appeared in
10. Workshop Farbbildverarbeitung 2004, Koblenz, Online-Proceedings
http://www.uni-koblenz.de/icv/fws2004/

Nonparametric Noise Removal in Color Images

Bogdan Smolka *

Polish-Japanese Institute of Information Technology
Koszykowa 86, 02-008, Warsaw
bsmolka@pjwstk.edu.pl

Abstract. In this paper the problem of nonparametric impulsive noise removal in multichannel images is addressed. The proposed filter class is based on the nonparametric estimation of the density probability function in a sliding filter window. The obtained results show good noise removal capabilities and excellent structure preserving properties of the new impulsive noise reduction technique.

1 Introduction

The majority of the nonlinear, multichannel filters are based on the ordering of vectors in a sliding filter window. The output of these filters is defined as the lowest ranked vector according to a specific vector ordering technique.

Let the color images be represented in the commonly used RGB color space and let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ be N samples from the sliding filter window W. Each of the $\mathbf{x}_i$ is an m-dimensional multichannel vector, (in our case $m = 3$). The goal of the vector ordering is to arrange the set of N vectors $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ belonging to W using some sorting criterion.

In [1, 2] the ordering based on the cumulative distance function $R(\mathbf{x}_i)$ has been proposed: $R(\mathbf{x}_i) = \sum_{j=1}^{N} \rho(\mathbf{x}_i, \mathbf{x}_j)$, where $\rho(\mathbf{x}_i, \mathbf{x}_j)$ is a function of the distance among $\mathbf{x}_i$ and $\mathbf{x}_j$. The ordering of the scalar quantities according to $R(\mathbf{x}_i)$ generates the ordered set of vectors. The most commonly used measure to quantify distance between two multichannel signals is the Minkowski norm $\rho_\gamma(\mathbf{x}_i, \mathbf{x}_j) = \left[\sum_{k=1}^{m} |x_{ik} - x_{jk}|^\gamma\right]^{1/\gamma}$. The Minkowski metric includes the city-block distance ($\gamma = 1$), Euclidean distance ($\gamma = 2$) and chess-board distance ($\gamma = \infty$) as the special cases.

One of the most important noise reduction filter is the vector median. In the case of gray scale images, given a set W containing N samples, the median of the set is defined as $x_{(1)} \in W$ such that

$$\sum_j \left|x_{(1)} - x_j\right| < \sum_j \left|x_i - x_j\right|, \quad \forall \;\; x_i, x_j \in W. \tag{1}$$

Median filters exhibit good noise reduction capabilities, (especially when long tailed distribution noise is involved) and outperform simple nonadaptive linear filters in preserving signal discontinuities. As in many applications the signal is multidimensional,

* This research has been supported by a grant No PJ/B/01/2004 from the Polish-Japanese Institute of Information Technology

in [3] the *Vector Median Filter* (VMF) was introduced, by generalizing the definition (1) using a suitable vector norm. Given a set W of N vectors, the vector median of the set is defined as $\mathbf{x}_{(1)} \in W$ satisfying

$$\sum_j \left\|\mathbf{x}_{(1)} - x_j\right\| < \sum_j \left\|\mathbf{x}_i - \mathbf{x}_j\right\|, \quad \forall\ \mathbf{x}_i, \mathbf{x}_j \in W\,. \tag{2}$$

The orientation difference between two vectors can also be used as their distance measure. This so-called vector angle criterion is used by the *Vector Directional Filters* (VDF), to remove vectors with atypical directions, [4]. The *Basic Vector Directional Filter* (BVDF) is a ranked-order, nonlinear filter which parallelizes the VMF operation. However, a distance criterion based on the angles between vectors is utilized. To improve the efficiency of the directional filters, another method called *Directional-Distance Filter* (DDF) was proposed. This filter retains the structure of the BVDF, but utilizes the combined distance criterions to order the vectors inside the processing window, [4, 5].

2 Nonparametric Estimation

Applying statistical pattern recognition techniques requires the estimation of the probability density function of the data samples. Nonparametric techniques do not assume a particular form of the density function since the underlying density of the real data rarely fits common density models.

Nonparametric Density Estimation is based on placing a kernel function on every sample and on the summation of the values of all kernel function values at each point in the sample space, [6]. The nonparametric approach to estimating multichannel densities can be introduced by assuming that the color space occupied by the multichannel image pixels is divided into m-dimensional hypercubes. If h_N is the length of an edge of a hypercube, then its volume is given by $V_N = h_N^m$. If we are interested in estimating the number of pixels falling in the hypercube of volume V_N, then we can define the window function $\phi(\mathbf{x_i}) = 1$, if $|x_{ij}| \leq 1/2$, $j = 1, \ldots, m$ and 0 otherwise, which defines a unit hypercube centered in the origin.

The function $\phi\left(\left\|\mathbf{x} - \mathbf{x}_i\right\|/h_N\right)$ is equal to unity if the pixel $\mathbf{x}_i$ falls within the hypercube V_N centered at $\mathbf{x}$ and is zero otherwise. The number of pixels in the hypercube with the length of edges equal to h_N is then $k_N = \sum_{i=1}^N \phi\left(\|\mathbf{x} - \mathbf{x}_i\|/h_N\right)$ and the estimate of the probability that a sample $\mathbf{x}$ is within the hypercube is $p_N = k_N/NV_N$, which gives

$$p_N(\mathbf{x}) = (NV_N)^{-1} \sum_{i=1}^N \phi\left(\|\mathbf{x} - \mathbf{x}_i\|/h_N\right). \tag{3}$$

This estimate can be generalized by using a smooth kernel function K in place of $\phi(\cdot)$ and the width parameter h_N satisfying: $K(\mathbf{x}) = K(-\mathbf{x})$, $K(\mathbf{x}) \geq 0$, $\int K(\mathbf{x})\, d\mathbf{x} = 1$ and $\lim_{N\to\infty} h_N = 0$, $\lim_{N\to\infty} h_N^m = \infty$.

The multivariate estimator in the m-dimensional case is defined as

$$p_N^*(\mathbf{x}) = \frac{1}{N}\sum_{i=1}^N \frac{1}{h_1 \ldots h_q} \mathcal{K}\left(\frac{|x_1 - x_{i1}|}{h_1}, \cdots, \frac{|x_q - x_{iq}|}{h_m}\right), \tag{4}$$

with $\mathcal{K}$ denoting a multidimensional kernel function $\mathcal{K}\colon \mathbb{R}^m \to \mathbb{R}$, $h_1, \ldots, h_m$ denoting bandwidths for each dimension and N being the number of samples in W. A common approach to build multidimensional kernel functions is to use a *product kernel* $\mathcal{K}(u_1, \ldots, u_m) = \prod_{i=1}^m K(u_i)$, where K is a one-dimensional kernel function

$$p_N^*(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^{N} \prod_{j=1}^{m} \left(\frac{|x_{ij} - x_j|}{h_i} \right) . \tag{5}$$

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.

If we chose the Gaussian kernel, then the density estimate of the unknown probability density function at $\mathbf{x}$ is obtained as a sum of kernels placed at each sample $\mathbf{x}_i$

$$p_N(\mathbf{x}, h) = \frac{1}{N \left(h\sqrt{2\pi} \right)^m} \sum_{i=1}^{N} \exp \left(-\frac{\|\mathbf{x} - \mathbf{x}_i\|^2}{2h^2} \right) . \tag{6}$$

The smoothing parameter h depends on the local density estimate of the sample data. The form of the data dependent smoothing parameter is of great importance for the non-parametric estimator.

Choosing the Gaussian kernel function for $\mathcal{K}$, the optimal bandwidth is

$$h^* = (4/(m+2))^{-\frac{1}{m+4}} \, \hat{\sigma} \, N^{-\frac{1}{m+4}} , \tag{7}$$

where σ denotes the approximation of the standard deviation of the samples. In one dimensional case (7) reduces to the well known, 'rule of thumb', $h^* = 1.06 N^{-\frac{1}{5}} \hat{\sigma}$, [6]. A version which is more robust against outliers in the sample set can be constructed if the interquartile range is used as a measure of spread instead of the variance. This modified estimator is $h^* = 0.79 \varrho N^{-\frac{1}{5}} \hat{\sigma}$, where ϱ is the inter-quartile range. Another robust estimate of the optimal bandwidth is $h^* = 0.9 A N^{-\frac{1}{5}} \hat{\sigma}$ with $A = \min(\hat{\sigma}, \varrho/1.34)$. Generally the simplified rule of choosing the optimal bandwidth h can be written as

$$h_1^* = C \, \hat{\sigma} \, N^{-\frac{1}{m+4}} , \tag{8}$$

where C is an appropriate weighting coefficient.

From the maximum likelihood principle and assuming independence of the samples, one can write the likelihood of drawing the complete dataset as the product of the densities of one sample

$$\mathcal{L}(h) = \prod_{j=1}^{N} p_N(\mathbf{x}_j, h) = \prod_{j=1}^{N} \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\left(h\sqrt{2\pi} \right)^m} \exp \left(-\frac{\|\mathbf{x}_j - \mathbf{x}_i\|^2}{2h^2} \right) . \tag{9}$$

As this likelihood function has a global maximum for h=0, in [7] a modified approach has been proposed

$$\mathcal{L}^*(h) = \left[\prod_{j=1}^{N} \frac{1}{N} \sum_{i=1,\, i \neq j}^{N} \frac{1}{\left(h\sqrt{2\pi} \right)^m} \exp \left(-\frac{\|\mathbf{x}_j - \mathbf{x}_i\|^2}{2h^2} \right) \right]^{\frac{1}{m}} . \tag{10}$$

This function has one maximum for h, which can be found by setting to 0 the derivative of the logarithm of $\mathcal{L}^*(h)$ with respect to h. A crude but rather fast way to obtain an approximate solution hereof is by assuming that the density estimate of Eq. (5) on a certain location **x** in the feature space is determined by the nearest kernel only, [7]. In this paper we use the optimal h derived from (10) defined as

$$h_2^* = C\left((mN)^{-1}\sum_{j=1}^{N}\|\tilde{\mathbf{x}}_j - \mathbf{x}_j\|^2\right)^{\frac{1}{2}}, \tag{11}$$

where $\tilde{\mathbf{x}}_i$ represents the nearest neighbor of the sample $\mathbf{x}_i$, and C is a tuning parameter.

3 Proposed Algorithm

Let us assume a filtering window W containing N image pixels, $\{x_1, \ldots, x_N\}$ and let us define the similarity function $\mu : [0;\infty) \to \mathbb{R}$ which is non-ascending and convex in $[0;\infty)$ and satisfies $\mu(0) = 1$, $\mu(\infty) = 0$. The similarity between two pixels of the same intensity should be 1, and the similarity between pixels with minimal and maximal gray scale values should be very close to 0. The function $\mu(x_i, x_j)$ defined as $\mu(x_i, x_j) = \exp\{-[(x_i - x_j)/h]^2\}$, where h is the bandwidth of the Gaussian kernel, defined by (8) or (11), satisfies the required conditions.

Let us additionally define the cumulated sum M of similarities between a given pixel and all other pixels belonging to window W. For the central pixel x_1 we introduce M_1 and for the neighbors of x_1 we define M_k as

$$M_1 = \sum_{j=2}^{N}\mu(x_1, x_j), \quad M_k = \sum_{j=2,\ j\neq k}^{N}\mu(x_k, x_j),\ k > 1, \tag{12}$$

which means that for x_k, which are neighbors of x_1, we do not take into account the similarity between x_k and x_1, which is the main idea of this algorithm. The omission of the similarity $\mu(x_k, x_1)$ when calculating M_k, privileges the central pixel, as in the calculation of M_1 we have $N-1$ similarities $\mu(x_1, x_k)$, $k > 2$ and for M_k, $k > 1$ we have only $N-2$ similarity values, as the central pixel x_1 is excluded from the calculation of M_k, [8, 9].

In the construction of the new filter, the reference pixel x_1 in the window W is replaced by one of its neighbors if $M_1 < M_k,\ k = 2, \ldots, N$. If this is the case, then x_1 is replaced by that x_{k^*} for which $k^* = \arg\max M_k,\ k = 2, \ldots, N$. In other words x_1 is detected as being corrupted if $M_1 < M_k,\ k = 2, \ldots, N$ and is replaced by its neighbors x_k which maximizes the sum of similarities M between all the pixels from W excluding the central pixel.

The basic assumption is that a new pixel must be taken from the window W, (introducing pixels, that do not occur in the image is prohibited like in the VMF). For this purpose μ must be convex, which means that in order to find a maximum of the sum of similarity functions M it is sufficient to calculate the values of M only in points $x_1, x_2, \ldots, x_N$.

The presented approach can be applied in a straightforward way to multichannel images using the similarity function defined as $\mu(\mathbf{x}_i, \mathbf{x}_j) = \exp\{-[\|\mathbf{x}_i - \mathbf{x}_j\|/h]^2\}$, where $\|\cdot\|$ denotes the specific vector norm and h denotes the bandwidth. Now in exactly the same way we can maximize the total similarity function M for the vector case.

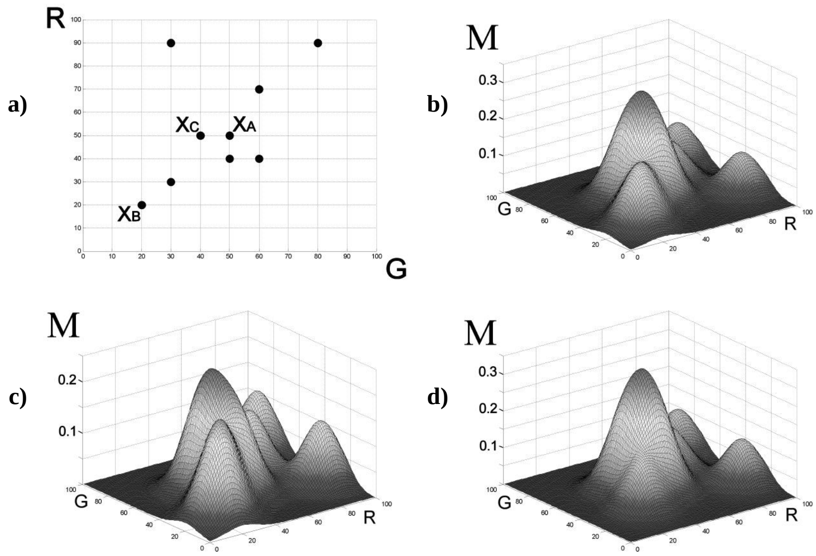

Fig. 1. Impulsive noise removal technique in the 2D case.

The working scheme of the new filter is presented in Fig. 1 for the two-dimensional data. Fig. 1 **a**) depicts the arrangement of pixels in W and Fig. **b**) their nonparametric probability density estimation. Figs. **c**) and **d**) present the density plots for the cases when the central pixels x_A and x_B are removed from W. It can be seen that in the first case **c**) the pixel $x_1 = x_A$ will be retained and in the second case **d**) the pixel $x_1 = x_B$ will be replaced by x_A. The pixel x_A will be preserved, as in Fig. **c**) the plot attains its maximum at x_C, but this maximum is less than the maximum for x_A in Fig. **b**). Regarding sample x_B, its rejection causes that the maximum is attained at x_A and this pixel will replace the central pixel x_B.

4 Results

The performance of the proposed impulsive noise reduction filters was evaluated using the widely used PSNR quality measure. Figure 2a) shows the dependence of the noise attenuation capability of the proposed filter class on the bandwidth type h_1^* and h_2^* defined by (8) and (11). Clearly the filter based on the h_2^* outperforms the technique based on the h_1 bandwidth for the whole range of used contamination probabilities p, ($p = 0.01$ - 0.1).

Figure 2b) presents the dependence of the PSNR restoration quality measure on the kind of the Minkowski norm. Surprisingly, the L_∞ norm yields significantly better results than the L_1 or L_2 norms. This is the result of the construction of the h_2^* bandwidth, which depends on the nearest neighbor in the sliding filter window. This behavior is advantageous, as the calculation of the L_∞ norm is much faster than the evaluation of distances determined by L_1, L_2 norms.

The efficiency of the filters based on adaptive h_1^* and h_2^* bandwidths are dependent, (especially for very small noise contamination) on the coefficient C in (8) and (11). Figure 2c) shows the dependence of PSNR for the filter based on h_2^* as a function of C in (11). For low noise intensity the parameter C should be significantly larger than for the case of images corrupted by heavy noise process. However, setting C to 4 is an acceptable trade-off, as can be seen in Fig. 2 d), which depicts the efficiency of the proposed filter in comparison with VMF, AMF and BVDF. It can be observed that although the $C = 4$ is not an optimal setting for the whole range of tested noise intensities, nevertheless the described filter yields much better results than the traditional techniques.

This is also testified by Fig. 3, which compares the filtering results obtained by the filter based on adaptive h_2^* bandwidth with the performance of the *reference* VMF, BVDF, DDF filter. As can be observed the new filtering has much better detail preserving properties than VMF, BVDF and DDF.

5 Conclusions

In this paper a new nonparametric technique of impulsive noise removal in multichannel images has been proposed. The described filter class is based on the estimation of the kernel bandwidth using the technique proposed in [7]. The experiments revealed, that the proposed algorithm yields the best results when applying the L_∞ norm, which makes the filter computationally very attractive. The obtained results show that the proposed technique excels significantly over the standard techniques like VMF, BVDF and DDF. The future work will focus on the automatic adjustment of the tuning parameter C in (8) and (11).

References

1. I. Pitas, P. Tsakalides, Multivariate ordering in color image processing, IEEE Trans. on Circuits and Systems for Video Technology, 1, 3, 247-256, 1991.
2. K. Tang, J. Astola, Y. Neuovo, Nonlinear multivariate image filtering techniques, IEEE Trans. on Image Processing, 4, 6, 788-797, 1995.
3. J. Astola, P. Haavisto, Y. Neuvo, Vector median filters, Proceedings of the IEEE, 78, 678-689, 1990.
4. P.E. Trahanias, A.N. Venetsanopoulos, Vector directional filters: a new class of multichannel image processing filters, IEEE Trans. on Image Processing, 2, 4, 528-534, 1993.
5. K.N. Plataniotis, A.N. Venetsanopoulos, "Color Image Processing and Applications", Springer Verlag, August 2000.
6. D.W. Scott, "Multivariate Density Estimation", New York, John Wiley, 1992.
7. M.A. Kraaijveld, A Parzen classifier with an improved robustness against deviations between training and test data, Pattern Recognition Letters, 17, 679-689, 1996.
8. B. Smolka, K.N. Plataniotis, A. Chydzinski, M. Szczepanski, A.N. Venetsanopulos, K. Wojciechowski, Self-adaptive algorithm of impulsive noise reduction in color images, Pattern Recognition, 35, 1771-1784, 2002.
9. B. Smolka, R. Lukac, A. Chydzinski, K.N. Plataniotis, K. Wojciechowski, Fast adaptive similarity based impulsive noise reduction filter, Real Time Imaging, 9, 261-276, 2003.

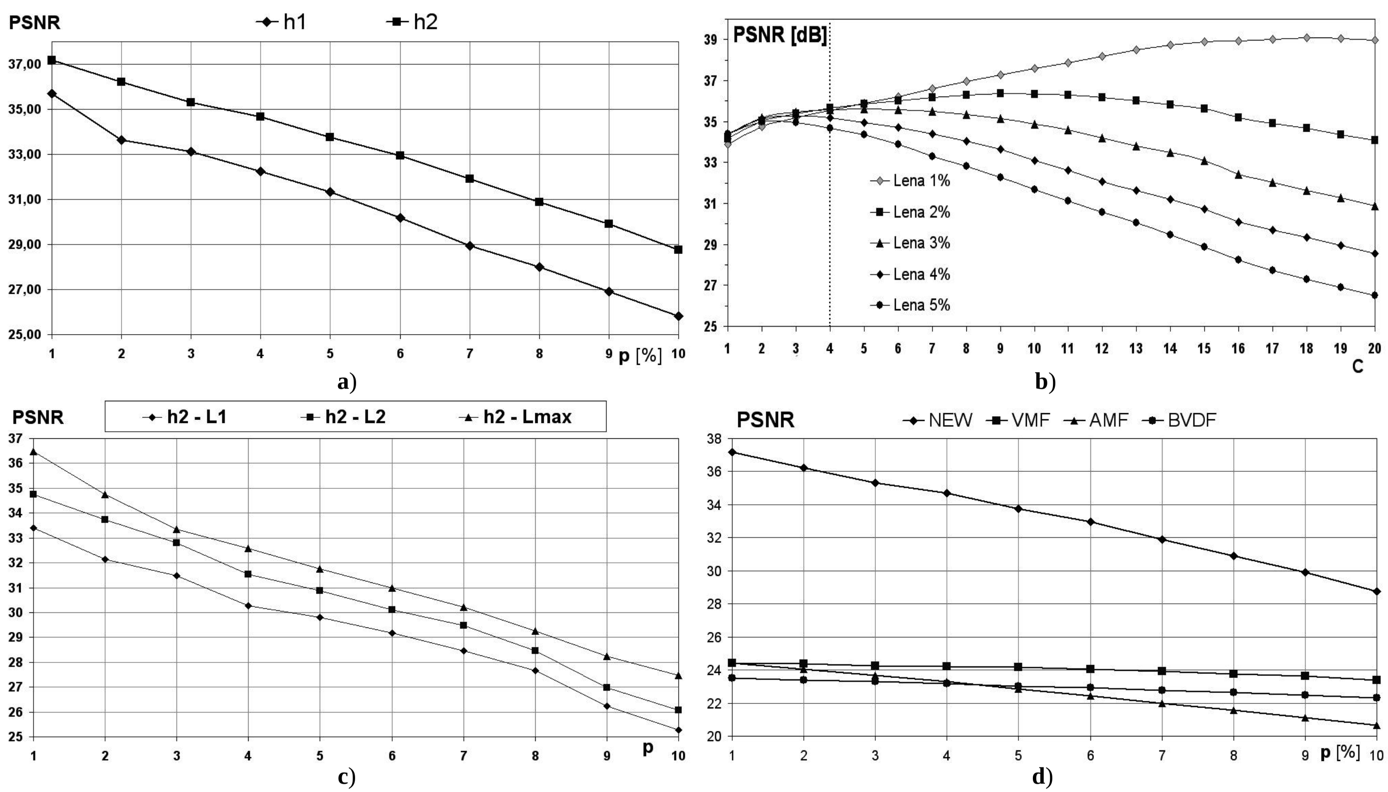

Fig. 2. Dependence of the efficiency of the proposed filtering scheme on h_1^* (8) and h_2^* (11) - (**a**), besides the dependence of the PSNR on the tuning parameter C in (11) - (**b**) and the dependence on the kind of Minkowski norm for the bandwidth h_2* - (**c**). Figure (**d**) shows the comparison of results obtained using the h_2^* bandwidth, L_∞ norm and $C = 4$ with the standard multichannel filters VMF and BVDF, (test were performed on the color image *LENA*); p denotes the probability of a pixel corruption - to RGB channels random, uniformly distributed values from the interval [0,255] were assigned.

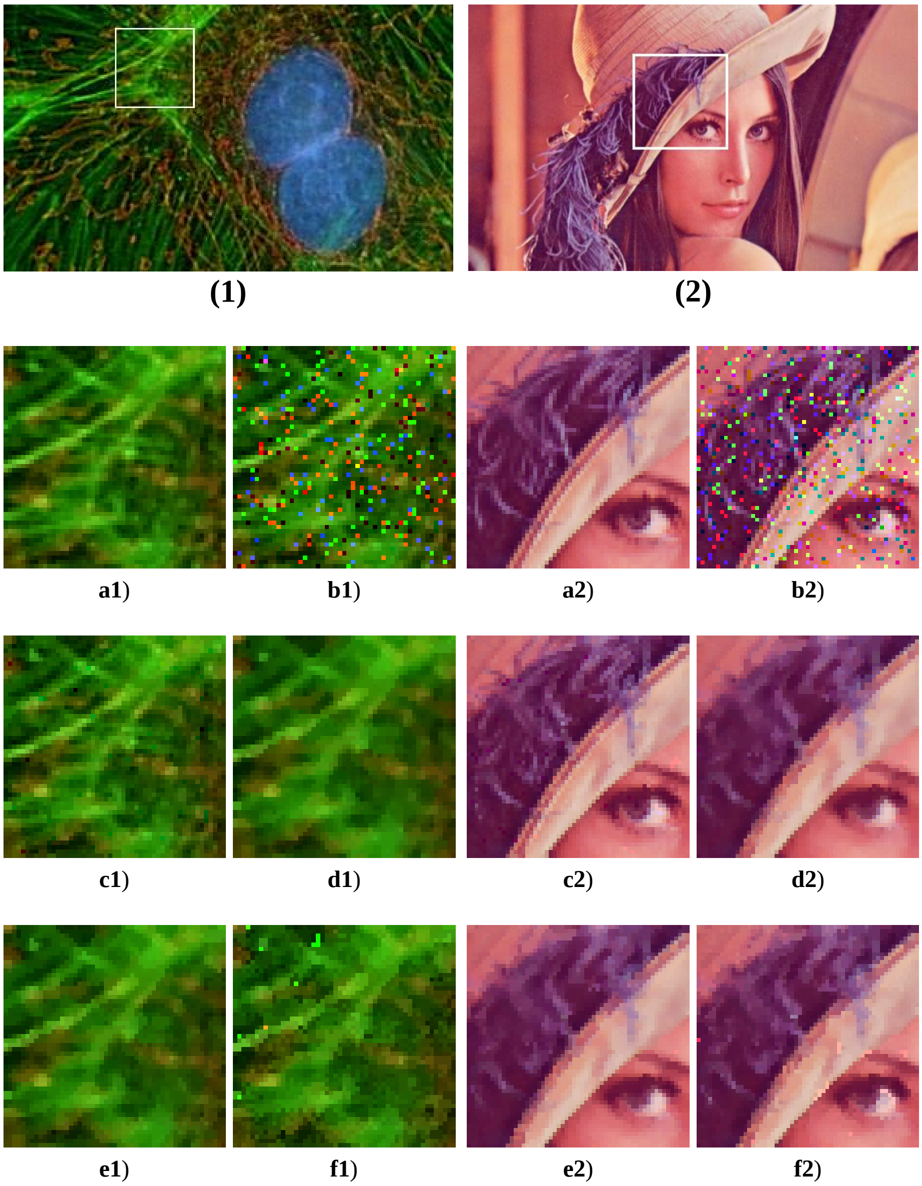

Fig. 3. Illustrative example of the efficiency of the proposed algorithm: **a**) zoomed parts of the test color images, **b**) image corrupted by 3% of impulsive noise, **c**) image after filtering with the proposed filter, **d**) VMF output, **e**) DDF output, **f**) BVDF output.