NEW CLASS OF IMPULSIVE NOISE REDUCTION FILTERS BASED ON KERNEL DENSITY ESTIMATION

B. Smolka[1] *K.N. Plataniotis*[2] *R. Lukac*[3] *A.N. Venetsanopoulos*[4]

[1] Department of Automatic Control, Silesian University of Technology, Poland
[2] The Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, 10 King's College Road, Toronto ON, M5S 3G4, Canada
[3]Slovak Image Processing Center, Jarkova 343, 049 25 Dobsina, Slovak Republic
[4]Faculty of Applied Science and Engineering, University of Toronto, 35 St. George Street, Toronto, ON, M5S 3G4, Canada

ABSTRACT

This paper presents a new filtering scheme for the removal of impulsive noise in color images. It is based on estimating the probability density function for color pixels in a filter window by means of the kernel density estimation method. A quantitative comparison of the proposed filter with the vector median filter shows its excellent ability to reduce noise while simultaneously preserving fine image details.

1. INTRODUCTION

The reduction of noise in multichannel signal processing has been the subject of extensive research during the last years, primarily due to its importance to color image processing, [7]. In order to achieve optimal filtering results, the knowledge of the underlying statistical distribution of the signal and noise is needed. These distributions are often unknown and must be estimated from the data to prevent unrealistic assumptions that deteriorate the filter performance. If no information on the shape of the density distribution is known, non-parametric density estimation can be used, [1, 2]. The filter proposed in this paper is based on the non-parametric technique of *Parzen* or *Kernel Density Estimation* (KDE) [3], which is widely used in the field of pattern recognition and classification.

2. VECTOR MEDIAN FILTER

Let the mapping: $Z^l \rightarrow Z^q$ represents a multichannel image, where l is an image dimension and q characterizes a number of channels ($q = 3$ for color images). Let $W = \{\mathbf{x}_i \in Z^l; i = 1, 2, \ldots, n\}$ represents the samples in the filter window. Each input vector $\mathbf{x}_i$ can be associated with the cumulative distance measure D_i given by

$$D_i = \sum_{j=1}^{n} ||\mathbf{x}_i - \mathbf{x}_j|| \,, \quad i = 1, \ldots, n \,, \qquad (1)$$

where $\mathbf{x}_i = (x_{i_1}, \ldots, x_{i_q})$ and $\mathbf{x}_j = (x_{j_1}, \ldots, x_{j_q})$ characterize two q-dimensional vectors and $\|\cdot\|$ denotes a chosen vector norm. Since $D_1, D_2, ..., D_n$ are scalar values, their ordered set can be written simply as $D_1 \leq D_2 \leq \ldots \leq D_n$. If the same ordering is implied to the input set $\mathbf{x}_1, \ldots, \mathbf{x}_n$, the ordered input set is described as $\mathbf{x}_{(1)}, \mathbf{x}_{(2)}, \ldots, \mathbf{x}_{(n)}$ and the vector median (VMF) output is given by the sample $\mathbf{x}_{(1)}$ from the input set that minimizes the sum of vector distances with other vectors, [4].

3. KERNEL DENSITY ESTIMATION

Density Estimation describes the process of modelling the probability density function $f(x)$ of a given sequence of sample values drawn from an unknown density distribution.

The simplest form of density estimation is the histogram: sample space is first divided into a grid, then the density at the center of the grid cells is approximated by the number of sample values that fall into one bin divided by the width of one grid cell. The main disadvantage of the histogram is the strong dependence of the histogram's appearance on the chosen bin-width and the origin of the grid.

Kernel Density Estimation, (KDE) avoids this disadvantage by placing a "bump" on every sample value in the sample space and then summing the heights of all bumps for every value in the sample space. This results in a smooth density estimates that are not affected by an arbitrarily chosen partition of the sample space.

The multivariate kernel density estimator in the q-dimensional case is defined as [1, 2]

$$\hat{f}_{\mathbf{h}}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_1 ... h_q} \mathcal{K} \left(\frac{x_{i_1} - x_1}{h_1}, ..., \frac{x_{i_q} - x_q}{h_q} \right), \quad (2)$$

with $\mathcal{K}$ denoting a multidimensional kernel function $\mathcal{K} : \mathbb{R}^q \rightarrow \mathbb{R}$ and $h_1, ..., h_q$ denoting bandwidths for each dimension. A common approach to build multidimensional kernel functions is to use a *product kernel*

$$\mathcal{K}(u_1, ..., u_q) = \prod_{i=1}^{q} K(u_i)\,, \qquad (3)$$

where K is a one-dimensional kernel function.

Intuitively, the kernel function determines the shape of the bumps placed around the sample values and the bandwidths $h_1, ..., h_q$ their width in each dimension. In case bandwidth is equal for all dimensions, multivariate radial-symmetric kernel functions can be used. Equation (2) then changes to

$$\hat{f}_h(\mathbf{x}) = \frac{1}{nh^q} \sum_{i=1}^{n} K\left(\frac{||\mathbf{x}_i - \mathbf{x}||}{h}\right). \qquad (4)$$

The shape of the approximated density function depends heavily on the bandwidth chosen for the density estimation. Small values of h lead to spiky density estimates showing spurious features. On the other hand too big values of h produce over-smoothed estimates that hide structural features.

The quality of a chosen bandwidth can in theory be determined by comparing the true probability density function f to the estimated density $\hat{f}$. Common criteria are the *Integrated Squared Error (ISE)*, its expected value, the *Mean Integrated Squared Error (MISE)* and the *Asymptotic Mean Squared Error (AMISE)* defined as

$$AMISE(h) = \frac{1}{nh}||K||_2^2 + h^{2k}\left(\frac{\mu_k(K)}{k!}\right)^2 ||f^k||_2^2, \quad (5)$$

with $||K||_2^2 = \int K^2(x)dx$ and $\mu_j(K) = \int x^j K(x)dx, j \in \mathbb{N}$ for any square integrable function K and order of the used kernel function k (usually $k = 2$). The optimal bandwidth can be chosen as the minimizer of $ISE(h)$, the minimizer of $MISE(h)$ or as the minimizer of $AMISE(h)$ - h_∞.

Since in practice the true density function f is unknown, it is not possible to compute the exact bandwidths from the sample values. Instead in most cases the bandwidth is computed by the following *rule of thumb*: differentiating Eq. (5) with respect to h and calculating the root of the derivative results in

$$h_\infty = \left(\frac{||K||_2^2}{\mu_2^2(K)||f^2||_2^2}\,,\right)^{\frac{1}{5}} n^{-\frac{1}{5}}\,, \qquad (6)$$

for $q = 1$ and the interesting case $k = 2$. The unknown density function f in this equation is now assumed to be the standard normal distribution re-scaled to have the same variance as the sample values. Choosing the Gaussian kernel function for K, the optimal bandwidth is in the one-dimensional case $h_{opt} = 1.06\hat{\sigma}n^{-\frac{1}{5}}$, where $\hat{\sigma}$ denotes the standard deviation, and respectively

$$h_{opt} = (4/(q+2))^{\frac{1}{q+4}}\, \hat{\sigma}\, n^{-\frac{1}{q+4}} \qquad (7)$$

for the q-dimensional case.

4. PROPOSED NOISE REDUCTION FILTER

The proposed filter is based on the idea of comparing image pixels contained in a filter window to their adjacent (neighbor) pixels. Filter output is that pixel in the filter window that is most similar to its neighborhood. The estimated probability density function therefore serves as a measure of similarity: usually the density estimate contains a clear maximum because adjacent pixels form a cluster in the color space, [5, 6]. If a pixel is similar to its neighborhood, the density estimation for that pixel results in a big value near the maximum. Noisy pixels on the other hand are almost always outliers from the cluster formed by adjacent pixels. Hence the density estimation for that pixels results in very small values.

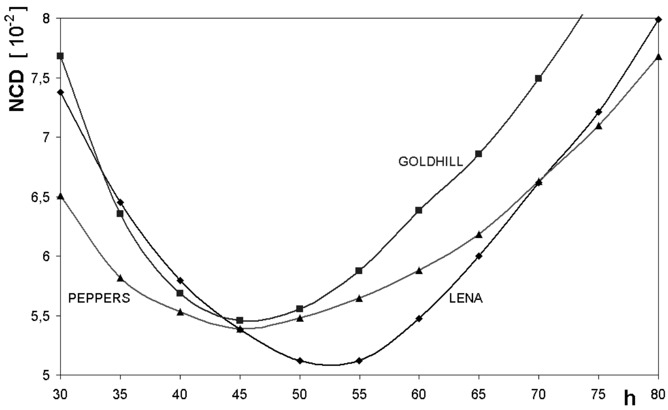


Fig. 1. Dependence of the filter efficiency (NCD) on the constant bandwidth of the Gaussian kernel, for test images *LENA*, *PEPPERS*, *GOLDHILL* contaminated by 5% impulsive noise. The value of $h = 55$ was used for the comparison with standard filtering techniques shown in Tab. 1

Given a set W of noisy image samples $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$ from the filter window as defined in section 2 let $\sim$ denotes the adjacency relation between two pixels contained in W. Assuming the 8-neighborhood system, the central pixel will have 8 adjacent neighbors, the pixels in the corners will have 3 adjacent neighbors and the remaining pixels in W will have 5 adjacent neighbors determined by the $\sim$ relation.

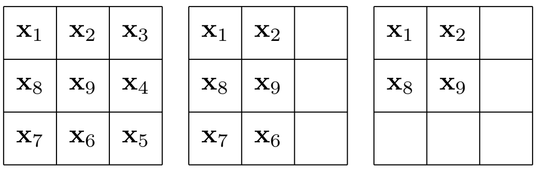
The probability density for sample $\mathbf{x}_i$ is then estimated as

$$\hat{f}_h(\mathbf{x}_i) = \sum_{\mathbf{x}_j \sim \mathbf{x}_i} K\left(\frac{||\mathbf{x}_j - \mathbf{x}_i||}{h}\right). \qquad (8)$$

The filter output is defined as that $\mathbf{x}_i$ for which $\hat{f}_h$ is maximal. In contrast to Eq. (4) the probability density is not normalized to bandwidth and number of sample values. The reason is that the values of $\hat{f}_h$ for different $\mathbf{x}_i$ are only used for comparison among each other and omission of normalization results in a significant performance gain as it priv-

Table 1. Filtering results achieved using test image *LENA* contaminated by impulsive noise using different kernels, (*G* denotes the Gaussian kernel, *Ep* the kernel of Epanechnikov, *Ex* the exponential kernel and *Tr* the linear, triangle kernel).

Noise p	0.05	0.05	0.05	0.10	0.10	0.10
Criterion	MAE	MSE	NCD	MAE	MSE	NCD
Noisy	2.54	393.3	0.0415	5.10	790.2	0.0838
VMF	**3.27**	**31.2**	**0.0387**	**3.42**	**34.2**	**0.0400**
BVDF	3.81	39.8	0.0400	3.95	44.2	0.0412
DDF	3.39	32.8	0.0389	3.51	35.4	0.0400
HDF	3.42	31.2	0.0399	3.55	33.9	0.0412
G, L_2, ad.	0.79	11.5	0.0093	0.98	20.2	0.0125
G, $h = 55$	0.42	11.8	0.0051	0.79	20.8	0.0100
G, L_1, ad.	0.82	14.8	0.0101	1.16	24.9	0.0149
Ep, L_2, ad.	1.17	15.3	0.0138	1.23	21.7	0.0151
Ex, L_2, ad.	0.43	10.59	0.0055	0.84	34.16	0.0128
Tr, L_2, ad.	0.45	14.01	0.0063	0.96	50.79	0.0159



a) **b)** **c)**

Fig. 2. Illustration of the adjacency concept $\sim$: the central pixel $\mathbf{x}_9$ has 8 neighbors belonging to the filtering window, the pixel $\mathbf{x}_8$ has then 5 adjacent neighbors and the pixel $\mathbf{x}_1$ has only three direct neighbors contained in W.

ileges the central sample, which has the largest number of neighbors, (Fig. 2).

The bandwidth is determined according to Eq. (7) and hence depends on the standard deviation $\hat{\sigma}$. Since $\hat{\sigma}$ is computed using only a few pixels from the filter window, the bandwidth is sensitive to noise if it is computed this way and may vary over a big range of values. As an option an experimentally chosen fixed value can be used as bandwidth to avoid this effect, (see Fig. 1).

5. FILTERING RESULTS

For evaluation purposes, the color test image LENA was corrupted with 1 to 10 percent impulsive noise defined by

$$\mathbf{x}_{ij} = \begin{cases} \mathbf{v}_{ij} & \text{with probability } p, \\ \mathbf{x}_{ij} & \text{else,} \end{cases} \tag{9}$$

where i, j define a pixel position, p describes the intensity of the noise process, $\mathbf{x}_{ij}$ denotes the original image pixel and $\mathbf{v}_{ij}$ denotes a pixel corrupted by the noise process $\mathbf{v}_{ij} = \{\nu_R, \nu_G, \nu_B\}$, where ν_R, ν_G, ν_B are random integer variables from the interval $[0, 255]$ updated for each corrupted pixel.

Filter quality is measured in *Mean Absolute Error* (MAE), *Mean Square Error* (MSE), *Signal to Noise Ratio* (SNR) and *Normalized Color Difference* (NCD), [7]. In general, these criteria reflect the filter capabilities of the signal detail preservation (MAE), the noise suppression (MSE, SNR) and the color chromaticity preservation (NCD).

Tab. 1 and Fig. 4 show the results of a quantitative and subjective comparison between the new filter scheme and the VMF as well as the Basic Vector Directional Filter (BVDF), Directional Distance Filter (DDF) and Hybrid Directional Filter (HDF), [7].

For experiments with fixed bandwidth a value of $h = 55$ was chosen, which brought subjectively good but not optimal results. As can be seen from Tab. 1 the noise reduction capability depends on the choice of the filter kernel.

Apart from the sometimes up to a few times lower MAE and NCD values compared with the vector median, the new filter shows enormous improvements in detail preservation for every used filter structure, (Figs. 3 and 4). The remarkably good results for the density estimation with fixed bandwidth indicate that the presented method of adaptive bandwidth selection does not work well enough for the impulsive noise and further research on this problem is needed.

6. CONCLUSIONS

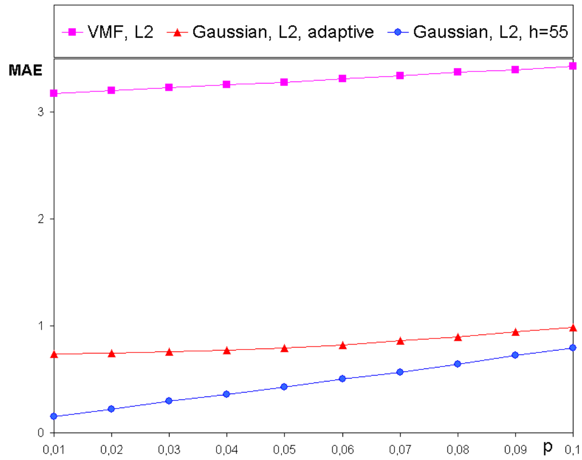
The experimental results show that the biggest advantage of the new filter is its excellent image detail preservation, (Tab. 1, Figs. 3, 4). The always very low values of MAE and NCD show that the new filter is clearly superior to VMF, BVDF, HDF and DDF in terms of detail preservation for all used filter kernels. Further, the comparison of different filter settings shows that the problem of choosing the bandwidth adaptive to the sample data is not yet completely solved and should be investigated in the future work.

Another advantage of the proposed filtering class is its low computational complexity compared to the VMF. For the VMF filtering the calculation of 36 distances between pixels are needed, whereas the new filter structure, with fixed bandwidth, requires only 20 different distances, which makes the new filter class interesting for real-time applications

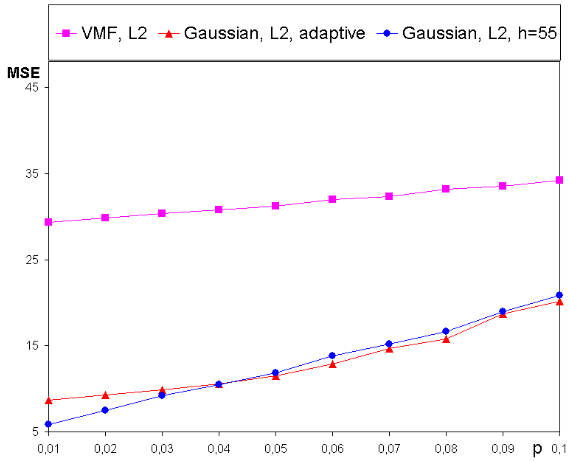
7. REFERENCES

[1] B.W. Silverman, "Density Estimation for Statistics and Data Analysis", London, Chapman and Hall, 1986

[2] D.W. Scott, "Multivariate Density Estimation", New York, John Wiley, 1992



a)



b)

Fig. 3. Results obtained with the new filtering framework in terms of MAE and MSE. The plots show the new filter performance for the Gaussian kernel with adaptive and constant bandwidth $h = 55$, in comparison with the vector median filter.

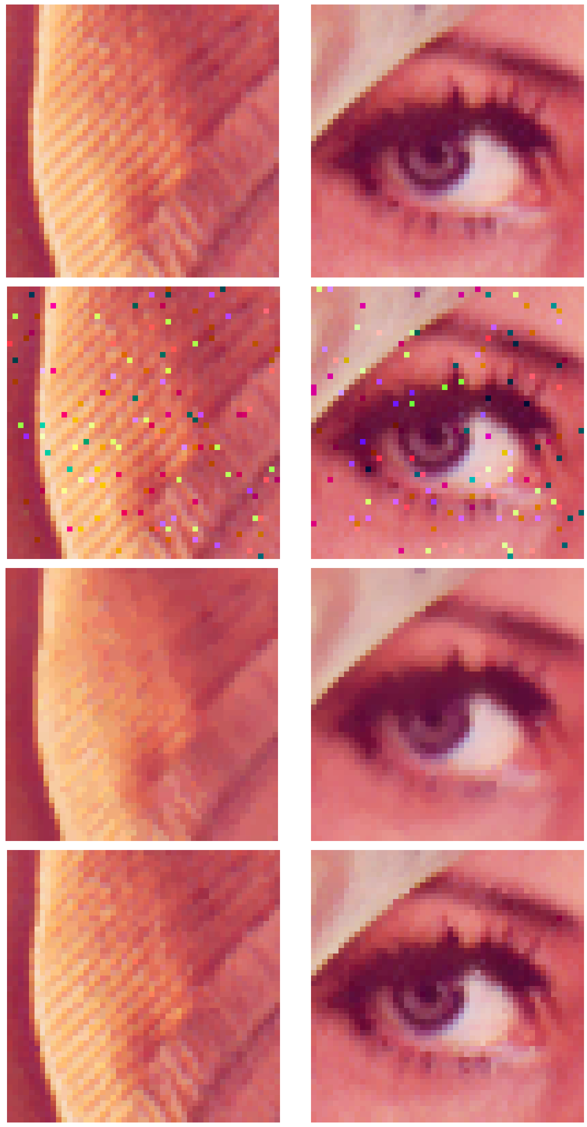
[3] E. Parzen, On estimation of a probability density function and mode, Ann. Math. Stat., Vol. 33, 1065-1076, 1962

[4] J. Astola, P. Haavisto, Y. Neuovo, Vector median filters, IEEE Proc. 78, 678-689, 1990

[5] A. Fotinos, N. Laskaris, G. Economou, S. Fotopoulos, Multidimensional fuzzy filtering using statistical parameters, Multidimensional Systems and Signal Processing, 10, 415-424, 1999

[6] D. Sindoukas, N. Laskaris, S. Fotopoulos., Algorithms for color image edge enhancement using potential functions, IEEE Signal Processing Letters, Vol. 4, No. 9, September 1997

[7] K.N. Plataniotis, A.N. Venetsanopoulos, ”Color Image Processing and Applications”, Springer Verlag, August 2000.



a)
b)
c)
d)

Fig. 4. Illustration of the efficiency of the new filter in comparison with the VMF: **a)** parts of test image *LENA*, **b)** test images corrupted by impulsive noise - $p = 0.05$, **c)** VMF output, **d)** new filter output using the Gaussian kernel and the adaptive scheme with the L_2 vector norm.