two stages. First, outliers are detected based on the minimizer of a cost-function composed of an ℓ_1 data-fidelity and an ℓ_2 regularization term. The computation of this minimizer is speed and the detection of outliers reliable. Then, only outliers are removed and replaced by the median of the nearest neighboring regular (uncorrupted) data samples. This method is justified by some recent theoretical results [13]. The numerical experiments show that our method is very efficient in a broad range of situations, including highly corrupted images.

1. INTRODUCTION

The unknown original image x^* and the data y are identified with vectors of R^p. We consider data y corrupted with impulsive noise. We call *outliers* all samples y_i for $i \in \Omega_y$ where $\Omega_y := \{i \in \{1,..,p\} : y_i \neq x_i^*\}$. Conversely, $\Omega_y^c = \{1,..,p\} \setminus \Omega_y$ addresses all *regular* (uncorrupted) data samples. It is naturally supposed that there are "enough" regular data samples and that outliers are quite *dissimilar* with respect to neighboring pixels. Since [1], many different order-statistic filters have been derived in order to process outliers [2, 3, 4, 5, 6, 7, 8], e.g. recursive median, hybrid median, center-weighted median (CWM), permutation weighted median (PWM). Being applied uniformly across the image, these filters tend to alter both corrupted and regular pixels. State-conditioned filters smooth pixels conditionally on a decision, taken over a small window, on whether or not it is an outlier [9, 10, 11]. Their success is closely dependent on the reliability of the outlier decision rule and of the window size.

Our approach is different. In [12, 13], we consider cost-functions composed of a non-smooth data-fidelity term and a smooth regularization term, and show that their minimizers fit exactly regular (uncorrupted) data entries and smooth outliers. We observed that outliers are reliably detected, even in highly corrupted images, but that their smoothing is not optimal. In this paper, we separate these tasks. For the detection of outliers we use an ℓ_1-ℓ_2 cost-function—see (2)—whose minimization is very easy (§4). Then detected outliers are smoothed using conventional median-based techniques. The numerical results are extremely encouraging.

2. THE PROPOSED METHOD

For every pixel $i \in \{1,\dots,p\}$, the symbol N_i denotes the set of the four, or the eight pixels adjacent to i. Recall that $i \cap N_i = \emptyset$ and that $j \in N_i \Leftrightarrow i \in N_j$. By extension, if ω is a connected subset on the grid of the image, its adjacent neighborhood N_ω reads

$$N_\omega = \left(\bigcup_{i \in \omega} N_i \right) \setminus \omega.$$

If $\omega = \{i\}$, we find $N_\omega = N_i$. The sought-after denoised image will be denoted $\hat{x}$. The proposed method is summarized below.

Step 1: Detection of outliers. The estimate $\hat{\Omega}_y$ of the locations of outliers in data y is defined as

$$\hat{\Omega}_y = \{i \in \{1,\dots,p\} : \tilde{x}_i \neq y_i\}, \tag{1}$$

where $\tilde{x}$ minimizes the cost-function $\mathcal{F}_y : \mathrm{R}^p \to \mathrm{R}$

$$\mathcal{F}_y(x) = \sum_{i=1}^{p} |x_i - y_i| + \frac{\alpha}{2} \sum_{i=1}^{p} \sum_{j \in N_i} (x_i - x_j)^2, \tag{2}$$

where $\alpha > 0$ is a fixed parameter. We then put

$$\hat{x}_i = y_i, \forall i \in \hat{\Omega}_y^c,$$

since $\hat{\Omega}_y^c$ is the estimated set of regular data points.

Step 2: Smoothing of outliers. Let us represent $\hat{\Omega}_y$ as a union of connected components, say ω_ℓ, for $\ell = 1,\dots,n$. We consider successively all ω_ℓ, for $\ell = 1,\dots,n$. If ω_ℓ is a singleton, say $\omega_\ell = \{i\}$, we take

$$\hat{x}_i = \text{median}(\{y_j : j \in N_i\}).$$

In practice, many ω_ℓs are singletons. Otherwise, for every $i \in \omega_\ell$, we determine $\tilde{N}_i \subset N_{\omega_\ell}$ as the subset of the m closest samples to i (e.g., $m = 4$) and calculate $\hat{x}_i = \text{median}(\{y_j : j \in \tilde{N}_i\})$. Thus all $\hat{x}_i$ are calculated based only on regular data samples.

3. THEORETICAL JUSTIFICATION

The statements given below follow from some more general results developed in [13]. For any $x \in \mathrm{R}^p$, and for any $i \in \{1,\dots,p\}$, let $\chi_i(x)$ be the mean of the neighbors of x_i:

$$\chi_i(x) = \frac{1}{\#N_i} \sum_{j \in N_i} x_j.$$

Let us emphasize that $\chi_i(x)$ does not depend on x_i.

Proposition 1 *For $y \in \mathrm{R}^p$, the function $\mathcal{F}_y : \mathrm{R}^p \to \mathrm{R}$ reaches its minimum at $\tilde{x} \in \mathrm{R}^p$ if and only if*

$$|\tilde{x}_i - \chi_i(\tilde{x})| \leq \frac{1}{2\alpha \# N_i}, \qquad \forall i \in \hat{\Omega}_y^c, \tag{3}$$

$$\tilde{x}_i - \chi_i(\tilde{x}) = \frac{\text{sign}(y_i - \chi_i(\tilde{x}))}{2\alpha \# N_i}, \qquad \forall i \in \hat{\Omega}_y, \tag{4}$$

where $\hat{\Omega}_y$ is defined as in (1).

Notice that in (3)we have $\tilde{x}_i = y_i$. By (3)-(4), if $|y_i - \chi_i(\tilde{x})| > 1/(2\alpha\#N_i)$, then y_i is estimated to be an outlier. So, the detection of outliers is based on the (implicit) comparison of each y_i with the relevant $\chi_i(\tilde{x})$. Let M_i be an extended neighborhood of i, say $M_i = N_i \cup \left(\bigcup_{j\in N_i} N_j\right) \setminus \{i\}$. Below we show that various subsets $\hat{\Omega}_y$ can arise as long as y varies.

Proposition 2 *Choose $\hat{\Omega}$ so that $M_i \in \hat{\Omega}^c$ for every $i \in \hat{\Omega}$. Then there is an open set $Y_{\hat{\Omega}} \subset \mathrm{R}^p$ such that for every $y \in Y_{\hat{\Omega}}$, the function $\mathcal{F}_y$ reaches its minimum at a point $\tilde{x}$ such that $\tilde{x}_i = y_i$, for all $i \in \hat{\Omega}^c$, whereas $\tilde{x}_i \neq y_i$ is given by (4) for all $i \in \hat{\Omega}$.*

For instance, suppose that $\tilde{x}$ is such that for some i we have $M_i \in \hat{\Omega}_y^c$. Based on (3)-(4), it is not difficult to see that

$$\begin{array}{lcll} |y_i - \chi_i(y)| \leq \dfrac{1}{2\alpha\#N_i} & \Leftrightarrow & \tilde{x}_i = y_i & \\ |y_i - \chi_i(y)| > \dfrac{1}{2\alpha\#N_i} & \Leftrightarrow & \tilde{x}_i = \chi_i(y) - \dfrac{\sigma_i}{2\alpha\#N_i}, & \end{array} \tag{5}$$

where $\sigma_i = \mathrm{sign}(y_i - \chi_i(y))$. In the second case, y_i is an outlier, because it is "too far" from the mean of its neighbors. When the image is highly corrupted, many outliers are adjacent neighbors; then $\hat{\Omega}_y$ is composed of connected components of different size. Let ω be a connected component of $\hat{\Omega}_y$. It can be shown that $\tilde{x}_i$, for $i \in \omega$, is calculated as an *increasing, affine combination of all neighboring regular samples*, y_j, for $j \in N_\omega \subset \hat{\Omega}_y^c$.

Next we focus on the stability of the detection of outliers proposed in step 1. For $i \in \hat{\Omega}_y$, put $\sigma_i = \mathrm{sign}(y_i - \chi_i(\tilde{x}))$. Based on (4), it can be shown that for all $y' \in \mathrm{R}^p$ such that

$$\begin{array}{rcll} y'_i & = & y_i & \forall i \in \hat{\Omega}_y^c, \\ \sigma_i y'_i & \geq & \sigma_i y_i & \forall i \in \hat{\Omega}_y, \end{array}$$

the relevant $\mathcal{F}_{y'}$ reaches its minimum at the same point $\tilde{x}$. This shows that if y_i is outlier, then $\tilde{x}_i$ is *independent* of the exact value of y_i, hence $\tilde{x}$ is insensitive to the magnitude of aberrant data.

Proposition 3 *For $y \in \mathrm{R}^p$ given, let $\mathcal{F}_y$ reach its minimum at $\tilde{x}$. Suppose that $\hat{\Omega}_y^c$, defined according to (1), is nonempty and that for every $i \in \hat{\Omega}_y^c$, the inequality in (3) is strict. Then there is an open subset $Y_{\hat{\Omega}} \subset \mathrm{R}^p$, with $y \in Y_{\hat{\Omega}}$, such that for every $y' \in Y_{\hat{\Omega}}$, the relevant $\mathcal{F}_{y'}$ reaches its minimum at an $\tilde{x}'$ such that*

$$\begin{array}{ll} \tilde{x}'_i = y'_i & \text{if} \quad i \in \hat{\Omega}_y^c, \\ \tilde{x}'_i \neq y'_i & \text{if} \quad i \in \hat{\Omega}_y. \end{array}$$

The estimation of outliers proposed in step 1 is *stable* since $\hat{\Omega}_y$ remains unchanged under small perturbations of the data. Reciprocally, the sets $Y_{\hat{\Omega}}$ in Propositions 2 and 3 being open, noisy data y do come across them and yield minimizers which fit *exactly* a certain number of the data entries, i.e. that $\hat{\Omega}_y \neq \emptyset$.

By (3)-(4) and (5) we see that decreasing α increases the threshold for the detection of outliers, and vice-versa. For images with complex features, detecting all outliers may require α to be pretty large. Then some regular data entries (especially those placed near the edges) may be falsely detected as outliers. This will not considerably decrease the quality of the restoration since at step 2 of our method, these entries are replaced by the median of their neighbors, which is a good edge-preserving operation. Practically, the method is quite stable with respect to the value of α.

4. MINIMIZATION ALGORITHM

For every $k = 1, 2, \ldots$, the iterate $x^{(k)}$ is obtained from $x^{(k-1)}$ by calculating successively all its components $x_i^{(k)}$, for $i = 1, \ldots, p$ (taken in any order), according to the scheme:

$$\begin{array}{rcl} & & \mathcal{F}_y(x_1^{(k)}, x_2^{(k)}, \ldots, x_{i-1}^{(k)}, x_i^{(k)}, x_{i+1}^{(k-1)}, \ldots, x_p^{(k-1)}) \\ & = & \min_{t\in\mathrm{R}} \mathcal{F}_y(x_1^{(k)}, x_2^{(k)}, \ldots, x_{i-1}^{(k)}, t, x_{i+1}^{(k-1)}, \ldots, x_p^{(k-1)}). \end{array}$$

Proposition 4 *The sequence $x^{(k)}$ converges, as $k \to \infty$, to a point $\tilde{x}$ such that $\mathcal{F}_y(\tilde{x}) \leq \mathcal{F}_y(x)$, for all $x \in \mathrm{R}^p$.*

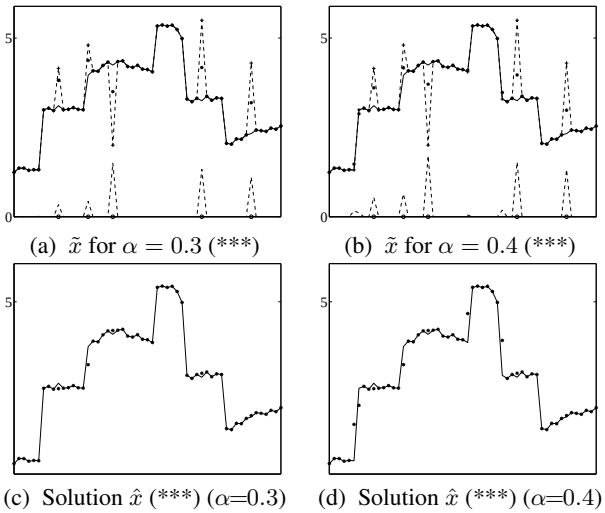
This result is obtained by applying Theorem 2 of [13]. Let the intermediate solution at step $i-1$ of iteration k be denoted $x^{(k,i-1)} = (x_1^{(k)}, x_2^{(k)}, \ldots, x_{i-1}^{(k)}, x_i^{(k-1)}, x_{i+1}^{(k-1)}, \ldots, x_p^{(k-1)})$. By Proposition 1, at the next step we calculate $x_i^{(k)}$ according to the rule:

$$\begin{array}{rcl} \xi_i^{(k)} & = & y_i - \chi_i^{(k)} \ \text{ where } \ \chi_i^{(k)} = \dfrac{1}{\#N_i} \displaystyle\sum_{j\in N_i} x_j^{(k,i-1)}, \\ \text{if } \left|\xi_i^{(k)}\right| & \leq & \dfrac{1}{2\alpha\#N_i} \ \Rightarrow \ x_i^{(k)} = y_i, \\ \text{if } \left|\xi_i^{(k)}\right| & > & \dfrac{1}{2\alpha\#N_i} \ \Rightarrow \ x_i^{(k)} = \chi_i^{(k)} + \dfrac{\mathrm{sign}(\xi_i^{(k)})}{2\alpha\#N_i}. \end{array}$$

Notice that updating each entry $x_i^{(k)}$ involves only the samples belonging to its neighborhood N_i. So, at one step we can update *simultaneously* any subset of entries $\{i_1, \ldots, i_K\} \subset \{1, \ldots, p\}$ provided that $i_j \cap N_{i_k} = \emptyset$ for all $j, k \in \{1, \ldots, K\}$ with $j \neq k$. Based on Proposition 3, we can expect that $\tilde{x}_i = y_i$ for a certain number of the entries of $\tilde{x}$. This suggests we initialize the algorithm with $x^{(0)} = y$. Because of the simplicity of the calculations, the algorithm is fast. The calculation time is of the order of the median filtering (quite faster if there is a few outliers).

5. EXPERIMENTS

5.1. Illustration of the method.



(a) $\tilde{x}$ for $\alpha = 0.3$ (***) (b) $\tilde{x}$ for $\alpha = 0.4$ (***)

(c) Solution $\hat{x}$ (***) (α=0.3) (d) Solution $\hat{x}$ (***) (α=0.4)

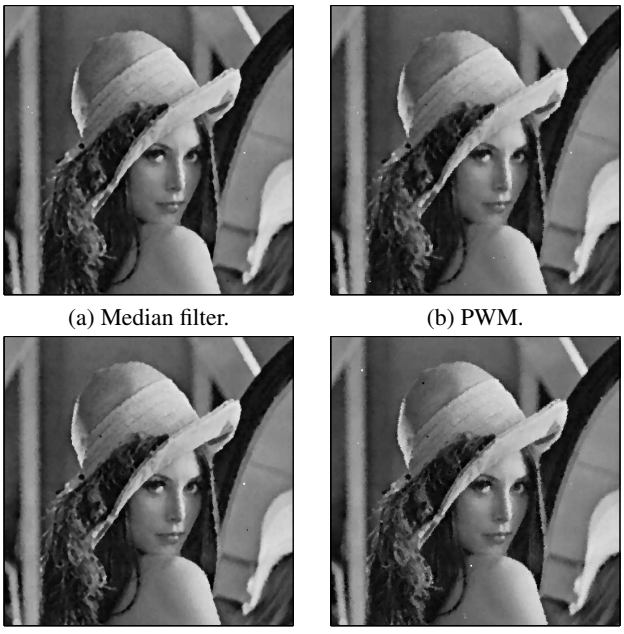
Fig. 1. Original (—), Data (- -), Outliers in y (o), Residual $|\tilde{x} - y|$ (- -).

We consider the denoising of the original signal x^* plotted with the solid line in Fig. 1 from the data y plotted with the dashed line in (a) and (b), which involve 5 outliers. For $\alpha = 0.3$, the minimizer $\tilde{x}$ of $\mathcal{F}_y$, plotted in (a) with "$*$", leads to $\hat{\Omega}_y = \Omega_y$, i.e. to the true set of outliers in y. Applying step 2 of our method yields the solution $\hat{x}$ plotted in (c) with "$*$". A larger value for α, as in (b), leads to an $\hat{\Omega}_y$ which contains Ω_y plus some falsely detected outliers. The resultant solution $\hat{x}$ is shown in (d).

5.2. Denoising of an image

The data in Fig. 2(right) contain 50 % salt-and-pepper (S-P) noise, uniformly distributed over the grid of the image.

Fig. 2. Original image and data with 50% S-P noise.



(a) Median filter. (b) PWM.

(c) CWM. (d) Recursive CWM.

Fig. 3. Median-based methods

In Fig. 3, different median-based techniques are applied by using the set of parameters (window size, number of iterations, recursivity, additional parameters) which lead to the best image denoising. In all cases, we have a $5\mathbf{x}5$ window. The image in (a) corresponds to 2 iterations of median filter. The solution in (b) corresponds to PWM with rank parameters 7 and 19. The image in (c) is obtained using 3 iterations of CWM with multiplicity parameter 5. The next (d) results from 5 iterations of recursive CWM with parameter 10. All these restorations are slightly blurred, especially the hairs and the feathers, the tissue of the hat lose detail, and there are defaults near the nose. Moreover, they still contain several outliers with high intensity.

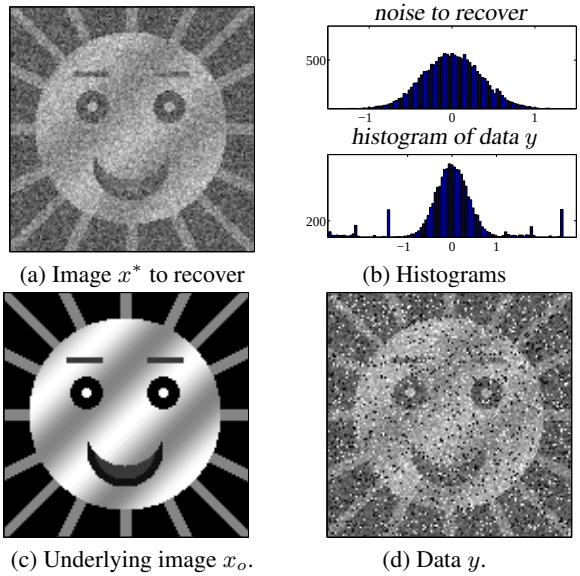
(a) The minimizer $\tilde{x}$ of $\mathcal{F}_y$ (b) Solution $\hat{x}$.

Fig. 4. The proposed method.

The image in Fig. 4 (a) is $\tilde{x}$, the minimizer of $\mathcal{F}_y$ for $\alpha = 0.0075$. For 45 % of its entries, $\tilde{x}_i = y_i$. The other entries correspond to important outliers. The image in Fig. 4 (b) corresponds to step 2 of our method. In is quite clean, edges are neat and the challenging features—hair, feathers, hat—are well recovered. This restoration is pretty satisfactory.

5.3. Preprocessing of noisy data

In different applications, data y result from outlier-free degradations (e.g., distortion, blurring, quantization errors, electronic noise), plus impulsive noise. A preliminary processing, aimed at removing the outliers, is usually needed before to apply reconstruction methods. Preprocessing is often realized using median-based filters. It is critical that preprocessing keeps intact all the information contained in the outlier-free data.

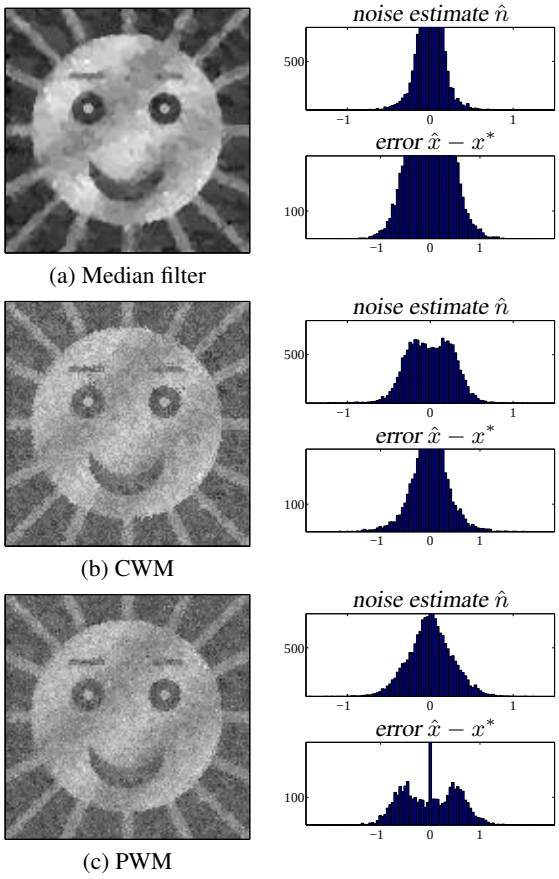


(a) Image x^* to recover (b) Histograms

(c) Underlying image x_o. (d) Data y.

Fig. 6. Data with two-stage degradation.

In our experiment, the sought image x^*, shown in Fig. 6(a), is related to x_o, shown in (c), by $x^* = x_o + n$, where n is white Gaussian noise with 20 dB SNR. The histogram of n is plotted in (b), up. Our goal is to restore x^* based on the data y, shown in (d), which contain 10 % S-P noise. Restoring x^* is a challenge since the white noise there must be preserved. The relevance of an estimate $\hat{x}$ is evaluated by both, the error $\hat{x} - x^*$ and the closeness of the histogram of the estimated noise, $\hat{n} = \hat{x} - x_o$, to the initial noise n. All results in Fig. 7 correspond to the parameters which lead to the best denoising. The image in (a) corresponds to one iteration of median filter over a $3\mathbf{x}3$ window. The image in (b) is calculated using CWM with a $5\mathbf{x}5$ window and multiplicity parameter 14. The image in (d) corresponds to one iteration of PWM, for a $5\mathbf{x}5$ window and rank parameters 4 and 22. In all these estimates, the distribution of the noise estimate $\hat{n}$ (shown on the right, up) is quite different from the distribution of n. Fig. 8 displays the issue of the proposed method, for $\alpha = 1.3$. It achieves a good preservation of the statistics of the noise in x^*, as seen from the histogram of the estimated noise $\hat{n}$—Fig. 8, right, up. Moreover, the error $\hat{x} - x^*$ is essentially concentrated around zero, as seen in Fig. 8, right, down.



(a) Median filter

(b) CWM

(c) PWM

Fig. 7. Median-based restorations.

6. CONCLUSIONS

We present a new framework for the removal of impulsive noise. An important advantage of our method is that the outlier decision is based on a global cost-function, taking into account a prior on the image. Hence the reliability of the outlier detection. The numerical results are promising. Further improvement of step 2 can be envisaged in order to take into account local edge information.

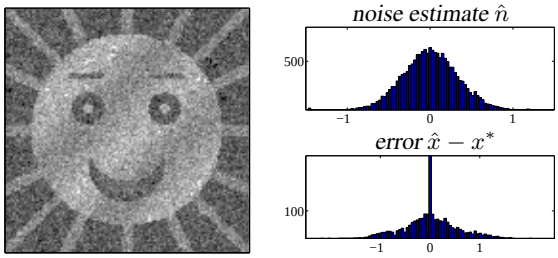


Fig. 8. The proposed method.

Method	$\|\hat{x}-x^*\|_2$	$\|\mathrm{hist}(\hat{n}-n)\|_2$
Median filter	45.79	$1.95\mathbf{x}10^3$
CWM	30.96	$1.08\mathbf{x}10^3$
PWM	34.70	$0.86\mathbf{x}10^3$
Proposed method	25.59	$0.43\mathbf{x}10^3$

Fig. 9. Errors: data and noise distribution.

7. REFERENCES

[1] J. W. Tukey, "Nonlinear (nonsuperimposable) methods for smoothing data," in *Conf. Rec.* Eacson, 1974.

[2] Sung-Jea Ko and Yong Hoon Lee, "Adaptive center weighted median filter," *IEEE Transactions on Circuits and Systems*, vol. 38, no. 9, pp. 984–993, Sep. 1998.

[3] T Sun and Y. Neuvo, "Detail-preserving based filters in image processing," *Pattern-Recognition Letters*, vol. 15, pp. 341–347, 1994.

[4] G. R. Arce, T. A. Hall, and K. E. Barner, "Permutation weighted order statistic filters," *IEEE Transactions on Image Processing*, vol. 4, no. 8, pp. 1070–1083, Aug. 1995.

[5] L. Yin, R. Yang, M. Gabbouj, and Y Neuvo, "Weighted median filters: a tutorial," *IEEE Transactions on Circuit Theory*, vol. 41, pp. 157–192, May 1996.

[6] John Russ, *The Image Processing Handbook*, CRC Press, Berlin, 3 edition, 1998.

[7] Alan C. Bovik, *Handbook of image and video processing*, Academic Press, New York, 2000.

[8] Windyga, "Fast impulsive noise removal," *IEEE Transactions on Image Processing*, vol. 10, no. 1, pp. 173–179, 2001.

[9] W. K. Pratt, *Digital Image Processing*, Wiley, New York, 1978.

[10] E. Abreu, M. Lightstone, S. K. Mitra, and K. Arakawa, "A new efficient approach for the removal of impulse noise from highly corrupted images," *IEEE Transactions on Image Processing*, vol. 5, no. 6, pp. 1012–1025, June 1996.

[11] A. Flaig, G. R. Arce, and Ken Barner, "Affine order statistics filters: "medianization" of linear fir filters," *IEEE Transactions on Signal Processing*, 1998.

[12] Mila Nikolova, "Minimizers of cost-functions involving nonsmooth data-fidelity terms. Application to the processing of outliers," *SIAM Journal on Numerical Analysis*, vol. 40, no. 3, pp. 965–994, 2001.

[13] Mila Nikolova, "Image and signal denoising via minimization of cost-functions with a non-smooth data-fidelity term," to appear in Journal of Mathematical Imaging and Vision.