

Robust Filtering of NMR Images

Petr Hotmar and Jaromír Kukal

Prague Institute of Chemical Technology, Faculty of Chemical Engineering,
Department of Computing and Control Engineering

INTRODUCTION

The development of nonlinear median-based filters in recent years has resulted in remarkable results and has highlighted some new promising research avenues. On account of its simplicity, edge preservation property and robustness to impulsive noise, the standard median filter remains the most popular for image processing applications. The median filter, however, is often apt to remove fine details in the image, such as thin lines and corners. In recent years, a variety of median-type filters (e.g. relaxed median) have been developed to overcome this drawback.

Despite Robust Statistics – as the fundamental mathematical tool dealing with deviations from idealized assumptions in statistics – had been initiated in 60's and subsequently extensively exploited in 80's, particularly by Peter J. Huber and Frank R. Hampel, it has been sporadically applied effectively in image processing ever since. Although M-filters have been on several occasions proven useful as signal smoothers in impulsive noise environments, they have not yet been sufficiently developed to constitute a powerful alternative for the robust filtering problem. In this paper, we have tried to shed some light onto the issue of utilization of M-estimates in image processing and hopefully filled some gaps in the mentioned field.

1. ROBUST BACKGROUND

The following section is based on theoretical background presented in [4], [2] and [1].

1.1. Definition of M-estimator. M-estimators are generalizations of the usual maximum likelihood estimates. Classically θ is the parameter value maximizing the likelihood function², i.e. we have in obvious notation

$$(1.1) \quad L(\Theta) = \prod_{i=1}^n f(x_i, \Theta) = \max!$$

²The value Θ^* of the parameter Θ satisfying

$$\forall \Theta : L(\Theta^*) \geq L(\Theta)$$

is said to be *maximum likelihood estimate* of parameter Θ . The maximum likelihood principle states that the set of model parameters that maximizes the apparent probability of a set of observations is the best set possible.

where $f(x_i, \Theta)$ stands for the probability density of the distribution of a random variable X , or equivalently

$$(1.2) \quad -\ln L(\Theta) = -\sum_{i=1}^n \ln f(x_i, \Theta) = \min!$$

Any estimate T_n , defined by a minimum problem

$$(1.3) \quad \sum_{i=1}^n \rho(x_i; T_n) = \min!,$$

or by an implicit equation

$$(1.4) \quad \sum_{i=1}^n \psi(x_i; T_n) = 0,$$

where ρ is an arbitrary function,

$$(1.5) \quad \psi(x; \rho) = \frac{\partial \rho(x; \theta)}{\partial \theta},$$

is called *M-estimate*, or *Maximum Likelihood Estimate*.

In other words, the maximum likelihood estimate of θ for an assumed underlying family of densities $f(x, \theta)$ is functional derived from 1.3, thus a solution of

$$(1.6) \quad \int \psi(x, \theta) dF_n(x) = 0$$

with

$$(1.7) \quad \psi(x, \theta) = \frac{\partial \ln f(x, \theta)}{\partial \theta}.$$

We are particularly interested in location estimates

$$(1.8) \quad \sum_{i=1}^n \rho(x_i - T_n) = \min!$$

or

$$(1.9) \quad \sum_{i=1}^n \psi(x_i - T_n) = 0.$$

This last equation can be written equivalently as

$$(1.10) \quad \sum_{i=1}^n w_i \cdot (x_i - T_n) = 0$$

with

$$(1.11) \quad w_i = \frac{\psi(x_i - T_n)}{x_i - T_n};$$

this gives a formal representation of T_n as a weighted mean

$$(1.12) \quad T_n = \frac{\sum w_i x_i}{\sum w_i}$$

with weights depending on the sample.

Remark 1.1. If ψ is not monotone, the situation is much more complicated. For computational reasons, in order to narrow down the choice of solutions we take the solution nearest to the sample median: start an iterative root-finding procedure (eg. Newton's method) at the sample median and accept whatever root it converges to. This way, the procedure inherits some favourable properties from the median.

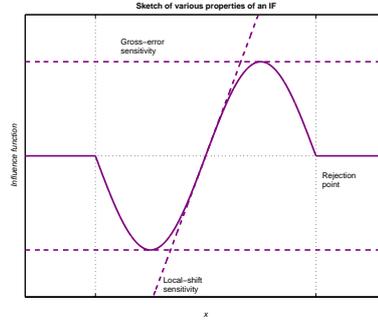


FIGURE 1. Sketch of various properties of an influence function

1.2. Robustness Measures. These measures basically try their best to answer the question of how greatly a small change in the underlying distribution changes the distribution of an estimate. Whereas the *breakdown point* is a global measure of the maximum fraction ϵ^* of arbitrary gross errors that an estimator can handle, the *influence function* measures the effect of infinitesimal perturbations on the estimator (formal definition comprises Gâteaux derivative and a Dirac function)

Let us now look at the important relation between the IF and the asymptotic variance. Making use of the first-order Taylor series expansion, the central limit theorem and the asymptotic normality, we conclude

$$(1.13) \quad V(T, F) = \int IF(x; T, F)^2 dF(x).$$

When estimating location in the model $\mathcal{X} = \mathfrak{R}$, $\Theta = \mathfrak{R}$, $F_\theta(x) = F(x - \theta)$ is seems natural to use ψ -functions of the type

$$(1.14) \quad \psi(x, \theta) = \psi(x - \theta).$$

Then at the model distribution F we obtain

$$(1.15) \quad IF(x; \psi, F) = \frac{\psi(x)}{\int \psi' dF}.$$

Hence the asymptotic variance at F can be calculated from (1.13), yielding

$$(1.16) \quad V(\psi, F) = \frac{\int \psi^2 dF}{(\int \psi' dF)^2}.$$

1.3. Other Robustness Measures Derived from the Influence Function.

- The *gross-error sensitivity* measures the worst possible influence on an estimator by an arbitrary infinitesimal contaminant

$$(1.17) \quad \gamma^* = \sup_x |IF(x; T, F)|.$$

- The *local-shift sensitivity* has to do with small fluctuations in the observations

$$(1.18) \quad \lambda^* = \sup_{x \neq y} \frac{|IF(y; T, F) - IF(x; T, F)|}{|y - x|}$$

- The *rejection point* allowing to reject extreme outliers entirely.

$$(1.19) \quad \rho^* = \inf \{r > 0; IF(x; T, F) = 0 \quad \text{when} \quad |x| > r\}$$

2. ROBUST ESTIMATORS BANK

2.1. Redescending M-estimators.

1. Hampel's three-part redescending M-estimator
2. Andrews' sine wave
3. Tukey's biweight function
4. Minimax hyperbolic tangent estimator

$$(2.1) \quad \psi(x) = \begin{cases} x & 0 \leq |x| \leq a \\ a \operatorname{sign}(x) & a \leq |x| \leq b \\ a \frac{r-|x|}{r-b} \operatorname{sign}(x) & b \leq |x| \leq r \\ 0 & r \leq |x| \end{cases}$$

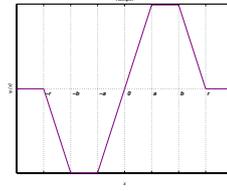


FIGURE 2. Hampel's M-estimator

$$(2.2) \quad \psi(x) = \begin{cases} \sin(x) & -\pi \leq x \leq \pi \\ 0 & \text{otherwise} \end{cases}$$

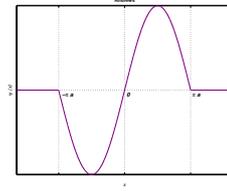


FIGURE 3. Andrews' M-estimator

$$(2.3) \quad \psi(x) = \begin{cases} x(1-x^2)^2 & |x| \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

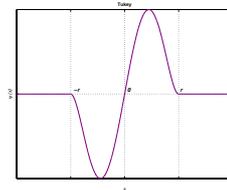


FIGURE 4. Tukey's M-estimator

$$(2.4) \quad \psi(x) = \begin{cases} x & 0 \leq x \leq a \\ b \tanh[\frac{1}{2}b(c-x)] & a \leq x \leq c \\ 0 & x \geq c \end{cases}$$

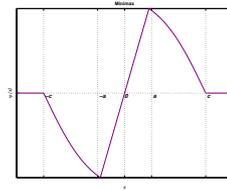


FIGURE 5. Minimax M-estimator

The redescending M-estimators have proven very successful in various studies, especially because they

- possess a γ^* which is rather low,
- possess a low local-shift sensitivity γ^* ,
- possess a finite rejection point ρ^* ,
- are much more efficient than the median.

2.2. Other M-estimators. The following estimators are also worth looking into in connection with image filtering.

- **Huber estimator** (see fig. 6), which was introduced by Huber as the solution of the minimax problem at $F = \Phi$,
- **Cauchy filter** is a flexible and efficient filter class for non-Gaussian impulsive environments that can appear in practice. They are defined as the MLE of location at the Cauchy distribution with dispersion γ , given by

$$f(x) = \frac{\gamma}{\pi} \frac{1}{\gamma^2 + x^2}$$

and thus generating an associated cost function

$$\rho(x) = \log[\gamma^2 + x^2].$$

In case of standard Cauchy distribution with $\gamma=1$, we can state that

$$(2.5) \quad \psi(x) = \begin{cases} x & |x| < b \\ b \operatorname{sign} x & |x| \geq b, \end{cases}$$

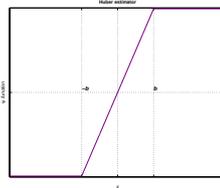


FIGURE 6. Huber estimator

$$(2.6) \quad \psi(x) = \frac{2x}{(1+x^2)}$$

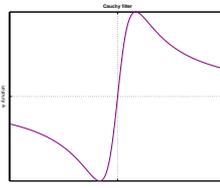


FIGURE 7. Cauchy filter

3. M-ESTIMATES AND IMAGE FILTERING

3.1. Computational challenge. Unfortunately, location M-estimators are usually not invariant with respect to scale, which is often a nuisance parameter. According to [4], this problem can be solved by defining T_n through

$$(3.1) \quad \sum_{i=1}^n \psi\left(\frac{x_i - T_n}{S_n}\right) = 0,$$

where S_n is a robust estimator of scale. It is advisable to determine S_n first, by means of the median absolute deviation as an auxiliary scale estimate⁴

$$(3.2) \quad S_n = 1.483 \operatorname{MAD}(x_i) = 1.483 \operatorname{med}_i\{|x_i - \operatorname{med}_j(x_j)|\}$$

which has maximal breakdown point $\epsilon^* = 50\%$. On the other hand, one could also compute T_n and S_n simultaneously as M-estimators of location and scale, as in Huber's proposal [4]. However, simulation has clearly shown the superiority of M-estimators with initial scale estimate given by the MAD, which is also much easier to use than the simultaneous version. Therefore, we recommend the use of initial MAD scaling for M-estimators, namely the following computational variant.

- *Modified Weights*

This approach apparently results from applying the plain iteration method on (1.9). Let

$$(3.3) \quad T^{(0)} = \operatorname{med}\{x_i\},$$

$$(3.4) \quad S^{(0)} = \operatorname{med}\{|x_i - T^{(0)}|\}.$$

$$(3.5)$$

Perform a few iterations of

$$(3.6) \quad T^{(m+1)} = \frac{\sum w_i^{(m)} x_i}{\sum w_i^{(m)}},$$

with

$$(3.7) \quad w_i^{(m)} = \frac{\psi[(x_i - T^{(m)})/S^{(0)}]}{(x_i - T^{(m)})/S^{(0)}}.$$

The iteration limit $T^{(\infty)}$ is unequivocally a solution of

$$(3.8) \quad \sum \psi\left(\frac{x_i - T}{S^{(0)}}\right) = 0.$$

⁴In most cases, it is convenient to standardize the estimates such that they are consistent at the ideal model distribution. In order to make MAD consistent at the normal distribution, we must multiply it by $\Phi(\frac{3}{4}) \cong 1.483$

3.2. Filtering windows. Several well-known two-dimensional fixed-sized filter masks (see fig. 8 and [3]) have been used in selecting image regions for the subsequent filter application. The borders of the image were coped with by duplicating the outmost values as many times as needed. Of the basic filter masks in Figure 8 the square mask is the least sensitive to image details. It filters out narrow lines and cuts corners of square-shaped objects. In image filtering the key factor is the trade-off between the amount of smoothing and the preservation of details.

It is easy to observe that the cross filter is able to preserve horizontal and vertical lines, whereas the X-shaped filter preserves only diagonal lines. For most applications the cross filter is preferred over the X-shaped filter since horizontal and vertical details are usually more important for a human observer than diagonal details.

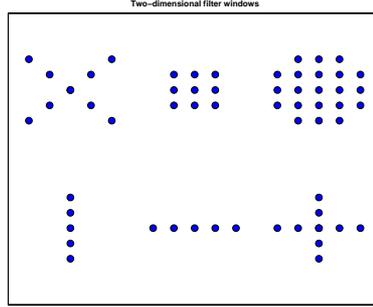


FIGURE 8. Filter masks: X-SHAPED, SQUARE, CIRCLE, VERTICAL, HORIZONTAL, CROSS

3.3. Preliminary experiments. In the experimental testing phase we processed a horizontal NMR slice (see Figure 9) conducted on the level of turbinate bone, outer auditory canal and brain stem. We tested four illustrative filters using circle mask with 2 pixels radius : Standard mean and median filter (to demonstrate the contrast between linear FIR filter and nonlinear, Cauchy filter (whose behavior can be tuned a simple parameter easily range from highly robust mode-type operation to the familiar average-type operation of the sample mean) robust filter) and finally Andrews' filter as a representative of redescending M-estimates.

Eventually, several quantitative measures were evaluated to epitomize both performance and efficiency of the filters, namely

1. Signal to Noise ratio

$$\text{SNR} = 10 \cdot \log \frac{\sum_{i,j} y_{ij}^2}{\sum_{i,j} (z_{ij} - y_{ij})^2}$$

2. Maximum error

$$\text{MAXE} = \max_{i,j} |x_{ij} - y_{ij}|,$$

3. Mean absolute error

$$\text{MAE} = \text{mean}_{i,j} |x_{ij} - y_{ij}|,$$

4. Sum squared error

$$\text{SSE} = \sum_{i,j} (x_{ij} - y_{ij})^2,$$

where (ij) stands for the coordinate location and x_{ij}, y_{ij} and z_{ij} are intensities of the original, denoised and corrupted image, respectively.

	SNR	MAXE	MAE	SSE
Mean	19.93	0.18	0.09	22.03
Median	23.77	0.09	0.01	1.14
Cauchy	20.23	0.18	0.09	20.88
Andrews	23.34	0.09	0.02	1.09

TABLE 1. Performance measures

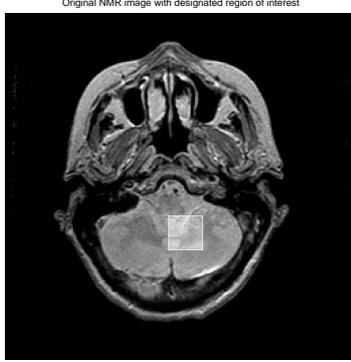


FIGURE 9. Original NMR image

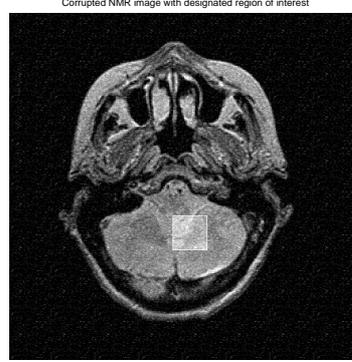


FIGURE 10. Corrupted NMR image

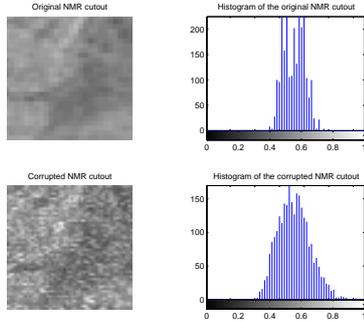


FIGURE 11. Image cutouts and corresponding histograms

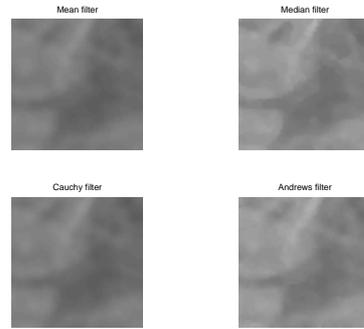


FIGURE 12. Filtering results

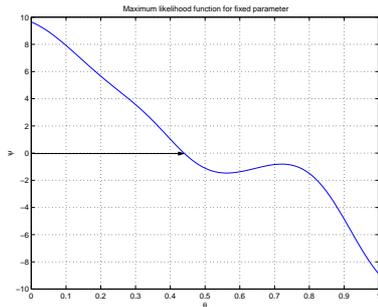


FIGURE 13. Location estimate

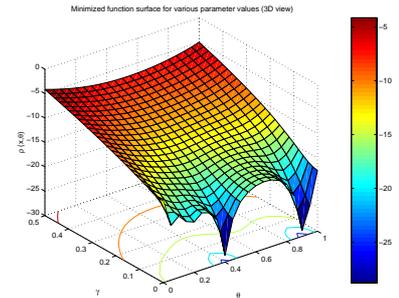


FIGURE 14. 3D view of minimized function

3.4. Demonstration of Numerical aspects. We studied the properties of the tuning parameter of Cauchy filter in such a way, that we used randomly selected pixel and applied horizontal mask with radius $r=2$. The maximum likelihood equation

$$\sum_i \frac{x_i - \theta}{\gamma^2 + (x_i - \theta)^2} = 0$$

was being solved by means of Newton iterative method until desired precision was attained to. At $\gamma=0$, the minimized function $\rho(X, \theta)$ has distinct minima at all the points included in filter mask. If γ is increased, the number of minima decreases.

	e	γ^*	κ^*	λ^*
Huber	0.9563	1.6749	3.6829	1.1889
Tukey	0.9100	1.6749	3.5528	1.4629
Andrews	0.9093	1.6749	3.5509	1.6749

TABLE 2. Comparison of Single-parameter Filters

3.5. Numerical analysis of single-parameter M-filters. In this part, we conducted the comparison of some redescending as well as monotone M-estimators. In order to make them comparable, their tuning constants were set up in such a way so that all the filters belong to Ψ_r with $r = 4$ and their gross-error sensitivities γ^* at Φ must be equal. Therefore, we decided to put $\gamma^* = 1.6749$, which is the value⁷ for the Tukey's biweight for $r = 4$. The estimators under study were the following :

- Andrews' Sine, a=1.142
- Tukey's Biweight, r=4
- Huber's Minimax, b=1.4088

Table 2 lists the asymptotic efficiencies (reciprocal value of asymptotic variance) of these estimators at normal model, as well as their gross-error, change-of-variance and local-shift sensitivities. It turns out that the three estimators exhibit similar behavior.

4. EXPERIMENTAL NOISING/DENOISING

In this key phase, we generated several types of noise coming from various distribution families, applied them to the NMR slice showing a brain tumor and finally cleared the noise out with our special-design filter based on hyperbolic functions.

4.1. Artificial Generation of Noise. In the beginning, the tumor close-up was corrupted with four different noise distributions, three of which being of additive nature and one multiplicative noise. The additive noises are distinguished by the following noise densities :

- Gaussian

$$(4.1) \quad f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2},$$

- Laplacian, as robust alternative to normal distribution

$$(4.2) \quad f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\sqrt{2}|x|/\sigma}$$

- and Cauchy, as yet another model for heavy-tailed noise

$$(4.3) \quad f(x) = \frac{1}{\sigma\pi \left(1 + \frac{x^2}{\sigma^2}\right)}$$

with σ as standard deviation for Gaussian and Laplacian noise, and as scale parameter for Cauchy noise.

4.2. Experimental Filter Design and Motivation. Our special-design, brand-new filter having been developed in accordance with robust statistics and validated in several experiments draws on Hampel's recommendations concerning M-estimates.

In [2], Hampel summarizes several observations regarding redescending estimators. He points out that two-part estimators as well as hyperbolic tangent estimators descend towards zero too steeply; therefore it seems advisable to investigate some approximations to combinations of these tanh estimators and two-part

⁷The value of γ^* for comparative purposes was suggested by [2].

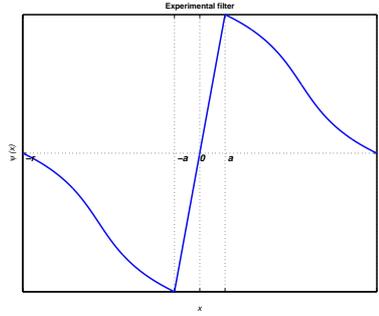


FIGURE 15. Experimental filter

descending estimators. Thus, we decided to retain the first linear part of two-part descending estimators, and replace the second one with function argument of hyperbolic sine, which comes from the family of hyperbolic functions – inverse functions to hyperbolic functions. So we took standard arsinh function, expressed in logarithmic form as

$$(4.4) \quad \operatorname{arsinh}(x) = \sinh^{-1}(x) = \ln(x + \sqrt{x^2 + 1}),$$

and shifted it $\frac{r-a}{2}$ units along x-axis and $\frac{a}{2}$ units along y-axis in order to link it to the first part of the estimator at the point $[a, a]$ and annul it at $[r, 0]$. Therefore, we multiply arsinh by constant k calculated from the equation

$$(4.5) \quad k \left(-\operatorname{arsinh} \left(\frac{r-a}{2} \right) \right) = -\frac{a}{2},$$

yielding

$$(4.6) \quad \forall x \in \langle a; r \rangle : k \left(-\operatorname{arsinh} \left(x - \frac{r-a}{2} - a \right) \right) + \frac{a}{2} = k \left(-\operatorname{arsinh} \left(x - \frac{r+a}{2} \right) \right) + \frac{a}{2},$$

as shown on figure 15. Likewise, we performed analogical operation in the left half-plane. Thus the filter is defined by

$$(4.7) \quad \psi(x) = \begin{cases} x & 0 \leq |x| \leq a \\ k \left(-\operatorname{arsinh} \left(x - \frac{r+a}{2} \right) \right) + \frac{a}{2} & a \leq |x| \leq r \\ 0 & \text{otherwise} \end{cases}$$

4.3. Testing Procedure. At the beginning of the testing phase we took noisy images generated in preceding noising procedures and engaged filtering programmes. In particular, we employed 2D binomial mask on sarcoma close-up 16 corrupted with Gaussian noise (Experiment no.1), Cauchy noise (Experiment no.2), Laplacian noise (Experiment no.3) and finally multiplicative noise (Experiment no.4), and performed filtration having used our experimental filter. Then, we applied wide spectrum of mask types on Laplacian-corrupted image, namely circle mask with 2 pixels radius (Experiment no.5), square mask with 1 pixel radius (Experiment no.6), X-shaped mask with 2 pixels radius (Experiment no.7), and eventually cross mask with 3 pixels radius (Experiment no.8). The resultant images 17 and 18 clearly exhibit rather successful noise elimination, thus justifying the use of our special-design filter. Note characteristic patterns in the individual filtered images (18, experiments no.4–8) leaking from corresponding mask shapes.

Remark 4.1. Filter parameters a and r were set to $a = 15, r = 150$ in case of additive noise, and $a = 10, r = 11$ in case of multiplicative speckle noise. These values were assigned heuristically after performing several testing experiments and visual/quantitative evaluation of the results, taking into account especially SNR .

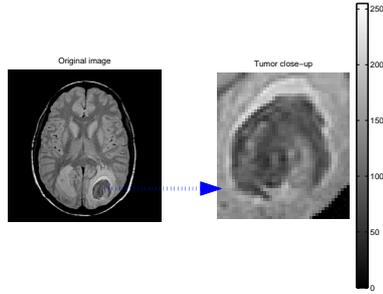


FIGURE 16. Sarcoma

	SNR	MAXE	MAE	SSE	MSE
Experiment no.1	24.6996	0.20432	0.039655	5.2046	0.0025216
Experiment no.2	16.3416	1.6173	0.053741	12.0105	0.005819
Experiment no.3	24.5328	0.21619	0.039109	5.2923	0.0025641
Experiment no.4	21.8824	1.1188	0.040783	6.9049	0.0033454
Experiment no.5	18.2131	0.34842	0.088493	23.3477	0.011312
Experiment no.6	24.012	0.22229	0.039974	5.5708	0.002699
Experiment no.7	20.2691	0.33154	0.048432	8.0968	0.0039228
Experiment no.8	19.71	0.27087	0.048278	8.5618	0.0041482

TABLE 3. Performance measures, testing experimental filter

Remark 4.2. Binomial masks flow from Pascal's triangle, i.e. the binomial expansion coefficients. Thus 2D binomial mask of 3×3 dimension is given by

$$(4.8) \quad \mathbf{B} = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$

. Not only did we select binomial masks because of their high reputation in image filtering (stemming mainly from their isotropic properties providing them with unique ability to considerably reduce noise environment), but above all to demonstrate incorporation of weighting framework into mask application.

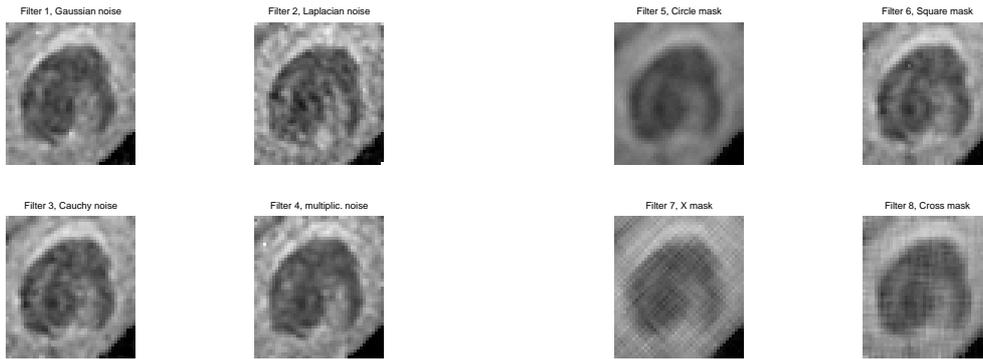


FIGURE 17. Filtered images, experiments 1-4

FIGURE 18. Filtered images, experiments 5-8

In order to have some quantitative appreciation of filtering results, we recalculated four measures of efficiency (see subsection 3.3 for definitions), added

- Mean squared error

$$\text{MSE} = \text{mean}_{i,j} |x_{ij} - y_{ij}|^2,$$

and summarized the results in table 3.

5. CONCLUSIONS

Experimental results (primarily signal to noise ratio) clearly demonstrate filter's high efficiency in Cauchy as well as Gaussian environment, whereas the Laplacian noise was not filtered out so effectively. One of the reasons for such behaviour lies in appropriate tuning of filter's parameters.

6. OPEN QUESTIONS AND FUTURE DIRECTIONS

The most auspicious and promising future work should be focused on making the comparison of the results with linear filters. In addition, more numerical aspects (continuity, limits, convergence etc.) of the filters in connection with their tuning parameters seem worth undergoing in-depth analysis and discussion. Parameters a and r of our special-design filter, as well as shapes and radiuses of the filtering windows, have obviously considerable impact on filtering results and therefore also deserve further scrutiny, including diligent research of parameter dependence on miscellaneous quantitative measures.

REFERENCES

1. Hamza A.B. and Krim H., *Image denoising: A nonlinear robust statistical approach*, IEEE Transactions on signal processing **49** (2001), no. 12.
2. Hampel F.R., *Robust statistics: The approach based on influence functions*, John Wiley, New York, 1986.
3. Kaiser J.F. Mitra S.K., *Handbook for digital signal processing*, John Wiley, 1993.
4. Huber P.J., *Robust statistics*, John Wiley, New York, 1981.

E-mail address: Petr.Hotmar@vscht.cz, Jaromir.Kukal@vscht.cz