Multivariate Ordering in Color Image Filtering

Ioannis Pitas and Panagiotis Tsakalides

Abstract—**Multivariate data ordering and its use in color image filtering are presented in this paper. Marginal ordering, reduced ordering, and the related multichannel filters are presented in detail. Several of the filters presented are extensions of the single-channel filters based on order statistics, e.g., median filters, trimmed mean filters. The statistical analysis of the marginal order statistics is presented for the p-dimensional case. The performance of the proposed filters is thoroughly examined theoretically and by simulation for additive and impulsive noise removal.**

I. Introduction

DIGITAL color image processing has very important applications, i.e., digital video image processing [1], [2], image synthesis [3], and high-definition TV (HDTV) [4]. However, until recently the development of color image processing techniques did not follow the pace of the development of black-and-white (BW) digital image processing. There are various reasons for this. BW digital image processing was the first natural candidate for development. Furthermore, the simplest approach to digital color image processing is to use BW techniques on each channel independently. Technological problems have also slowed down the development of digital color image processing, e.g., the availability and cost of color frame grabbers, the large memory, and computation speed required. In past years, such technological problems have been solved. BW digital image processing techniques have reached a maturity that allowed scientists to target their research on color image processing. Furthermore, multichannel signal processing has produced significant results that can be used in color image processing. An approach to nonlinear multichannel (color) image processing is presented in this paper.

In his classical theory, Thomas Young stated that any color can be reproduced by mixing an appropriate set of three primary colors [5]. Therefore, any color image can be considered as a two-dimensional three-channel sequence. Furthermore, in cases of constant brightness of a two-variable scheme can be developed by factoring out luminance while keeping only the chromaticity attributes—e.g., hue and saturation—thus projecting the three-dimensional color space on a plane. In this case, we can treat color image as a two-dimensional two-channel sequence. Therefore, color image processing is essentially multichannel two-dimensional signal processing. Until recently, little attention was paid to the multichannel nature of color images. Thus, BW image processing techniques were applied to each channel separately. Later, transformation techniques, e.g., the Karhunen–Loeve transformation, have been used to decorrelate the three channels in order to apply BW image processing techniques afterward [6]. Although this is a valid approach, we think that a more natural way to deal with color image processing is to apply multichannel signal processing techniques [7]–[9].

In the following, the use of ordering of multivariate data in multichannel signal processing is investigated with particular application to color image processing. The work presented in this paper extends previously reported work [10]–[12]. We are concerned mainly with two ordering techniques—marginal ordering (independent data ordering in each channel) and reduced ordering—that are based on distance functions of multivariate data. These ordering techniques are described in Section II. The probability distributions of the p-dimensional marginal order statistics are investigated in Section III. A family of multichannel filters based on multivariate data ordering, such as the marginal median, the vector median, the marginal α-trimmed mean, and the multichannel modified trimmed mean filter are described in Section IV. The performance of the marginal median and the vector median filters in impulsive noise filtering is investigated in Section V. Simulation examples concerning the performance of the filters under study are described in Section VI. Conclusions are drawn in Section VII.

II. Multivariate Data Ordering

Let us denote by $\mathbf{X}$ a p-dimensional random variable, i.e., a p-dimensional vector of random variables $\mathbf{X} = [X_1, \ldots, X_p]^T$. We denote by $f(\mathbf{X})$ and $F(\mathbf{X})$ the probability density function (pdf) and the cumulative density function (cdf) of this p-dimensional random variable, respectively. Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be n random samples from a p-dimensional distribution having mean $\mathbf{m}$ and dispersion matrix Σ [13]. The notion of data ordering cannot be extended in a straightforward manner from the one-dimensional case to the case of multivariate data. An excellent treatment of multivariate data ordering was given in [14]. It is shown that there are several ways to order multivariate data, but none of them is unambiguous nor universally agreeable. The following so-called subordering principles are discussed in the literature: 1) marginal ordering, 2) reduced (aggregate) ordering, 3) partial ordering, and 4) conditional ordering [14]. We are mainly interested in marginal and reduced ordering because they can be used easily to develop nonlinear filters for color image processing.

In marginal ordering, the multivariate samples are ordered along each of the p dimensions:

$$\begin{aligned} x_{1(1)} &\le x_{1(2)} \le \ldots \le x_{1(n)} \\ x_{2(1)} &\le x_{2(2)} \le \ldots \le x_{2(n)} \\ x_{p(1)} &\le x_{p(2)} \le \ldots \le x_{p(n)} \end{aligned} \quad (1)$$

Manuscript received December 12, 1990; revised June 27, 1991.

The authors are with the Department of Electrical Engineering, University of Thessaloniki, Thessaloniki 54006, Greece.

IEEE Log Number 9102805.

1051-8215/91/0900-0247$01.00 ©1991 IEEE

i.e., ordering is performed in each channel of the multichannel signal independently. $x_{1(1)}, x_{2(1)}, \ldots, x_{p(1)}$ are the minimal elements in each dimension. $x_{1(n)}, x_{2(n)}, \ldots, x_{p(n)}$ are the maximal elements in each dimension. $\mathbf{x}_{(\nu+1)} = [x_{1(\nu+1)}, \ldots, x_{p(\nu+1)}]^{\mathbf{T}}$ is the (marginal) median of the multivariate data for $n = 2\nu + 1$. The ith marginal-order statistic is the vector $\mathbf{x}_{(i)} = [x_{1(i)}, \ldots, x_{p(i)}]^{\mathbf{T}}$. The vector $[x_{1(r_1)}, x_{2(r_2)}, \ldots, x_{p(r_p)}]^{\mathbf{T}}$ for arbitrary r_i, $1 \le r_i \le n$, $i = 1, \ldots, p$ is a marginal-order statistic too. Needless to say that the median or any ith marginal-order statistic may not correspond to any of the samples $\mathbf{x}_1, \ldots, \mathbf{x}_n$. In contrast, in the one-dimensional case there is a one-to-one correspondence between the samples $\mathbf{x}_1, \ldots, \mathbf{x}_n$ and the order statistics $x_{(1)}, \ldots, x_{(n)}$ [15]. The statistical analysis of the p-dimensional marginal-order statistics is described in Section III.

The reduced ordering (R ordering) is based on the generalized distance

$$d = (\mathbf{x} - \mathbf{a})^{\mathbf{T}} \cdot \Gamma^{-1} \cdot (\mathbf{x} - \mathbf{a}) \tag{2}$$

of a sample $\mathbf{x}$ from a point $\mathbf{a}$ that may be either the origin or the sample arithmetic mean $\bar{\mathbf{x}}$:

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \tag{3}$$

or the marginal median $\mathbf{x}_{(\nu+1)}$. Γ may be the identity matrix, the dispersion matrix Σ, or the sample dispersion matrix $\mathbf{S}$:

$$\mathbf{S} = \frac{1}{n-1} \sum_{i=1}^{n} (\mathbf{x}_i - \bar{\mathbf{x}}) \cdot (\mathbf{x}_i - \bar{\mathbf{x}})^{\mathbf{T}}. \tag{4}$$

The various data $\mathbf{x}_i$ are ordered according to their distances d_i from $\mathbf{a}$. Thus, multivariate ordering is reduced to one-dimensional ordering.

Both the reduced and marginal ordering can be used to develop a family of nonlinear filters. Before the description of this family, we concentrate on the study of some statistical properties of the p-dimensional marginal-order statistics. This study is a continuation of the study reported in [11].

III. Probability Distributions of the Marginal-Order Statistics

The study of the probability distribution of marginal-order statistics started relatively early, and the probability density function of the marginal median was investigated [16]. Recurrence relations for the probability distribution of two-dimensional marginal-order statistics were given [17]. The asymptotic probability distribution of the p-dimensional marginal-order statistics was also studied [18]. In the following, expressions for the cdf and the pdf of the p-dimensional order statistics are derived. The work presented here is a generalization of the work reported in [11] for the two- and three-dimensional cases. When n data samples are available, the cdf of the two-dimensional order statistic $[X_{1(r_1)}, X_{2(r_2)}]^{\mathbf{T}}$ is given by [11]:

$$\begin{aligned} F_{(r_1, r_2)}(x_1, x_2) &= \sum_{i_1=r_1}^{n} \sum_{i_2=r_2}^{n} P\{i_1 \text{ of } X_{1i} \le x_1, i_2 \text{ of } X_{2i} \le x_2\} \\ &= \sum_{i_1=r_1}^{n} \sum_{i_2=r_2}^{n} \sum_{n_0=\max(0,\, i1+i2-n)}^{\min(i_1,\, i_2)} \\ &\quad \times \frac{n!}{n_0!(i_1-n_0)!(i_2-n_0)!(n-i_1-i_2+n_0)!} \\ &\quad \cdot F_0^{n_0}(x_1, x_2) \cdot F_1^{i_2-n_0}(x_1, x_2) \cdot F_2^{i_1-n_0}(x_1, x_2) \\ &\quad \cdot F_3^{n-i_1-i_2+n_0}(x_1, x_2) \end{aligned} \tag{5}$$

where $F_i(x_1, x_2)$, $i = 0, \ldots, 3$ are the probability masses on the four regions of the (x_i, x_2) plane and they are given by

$$\begin{aligned} F_0(x_1, x_2) &= P\{X_1 \le x_1, X_2 \le x_2\} = F(x_1, x_2) \\ F_1(x_1, x_2) &= P\{X_1 > x_1, X_2 \le x_2\} \\ F_2(x_1, x_2) &= P\{X_1 \le x_1, X_2 > x_2\} \\ F_3(x_1, x_2) &= P\{X_1 > x_1, X_2 > x_2\}. \end{aligned} \tag{6}$$

$F(x_1, x_2)$ is the joint cdf of the random vector $\mathbf{x} = [X_1, X_2]^{\mathbf{T}}$. Equation (5) is relatively complicated and does not give analytic expressions of $F_{(r_1, r_2)}(x_1, x_2)$ for arbitrary data cdf $F(x_1, x_2)$. However, it can be easily computed numerically.

We proceed now to the derivation of an expression for the cdf $F_{(r_1, \ldots, r_p)}(x_1, \ldots, x_p)$ of the p-dimensional order statistic $[X_{1(r_1)}, X_{2(r_2)}, \ldots, X_{p(r_p)}]^{\mathbf{T}}$:

$$\begin{aligned} &F_{(r_1, \ldots, r_p)}(x_1, \ldots, x_p) \\ &\quad = P\{X_{1(r_1)} \le x_1, \ldots, X_{p(r_p)} \le x_p\} \\ &\quad = \sum_{i_1=r_1}^{n} \cdots \sum_{i_p=r_p}^{n} \\ &\qquad \cdot P\{i_1 \text{ of } X_{1i} \le x_1, \ldots, i_p \text{ of } X_{pi} \le x_p\}. \end{aligned} \tag{7}$$

It is proven in Appendix A that the cdf for the p-dimensional case is given by

$$\begin{aligned} &F_{(r_1, \ldots, r_p)}(x_1, \ldots, x_p) \\ &\quad = \sum_{i_1=r_1}^{n} \cdots \sum_{i_p=r_p}^{n} \sum_{n_0} \cdots \sum_{n_{2^p-1}} \frac{n!}{\prod_{i=0}^{2^p-1} n_i!} \\ &\qquad \times \prod_{i=0}^{2^p-1} F_i^{ni}(x_1, \ldots, x_p) \end{aligned} \tag{8}$$

subject to the constraints

$$n_i \ge 0 \qquad i = 0, \ldots, 2^p - 1 \tag{9}$$

$$\sum_{i=0}^{2^p-1} n_i = n \tag{10}$$

$$\sum_{j_0=0} n_j = i_1, \ldots, \sum_{j_{p-1}=0} n_j = i_p. \tag{11}$$

The probability density function $f_{(r_1, \ldots, r_p)}(x_1, \ldots, x_p)$ is given by:

$$f_{(r_1, \ldots, r_p)}(x_1, \ldots, x_p) = \frac{\partial^p F_{(r_1, \ldots, r_p)}(x_1, \ldots, x_p)}{\partial x_1 \cdots \partial x_p}. \tag{12}$$

We describe an alternative way to find the pdf of the marginal-order statistics as follows. We shall focus on the two-dimensional case. Let us consider the geometry in Fig. 1. A marginal-order statistic $[x_{1(r1)}, x_{2(r2)}]^T$ at point (x_1, x_2) can come from data points lying either on (x_1, x_2) or on the regions $R'_0, \ldots, R'_3$. There exist five such possibilities:

1) 1 point lies on (x_1, x_2)
2) 1 point lies in R'_0, 1 point lies in R'_2
3) 1 point lies in R'_1, 1 point lies in R'_2
4) 1 point lies in R'_1, 1 point lies in R'_3
5) 1 point lies in R'_3, 1 point lies in R'_0

If $F'_0, \ldots, F'_3$ denote the probability masses in $R'_0, \ldots, R'_3$, the two-dimensional pdf is given by

$$f_{(r_1, r_2)}(x_1, x_2) = P\{X_{1(r_1)} = x_1, X_{2(r_2)} = x_2\} = \sum_{i=1}^{5} f_i \tag{13}$$

$$f_1 = f(x, y) \sum_{n_0} \frac{n!}{\prod_{i=0}^{3} n_i!} \prod_{i=0}^{3} F_i^{n_i}(x_1, x_2) \tag{14}$$

where

$$\begin{aligned} \sum_{i=0}^{3} n_i &= n - 1 \\ n_0 + n_2 &= r_1 - 1 \\ n_0 + n_1 &= r_2 - 1 \end{aligned} \tag{15}$$

$$f_2 = F'_0 F'_2 \sum_{n_0} \frac{n!}{\prod_{i=0}^{3} n_i!} \prod_{i=0}^{3} F_i^{n_i}(x_1, x_2) \tag{16}$$

where

$$\begin{aligned} \sum_{i=0}^{3} n_i &= n - 2 \\ n_0 + n_1 &= r_2 - 2 \\ n_0 + n_2 &= r_1 - 2 \end{aligned} \tag{17}$$

$$f_3 = F'_2 F'_1 \sum_{n_0} \frac{n!}{\prod_{i=0}^{3} n_i!} \prod_{i=0}^{3} F_i^{n_i}(x_1, x_2) \tag{18}$$

where

$$\begin{aligned} \sum_{i=0}^{3} n_i &= n - 2 \\ n_0 + n_1 &= r_2 - 1 \\ n_0 + n_2 &= r_1 - 2 \end{aligned} \tag{19}$$

$$f_4 = F'_1 F'_3 \sum_{n_0} \frac{n!}{\prod_{i=0}^{3} n_i!} \prod_{i=0}^{3} F_i^{n_i}(x_1, x_2) \tag{20}$$

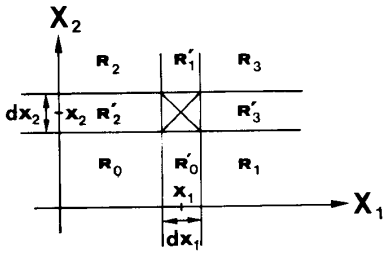


Fig. 1. Regions used in the derivation of the pdf of the two-dimensional marginal order statistics.

where

$$\begin{aligned} \sum_{i=0}^{3} n_i &= n - 2 \\ n_0 + n_1 &= r_2 - 1 \\ n_0 + n_2 &= r_1 - 1 \end{aligned} \tag{21}$$

$$f_5 = F'_3 F'_0 \sum_{n_0} \frac{n!}{\prod_{i=0}^{3} n_i!} \prod_{i=0}^{3} F_i^{n_i}(x_1, x_2) \tag{22}$$

where

$$\begin{aligned} \sum_{i=0}^{3} n_i &= n - 2 \\ n_0 + n_1 &= r_2 - 2 \\ n_0 + n_2 &= r_1 - 1. \end{aligned} \tag{23}$$

The proof of (13)–(23) is included in Appendix B. This approach to the derivation of the pdf of marginal-order statistics can be extended to higher dimensions. However, it is not recommended due to its increasing complexity.

The joint cdf of the p-dimensional order statistics $[X_{1(r_1)}, \ldots, X_{p(r_p)}]^T$ and $[X_{1(s_1)}, \ldots, X_{p(s_p)}]^T$ can be defined as follows:

$$\begin{aligned} &F_{(r_1, \ldots, r_p)(s_1, \ldots, s_p)}(x_1, \ldots, x_p, t_1, \ldots, t_p) \\ &\quad = P\{X_{1(r_1)} \le x_1, \ldots, X_{p(r_p)} \le x_p, X_{1(s_1)} \\ &\quad \le t_1, \ldots, X_{p(s_p)} \le t_p\} \\ &\quad = P\{\text{at least } r_i \ X_{1i} \le x_1, \ldots, \text{at least } r_p \ X_{pi} \le x_p \\ &\qquad \text{at least } s_1 \ X_{1i} \le t_1, \ldots, \text{at least } s_p \ X_{pi} \le t_p\} \\ &\quad = \sum_{j_1=s_1}^{n} \sum_{i_1=r_1}^{j_1} \cdots \sum_{j_p=s_p}^{n} \sum_{i_p=r_p}^{j_p} P\{i_1 \text{ of } X_{1i} \\ &\quad \le x_1, j_1 \text{ of } X_{1i} \le t_1, \ldots, \\ &\qquad i_p \text{ of } X_{pi} \le x_p, j_p \text{ of } X_{pi} \le t_p\}, \\ &\qquad x_i < t_i \qquad i = 1, \ldots, p \end{aligned} \tag{24}$$

for $r_1 < s_1, \ldots, r_p < s_p$. The two points $(x_1, \ldots, x_p)$ and $(t_1, \ldots, t_p)$ define hyperplanes parallel to the p axes that divide the p-dimensional space in 3^p subspaces. If n_i, F_i, and $i = 0, \ldots, 3^p - 1$ denote the number of data points and the probability masses in each subspace, it can be easily

proven along similar lines with the proof in Appendix A that

$$P\{i_1 \text{ of } X_{1i} \le x_1,\ j_1 \text{ of } X_{1i} \le t_1, \ldots, i_p \text{ of } X_{pi} \le x_p,\ j_p \text{ of } X_{pi} \le t_p\} = \sum_{n_0} \cdots \sum_{n_{3^p-1}} \frac{n!}{\prod_{i=0}^{3^p-1} n_i!} \prod_{i=0}^{3^p-1} F_i^{n_i} \tag{25}$$

under the constraints

$$\sum_{i=0}^{3^p-1} n_i = n$$

$$\sum_{I_0=0} n_i = i_1, \ldots, \sum_{I_{p-1}=0} n_i = i_p$$

$$\sum_{I_0=0,1} n_i = j_1, \ldots, \sum_{I_{p-1}=0,1} n_i = j_p \tag{26}$$

where $i = (I_{p-1}, I_{p-2}, \ldots, I_0)_3$ is the arithmetic representation of number i with base 3. Equations (25)-(26) give a numerically tractable way to calculate the joint cdf of the p-dimensional order statistics.

Having derived the probability distributions of the p-dimensional marginal-order statistics, we proceed to their use in multichannel signal filtering.

IV. Multichannel Filters Based on Data Ordering

One-dimensional order statistics (especially the median) and their linear combinations such as the α-trimmed mean and the L estimators have been used extensively as estimators of one-dimensional location [10], [15], [19], [20]. The definitions of L estimators can be easily extended to the p-dimensional case by using marginal-order statistics. Thus, the p-dimensional marginal L estimator is given by [11]:

$$\mathbf{T}_n = \sum_{i_1=1}^{n} \cdots \sum_{i_p=1}^{n} \mathbf{A}_{i_1, \ldots, i_p} \cdot \mathbf{x}_{(i_1, \ldots, i_p)} \tag{27}$$

where $\mathbf{x}_{(i_1, \ldots, i_p)} = [x_{1(i_1)}, \ldots, x_{p(i_p)}]^T$ are the marginal-order statistics and $\mathbf{A}_{i_1, \ldots, i_p}$ are $p \times p$ matrices. The p-dimensional marginal median is a special case of the L estimator (27), for the following choice of matrices $\mathbf{A}_{i_1, \ldots, i_p}$:

$$\mathbf{A}_{i_1, \ldots, i_p} = \mathbf{0} \qquad i_j \neq \nu + 1$$

$$\mathbf{A}_{\nu+1, \ldots, \nu+1} = \begin{bmatrix} 1 & & & 0 \\ & 1 & & \\ & & \ddots & \\ 0 & & & 1 \end{bmatrix} \tag{28}$$

The cdf of the p-dimensional marginal median can be easily derived from (8)-(11) by setting $r_1 = \cdots = r_p = \nu + 1$, for odd data number $n = 2\nu + 1$.

The p-dimensional marginal α-trimmed mean is another special case of the p-dimensional L estimator and is defined as follows [13]:

$$\mathbf{T}_n = \begin{bmatrix} \dfrac{1}{n(1-2\alpha_1)} & \displaystyle\sum_{i=\alpha_1 n+1}^{n-\alpha_1 n} x_{1(i)} \\ \dfrac{1}{n(1-2\alpha_p)} & \displaystyle\sum_{i=\alpha_p n+1}^{p} x_{p(i)} \end{bmatrix}. \tag{29}$$

Robustness measures, e.g., influence function, gross-error sensitivity [20] of the L estimators in the presence of outlying observations were presented in [11]. The influence function of the two-dimensional marginal median was also derived. It was found that the marginal median is a B-robust estimator because its gross error sensitivity is finite [11].

Another definition of the multichannel median filter has been proposed for color image filtering [12], [21]. It is called vector median filter and its output is the vector that minimizes the L_2 error norm:

$$\sum_{i=1}^{n} |\mathbf{x}_i - \mathbf{x}_{\text{med}}| \tag{30}$$

where $|\ |$ denotes the Euclidean distance. The vector $\mathbf{x}_{\text{med}}$ may or may not be one of the data $\mathbf{x}_i$, $i = 1, \ldots, n$. The vector median (30) is related to reduced ordering rather than to marginal ordering since it also uses distance measures. Its implicit definition makes difficult its use in color image filtering because an iterative optimization algorithm is required to produce each output image pixel. The computational load is considerably reduced by forcing the median $\mathbf{x}_{\text{med}}$ to belong to the set $\{\mathbf{x}_i,\ i = 1, \ldots, n\}$. In this case, the L_2 norm of (30) is computed for x_{med} equal to every $\mathbf{x}_i$, $i = 1, \ldots, n$. The median is the data sample that minimizes (30). Even in this case the computational complexity of the vector median filter is relatively high because we have to calculate $n(n-1)/2$ p-dimensional norms of the form $|\mathbf{x}_i - \mathbf{x}_j|$ and to perform $[n(n-1)/2] - 1$ comparisons. The computation of the marginal median requires only $O(pn \log n)$ comparisons if the QUICKSORT algorithm is used for its computation on each channel [10]. The computational complexity of the vector median filter is relatively small for $n = 3$. In this case the vector median is given by:

$$\mathbf{x}_{\text{med}} = \begin{cases} \mathbf{x}_1 & \text{if } |\mathbf{x}_2 - \mathbf{x}_3| \ge |\mathbf{x}_1 - \mathbf{x}_2| \\ & \quad \text{and } |\mathbf{x}_2 - \mathbf{x}_3| \ge |\mathbf{x}_1 - \mathbf{x}_3| \\ \mathbf{x}_2 & \text{if } |\mathbf{x}_1 - \mathbf{x}_3| \ge |\mathbf{x}_2 - \mathbf{x}_1| \\ & \quad \text{and } |\mathbf{x}_1 - \mathbf{x}_3| \ge |\mathbf{x}_2 - \mathbf{x}_3| \\ \mathbf{x}_3 & \text{if } |\mathbf{x}_1 - \mathbf{x}_2| \ge |\mathbf{x}_3 - \mathbf{x}_1| \\ & \quad \text{and } |\mathbf{x}_1 - \mathbf{x}_2| \ge |\mathbf{x}_3 - \mathbf{x}_2| \end{cases}. \tag{31}$$

Only the calculation of three p-dimensional norms, as well as two comparisons, are required [21].

The definition of the vector median (30) is equivalent to the definition of the marginal median only for the one-dimensional case. Generally, marginal and vector medians are different for dimensions $p \ge 2$. However, the marginal and the vector median tend to have approximately the same values and similar behavior for multivariate data that are strongly correlated among their different dimensions. For example, if all multivariate data samples lie on a line on the p-dimensional space, multivariate ordering is reduced to one-dimensional ordering. In this case both marginal median and vector median have the same value. For illustration purposes, we shall consider the two dimensional case in more detail for data $\mathbf{x} = [x_1, x_2]^T$ having Gaussian distribution with probability density function $N(\mathbf{m}, \mathbf{C}) = N(m_1, m_2, r,$

σ_1, σ_2) given by

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{(1-r^2)}} \cdot \exp\left\{-\frac{1}{2(1-r^2)}\left[\frac{(x_1-m_1)^2}{\sigma_1^2} - 2r\frac{(x_1-m_1)(x_2-m_2)}{\sigma_1\sigma_2} + \frac{(x_2-m_2)^2}{\sigma_2^2}\right]\right\} \quad (32)$$

m_1, m_2, and σ_1, σ_2 are the expected values and standard deviations along the two dimensions. Variable r is the correlation coefficient. We assume that $\sigma_1 = \sigma_2 = \sigma$ In this case the marginal median and the vector median have equal values if the data points $\mathbf{x}_i$ $1 \le i \le n$ satisfy $x_{1i} = -x_{2i}$ (or $x_{1i} = -x_{2i}$) for $1 \le i \le n$. As it is proven in Appendix C, the probability that the random variable x_1 is in the neighborhood of x_2 or $-x_2$, when $\sigma_1 = \sigma_2 = \sigma$, is proportional to the quantity

$$\Pr\{x_1 \cong x_2 \text{ or } x_1 \cong -x_2\} \propto \frac{\sqrt{\pi}\cdot\sigma\cdot\left[\sqrt{(1+r)} + \sqrt{(1-r)}\right] - 1}{2\pi\cdot\sigma^2\cdot\sqrt{1-r^2}}. \quad (33)$$

Fig. 2 shows the above quantity when the correlation coefficient r varies between -1 and 1 and for $\sigma = 1, 5$. As expected, the probability $\Pr\{x_1 \cong x_2 \text{ or } x_1 \cong -x_2\}$ attains its maximum when the correlation coefficient reaches ± 1, and it is larger for small values of the standard deviation σ. The probability that all data points satisfy $x_{1i} \cong x_{2i}$, $i = 1, \ldots, n$ attains a maximum for $r \to 1$:

$$\Pr\{x_{1i} \cong x_{2i},\ i = 1, \ldots, n\} \propto \frac{1}{\left(2\sigma\sqrt{\pi}\sqrt{(1-r)}\right)^n} \quad (34)$$

Similarly, the probability that all data points satisfy $x_{1i} \cong -x_{2i}$, $i = 1, \ldots, n$ attains a maximum for $r \to -1$:

$$\Pr\{x_{1i} \cong -x_{2i},\ i = 1, \ldots, n\} \propto \frac{1}{\left(2\sigma\sqrt{\pi}\sqrt{(1+r)}\right)^n}.$$

In conclusion, we can say that the performances of the two medians are nearly identical when the correlation among the two channels is large ($r \to \pm 1$) and the standard deviation of noise takes small values and it is equal on both channels. This conclusion will explain some simulation results shown in Section VI.

Both marginal median and vector median can be used as filters of multichannel images. Their performance will be assessed in subsequent sections. Another important class of nonlinear filters for image processing can be introduced as a compromise between the median and the moving average filter. These filters are the trimmed mean filters [10], [22]. An extension of the α-trimmed mean filter [22] to the multichannel case is given in (29) by using the marginal-order statistics. Lee and Kassam [23] have introduced a modification of the trimmed mean filter. They defined the modified trimmed mean filter (MTM), which computes the sample median in a window, determines an interval $[\text{med}(x_i) - q, \text{med}(x_i) + q]$ for an appropriately chosen value q, and finally averages only the samples lying inside the filter window and having values in this interval. Thus, the MTM filter can be thought of as a data-dependent version of the α-trimmed mean filter. For small values of q the MTM filter behaves as the median, whereas for large values of q its behavior approaches that of the moving average. Lee and Kassam have also introduced the double-window MTM filter (DW-MTM) as an extension of the MTM filters [23], [24]. This filter uses two windows of different size. First, the sample median is computed in the small window and then only those samples of the large window close to the sample median are averaged to give the filter output. A DW-MTM filter suppresses additive Gaussian noise effectively because it uses a fairly large window for the averaging. It preserves image edges because it rejects pixels having values far away from the median of the small window. The extension of the MTM and DW-MTM filters to the multichannel case can be achieved by using the reduced ordering. The multichannel MTM filter is defined as follows. First, the marginal median inside each window is determined along each channel and the vector of the medians $\mathbf{x}_{\text{med}}$ is formed. Then, the input vectors $\mathbf{x}_i$ are ordered by means of their generalized distance $d_i = (\mathbf{x}_i - \mathbf{x}_{\text{med}})^{\mathbf{T}} \cdot \Gamma^{-1} \cdot (\mathbf{x}_i - \mathbf{x}_{\text{med}})$. Vectors lying in the filter window and having generalized distances d_i smaller than a threshold d are averaged and form the filter output. The threshold value d may be selected as the average of the generalized distances d_i:

$$\mathbf{y}_{ij} = \frac{\Sigma_A \Sigma a_{rs}\mathbf{x}_{i+r,\,j+s}}{\Sigma_A \Sigma a_{rs}}$$

$$a_{rs} = \begin{cases} 1 & (\mathbf{x}_i - \mathbf{x}_{\text{med}})^{\mathbf{T}} \cdot \Gamma^{-1} \cdot (\mathbf{x}_i - \mathbf{x}_{\text{med}}) < d \\ 0 & \text{otherwise} \end{cases}. \quad (35)$$

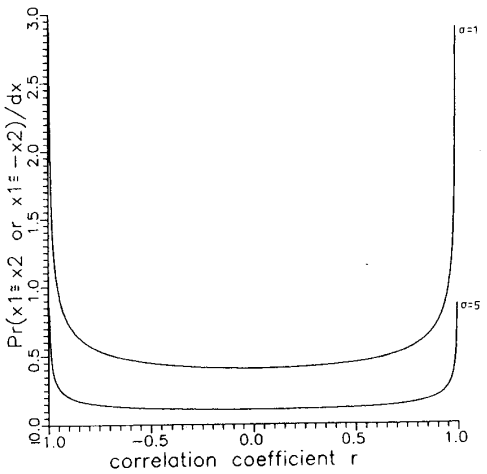


Fig. 2. Probability that the coordinates of a two-dimensional random variable are equal or opposite in value as a function of the correlation coefficient r for two-dimensional Gaussian distribution having $\sigma_1 = \sigma_2 = \sigma$.

A similar generalization can be done in order to define the multichannel DW-MTM filter. As in the single-channel case, two window sizes are used. The marginal median is computed first inside the small window. Then, the data vectors lying in the large window and having generalized distances d_i less than some value d are averaged to give the filter output.

The matrix Γ in the expression of the generalized distance d_i may be the identity matrix $\mathbf{I}$ or the sample dispersion matrix Σ. If we use $\Gamma = \mathbf{I}$, the data vectors are ordered according to their Euclidean distance from $\mathbf{x}_{(\nu+1)}$. However, it is well known that color distances are not Euclidean in the Commission International de l'Eclairage (CIE) primary system red-green-blue (RGB) [1]. Color distances are approximated by the Euclidean distance in the other color coordinate systems, e.g., in the modified universal camera site (UCS) system and in the L^*, a^*, b^* system [1]. Thus it is reasonable to perform DW-MTM filtering on a transformed image. Such an approach has been investigated by simulations. If the matrix $\Gamma = \Sigma$ is used, the data are ordered according to their Mahalanobis distance from $\mathbf{x}_{\text{med}}$. The Mahalanobis distance takes into account the relative dispersions and correlations among the channels of the multichannel signal. Furthermore, for the family of the elliptically symmetric distributions, the Mahalonobis distance can be interpreted as a probabilistic distance where equal distances imply equal likelihoods.

V. Performance of Marginal Median and Vector Median Filters in the Presence of Impulsive Noise

It is well known that median filters have very good performance in impulsive noise filtering in the single-channel case [10]. Therefore it is desirable to explore the performance of median filters in multichannel impulsive noise filtering. Multichannel impulsive noise occurs in color image transmission through noisy telecommunication channels, e.g., in TV image transmission. A serious problem in multichannel impulsive noise filtering is the lack of appropriate probability model for this type of noise. We extended the model of the single-channel impulsive noise to the multichannel case to solve this problem. In this section we study the performance of marginal median and vector median filters in the presence of impulsive noise. We focus our analysis on the two-channel case because the noise model is relatively simple and the theoretical analysis of the performance of the median filters is tractable. The noise model used is the following:

$$\mathbf{x}_i = \begin{cases} s_i & \text{with probability } 1 - p \\ (d, s_{2i}) & \text{with probability } p_1 \cdot p \\ (s_{1i}, d) & \text{with probability } p_2 \cdot p \\ (d, d) & \text{with probability } p \cdot (1 - p_1 - p_2) \end{cases} \tag{36}$$

where $\mathbf{s}_i = [s_{1i}, s_{2i}]$ is the noiseless, constant signal, d is the impulse value, and $\mathbf{x}_i = [x_{1i}, x_{2i}]$ is the noisy signal. Equation (36) describes impulsive noise having a single type of impulses (positive or negative) that depends on the impulse value d. We assume that $d \gg s_{1i}, s_{2i}$. Impulses may corrupt either channel with probabilities pp_1 and pp_2, respectively, or both of them with probability pp_3 where $p_3 = (1 - p_1 - p_2)$.

The derivation of a general expression for the probability of correct reconstruction in the case of a filter of arbitrary length is very difficult. Therefore, we study filters having a small window sizes that are usually encountered in practice. In the following, we derive the rate of failure for vector and marginal median filters having length $n = 3$ and we give the expressions for lengths $n = 5, 9$. The proof for filter length $n = 5$ can be found in Appendix C. We use the enumeration of the favorable events method.

The probability of correct reconstruction P_g is calculated as the sum of the probabilities of all combinations of favorable input vectors. Favorable input vectors are the ones for which the filters produce the noiseless signal vector $\mathbf{s} = [s_1, s_2]$ as output. The vector median filter gives $\mathbf{s}$ as output, if s minimizes the norm (30) On the other hand, the marginal median filter gives $\mathbf{s}$ as output, if in each channel there are at least $\nu + 1$ points of value s, where $n = 2\nu + 1$ is the filter length. In the following, we give the favorable input vectors and their probability of occurrence of filter length $n = 3$. For this filter length the analysis is identical for both filters:

1) Input vectors $[s_1, s_2]$, $[s_1, d]$, $[d, s_2]$. Their probability of occurrence is given by

$$P_{g1} = \frac{3!}{1!1!1!} \cdot (1 - p) \cdot pp_2 \cdot pp_1 \tag{37}$$

2) Two vectors of the form $[s_1, s_2]$ and one of the form $[s, d]$ $[d, s_2]$, or $[d, d]$ with probability:

$$P_{g2} = \frac{3!}{2!1!} \cdot (1 - p)^2 \cdot p \tag{38}$$

3) Three vectors of the form $[s_1, s_2]$ with probability:

$$P_{g3} = (1 - p)^3. \tag{39}$$

Thus, the probability of correct reconstruction of the two-channel signal is given by:

$$P_{g\,\text{vector}}^{n=3} = P_{g\,\text{marginal}}^{n=3} = 6p^2(1 - p)p_1p_2 + 3p(1 - p)^2 + (1 - p)^3. \tag{40}$$

The following expressions are found for filter lengths $n = 5, 9$ with similar reasoning:

$$P_{g\,\text{vector}}^{n=5} = 15p^3(1 - p)^2[2p_1p_2(1 + p_3) + (1 - p_3)p_3^2] + 10p^2(1 - p)^3 + 5p(1 - p)^4 + (1 - p)^5 \tag{41a}$$

$$P_{g\,\text{marginal}}^{n=5} = 30p^4(1 - p)p_1^2p_2^2 + 30p^3(1 - p)^2p_1p_2(1 + p_3) + 10p^2(1 - p)^3 + 5p(1 - p)^4 + (1 - p)^5 \tag{41b}$$

$$
\begin{aligned}
P_{g\ \text{vector}}^{n=9} = {} & 840p^6(1-p)^3[9p_1^2p_2^2p_3^2 + 6p_1^2p_2^2(p_1+p_2) \\
& \cdot p_3 + 2p_1^3p_2^3 + 3p_1p_2(p_1+p_2)p_3^3] \\
& + 315p^5(1-p)^4[2p_1p_2(p_1^3+p_2^3) \\
& + 4p_1^2p_2^2(p_1+p_2) + 8p_1p_2(p_1^2+p_2^2)p_3 \\
& + 4(p_1^3+p_2^3)p_3^2 + 12p_1^2p_2^2p_3 \\
& + 12p_1p_2(p_1+p_2)p_3^2 + 4(p_1^2+p_2^2)p_3^3 \\
& + 8p_1p_2p_3^3 + (p_1+p_2)p_3^4] + 126p^4 \\
& \cdot(1-p)^5 + 84p^3(1-p)^6 \\
& + 36p^2(1-p)^7 + 9p(1-p)^8 + (1-p)^9
\end{aligned} \tag{42a}
$$

$$
\begin{aligned}
P_{g\ \text{marginal}}^{n=9} = {} & 630p^8(1-p)p_1^4p_2^4 + 1260p^7(1-p)^2 \\
& \cdot p_1^3p_2^3(1+3p_3) + 840p^6(1-p)^3 \\
& \cdot[9p_1^2p_2^2p_3^2 + 6p_1^2p_2^2(p_1+p_2)p_3 + 2p_1^3p_2^3] \\
& + 315p^5(1-p)^4[2p_1p_2(p_1^3+p_2^3) \\
& + 4p_1^2p_2^2(p_1+p_2) + 8p_1p_2(p_1^2+p_2^2)p_3 \\
& + 12p_1^2p_2^2p_3 + 12p_1p_2(p_1+p_2)p_3^2 \\
& + 8p_1p_2p_3^3] \\
& + 126p^4(1-p)^5 \\
& + 84p^3(1-p)^6 + 36p^2(1-p)^7 \\
& + 9p(1-p)^8 + (1-p)^9.
\end{aligned} \tag{42b}
$$

The rate of failure, i.e., the probability of erroneous reconstruction

$$R = 1 - P_g$$

is displayed in Fig. 3 as a function of p for filter length $n = 3$ and for various probabilities p_1, p_2. It is observed that impulses appearing simultaneously on both channels ($p_1 = p_2 = 0$) produce higher failure rates than impulses appearing on just one channel. However, it has been found in simulations that the dependence of the rate of failure on p_1, p_2 is not very strong when p_1 and p_2 are different from zero for filter length $n = 9$. As expected, large filter windows improve the performance of both filters in impulsive noise removal. Filter windows having size 3×3 ($n = 9$) can effectively remove impulses having probability of occurrence approximately up to 0.3.

It is of interest to compare the performances of the marginal median and vector median filters in impulsive noise removal. For $p_1 = p_2 = 0$, the performance of the two filters is identical regardless of the filter length, according to the theoretical analysis. Fig. 4 shows the rate of failure of the two median filters for window size $n = 5$ and for three values of $p_1 = p_2$ (0.33, 0.4, 0.5). We observe that the vector median performs better than the marginal median, as the probability p_3 that both channels are corrupted by impulses rises. However, in general, the marginal median behaves better than the vector median filter for large impulse probability p. It has also been observed that the vector median filter performs slightly better than the marginal median filter in the typical case when we have impulse probability $p \leq 0.5$, equal probabilities $p_1 = p_2 = 0.33$, and filter length $n < 9$.

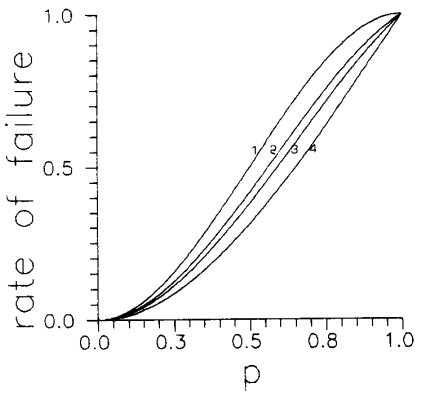


Fig. 3. Rate of failure of the marginal median filter for filter length $n = 3$, and probabilities $p_1, p_2 = 0$ (curve 1), $p_1, p_2 = 0.33$ (curve 2), $p_1, p_2 = 0.4$ (curve 3), and $p_1, p_2 = 0.5$ (curve 4).

VI. Simulation Examples

In this section, we present two sets of experiments in order to assess the performance of the filters we have discussed so far. In the first set of experiments, a two-channel one-dimensional input signal has been used. The following filters were tested: 1) the moving average filter, 2) the marginal median, 3) the vector median, 4) the α-trimmed mean, and 5) the multichannel MTM and DW-MTM filters. Two input sample distributions have been used: the two-dimensional Gaussian distribution $N(\mathbf{m}, \mathbf{C})$ (32), as well as the so-called contaminated Gaussian distribution:

$$f(\mathbf{x}) = (1-\epsilon) \cdot N(\mathbf{x}, \mathbf{m}, \mathbf{C}) + \epsilon \cdot g(\mathbf{x}). \tag{43}$$

Variable ϵ is the percentage of the contamination (outliers) and $g(\mathbf{x})$ is the distribution of the outliers. The performance criterion used is the standard deviation of the output in each channel:

$$s_{oi} = \left\{ \frac{1}{N-1} \sum_{j=1}^{N} (x_{ji} - \bar{x}_i)^2 \right\}^{1/2} \qquad i = 1, 2 \tag{44}$$

where $\bar{x}_i = (1/N)\sum_{j=1}^{N} x_{ji}$ is the output arithmetic mean. The smaller the output deviation is, the better is the performance of the filter. The lengths of the filters tested were $n = 5, 11$. The parameter α of the α-trimmed mean filter was chosen $\alpha = 0.2$. In the case of the DW-MTM multichannel, we have performed the experiments with windows having sizes 5, 7, and 11, 15, respectively.

First, we examine input distributions of the form $N(m_1 = 0, m_2 = 0, r, \sigma_1 = 1, \sigma_2 = 1)$ for correlation coefficient r varying in the range $0 \leq r \leq 1$. The output standard deviation of the first channel is shown in Fig. 5 as a function of the correlation coefficient r for filter size $n = 5$. As expected for

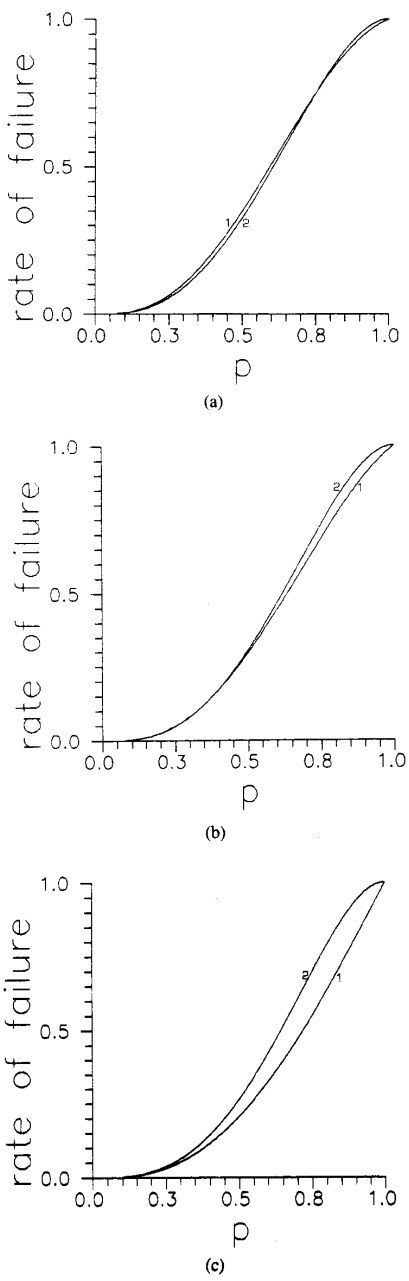


Fig. 4. Comparison of the rate of failure of the marginal median filter (curve 1) and of the vector median filter (curve 2) for filter length $n = 5$. (a) $p1 = p2 = .333$; (b) $p1 = p2 = 0.4$; (c) $p1 = p2 = 0.5$.

the Gaussian noise, the moving average filter has the smallest output deviation and thus performs better than any other filter. The α-trimmed mean is slightly inferior to the moving average filter. The marginal median and DW-MTM filters have similar performance and are inferior to both the moving average and the α-trimmed mean filter. Vector median and MTM filter have the worst performance. However, the vector median has better performance than the MTM filter for $r \geq 0.5$. The performances of all filters but the vector median filter are independent of the correlation between channels. When the correlation coefficient r tends to one $r \to 1$ the output variances of the marginal median and vector median filters are identical. This fact follows from the analysis given in Section IV. Similar observations can be made for the output of the second channel. The results are similar for negative correlation coefficients as well as for large filter windows. The only difference is that the output variance decreases with increasing filter length.

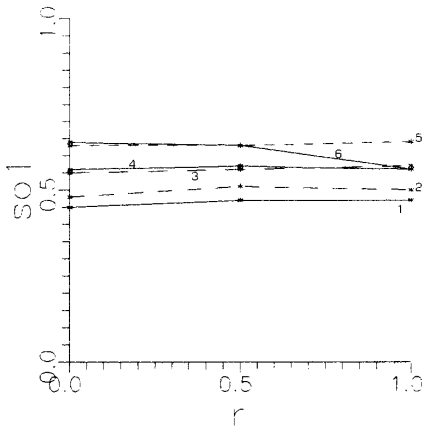


Fig. 5. Output standard deviation of the first channel as a function of the correlation coefficient for filter length $n = 5$. Curves 1-6 correspond to the moving average filter (1), α-trimmed mean filter (2), DW-MTM filter (3), marginal median filter (4), MTM filter (5), and vector median filter (6), respectively. The DW-MTM filter has window sizes 5 and 7.

This set of experiments is continued with input signals having contaminated Gaussian distributions of the form:

$$f(\mathbf{x}) = (1 - \epsilon) \cdot N(0, 0, 0, 1, 1) + \epsilon \cdot N(0, 0, 0, \sigma, \sigma) \quad (45)$$

with varying percentage ϵ and equal standard deviations $\sigma_1 = \sigma_2 = \sigma$ of the outliers. The output standard deviation of the first channel is shown as a function of the percentage ϵ in Fig. 6 for filter length $n = 5$ and outlier standard deviation $\sigma_1 = \sigma_2 = 4$. It is also shown as a function of the outlier standard deviation $\sigma_1 = \sigma_2 = \sigma$ in Fig. 7. In both cases the moving average filter has the worst performance. The second worst is the vector median filter. The performance of the DW-MTM filter is the best. Furthermore, it is nearly immune to the increase of the percentage ϵ and deviation σ of the contamination. This simulation verifies the robustness of the DW-MTM filter in the presence of outliers. The α-trimmed mean filter has also good performance in the case of small percentage ϵ or small deviation σ for trimming factor $\alpha = 0.2$. Its performance can be improved for large ϵ or σ by increasing the trimming factor α.

The above observations are clearer when the outliers are impulses of value ± 10, whereas the data distribution is Gaussian $N(0, 0, 0, 1, 1)$. The output variance is shown in Fig. 8 as a function of the outlier percentage ϵ for filter

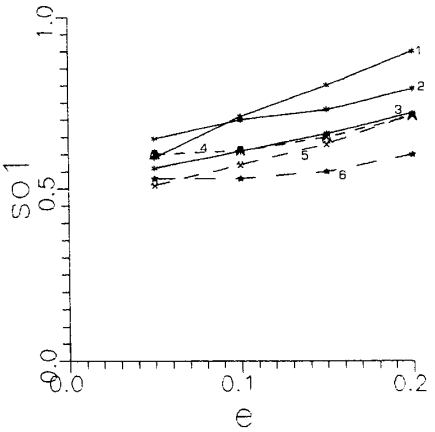


Fig. 6. Output standard deviation of the first channel as a function of the outlier percentage for filter length $n = 5$. Curves 1-6 correspond to the moving average filter (1), vector median filter (2), marginal median filter (3), MTM filter (4), α-trimmed mean filter (5), and DW-MTM filter (6), respectively. The DW-MTM filter has window sizes 5 and 7.

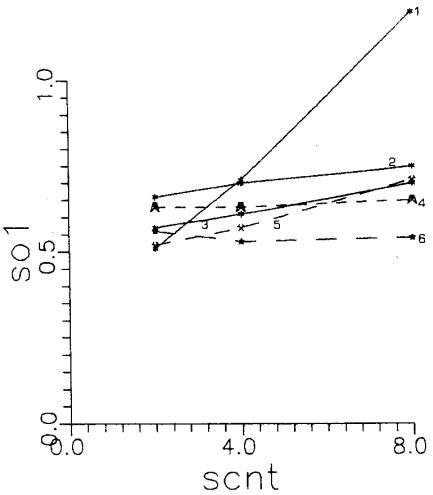


Fig. 7. Output standard deviation of the first channel as a function of the outlier standard deviation for filter length $n = 5$ ($\epsilon = .1$). Curves 1-6 correspond to the moving average filter (1), vector median filter (2), marginal median filter (3), MTM filter (4), α-trimmed mean filter (5), and DW-MTM filter (6), respectively. The DW-MTM filter has window sizes 5 and 7.

length $n = 5$. The DW-MTM filter has the best performance.

The second set of experiments has been performed on noisy color images. The quantitative criterion that we used was the enhancement factor F for color images defined as follows:

$$F = \frac{\sum_k \sum_l |\mathbf{u}(k,l)\mathbf{x}(k,l)|^2}{\sum_k \sum_l |\mathbf{u}(k,l) - \mathbf{y}(k,l)|^2} \tag{46}$$

where $\mathbf{u}(k,l)$ is the noise-free image, $\mathbf{x}(k,l)$ is the noisy input image, and $\mathbf{y}(k,l)$ is the filtered color image vector. The values in decibels of F for different filters and noise models are given in Table I.

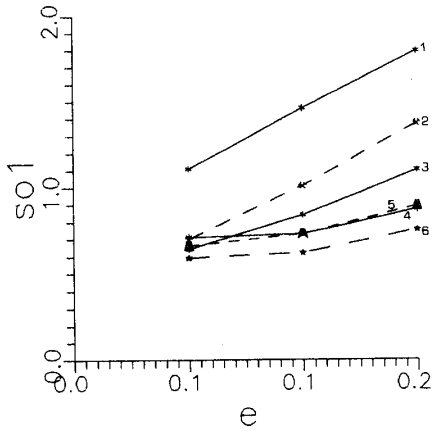


Fig. 8. Output standard deviation of the first channel as a function of the outlier percentage ϵ for filter length $n = 5$. Outliers are impulses of magnitude ± 10. Curves 1-6 correspond to the moving average filter (1), α-trimmed mean filter (2), marginal median filter (3), vector median (4), MTM filter (5), and DW-MTM filter (6), respectively. The DW-MTM filter has window sizes 5 and 7.

TABLE I
ENHANCEMENT FACTOR F (IN DECIBELS) FOR DIFFERENT FILTERS AND NOISE MODELS

The test image was corrupted by 1) additive, zero mean, Gaussian noise having standard deviation $\sigma = 20$ (first column), 2) impulsive noise where 5% of the image pixels in every channel were contaminated by impulses of value 0 or 255 (second column), 3) mixed Gaussian ($\sigma = 30$) and impulsive noise (3% of the pixels contaminated) (third column).

	Noise Model		
Filter	Gaussian Noise ($\sigma = 20$)	Impulsive Noise	Gaussian ($\sigma = 30$) and Impulsive Noise
Mean 3×3	5.403	7.185	7.087
Marginal median 3×3	4.967	12.633	7.714
Vector median 3×3	3.316	11.507	5.638
α-Tm 3×3, $\alpha = 0.2$	5.479	10.767	8.167
DW-MTM $^{5w=3\times3}_{1w=5\times5}$	5.551	10.071	8.642

The test image ("Lenna") having size 256×256 pixels is shown in Fig. 9(a). The filters used have a window size of 3×3. The large window of the DW-MTM filter has dimension 5×5. The test image corrupted by additive zero-mean white Gaussian noise having variance $\sigma^2 = 400$ on each channel is shown in Fig. 9(b). The output of the moving average filter and of the marginal median filter is shown in Fig. 9(c) and 9(d), respectively. As expected, the moving average filter performs better than the median filters in the homogeneous image regions, whereas the marginal median filter preserves edge information. Streaking and blotching effects are more visible in the vector median filter output, as seen in Fig. 9(e). This behavior is partially explained by the fact that the vector median output is equal to one of the samples in the filter window. Thus, the probability to have equal outputs in adjacent window positions is higher for the vector median than for the marginal median filter. The DW-MTM filter has the best performance, as shown in Fig.

Fig. 9. (a) Test image. (b) Image corrupted by multichannel Gaussian noise. (c) Output of a 3×3 moving average filter. (d) Output of a 3×3 marginal median filter. (e) Output of a 3×3 vector median filter. (f) Output of a 3×3, 5×5 DW-MTM filter. (For color supplement see p. 295.)

9(f). The above subjective observations are in accordance with the measured values of the enhancement factor given in the first column of Table I.

The test image corrupted by impulses having value 0 or 255 on all three channels, with probability of occurrence $p = 0.05$, is shown in Fig. 10(a). We can observe in Fig. 10(b) and (c) that the marginal median filter rejects the impulses slightly better than the vector median filter. Furthermore, the vector median filter has more intense streaking and blotching effects.

Finally, the original image corrupted by additive zero-mean white Gaussian noise having variance $\sigma^2 = 900$, mixed with impulses with probability of occurrence $p = .03$ is shown in Fig. 11(a). The superiority of the DW-MTM (18) filter over the marginal median filter is clear, as can be seen in Fig. 11(b) and (c). The DW-MTM filter has also been applied directly on the RGB channels. Alternatively it has been applied on the XYZ and on the modified (USC) channels after suitable color transformations [1] to check its sensitivity on various color distances. However, the results were very similar on all color coordinate systems used.

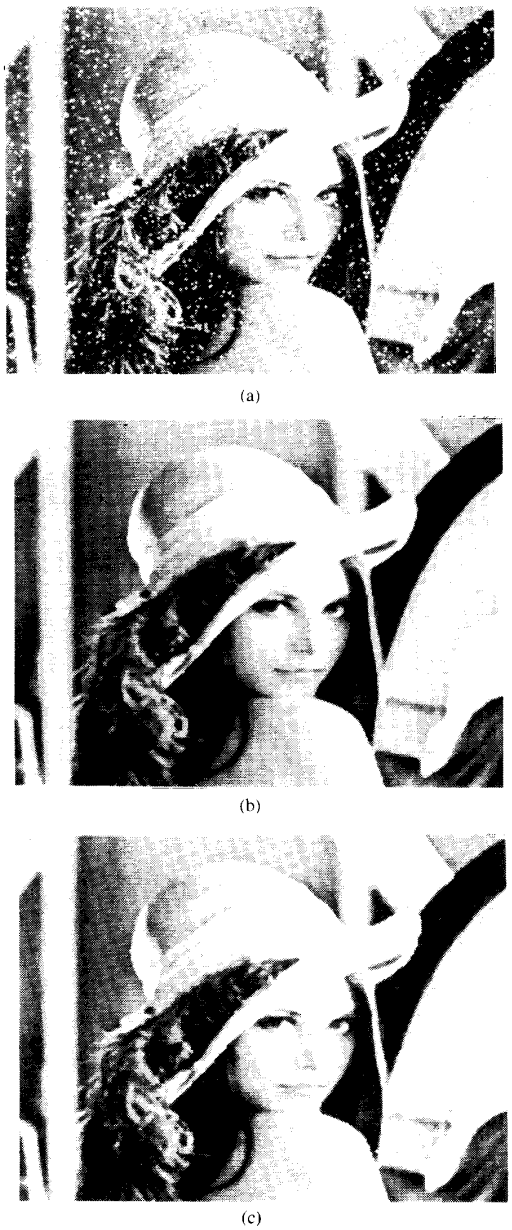


Fig. 10. (a) Test image corrupted by impulsive noise. (b) Output of a 3×3 marginal median filter. (c) Output of a 3×3 vector median filter. (For color supplement see p. 296.)

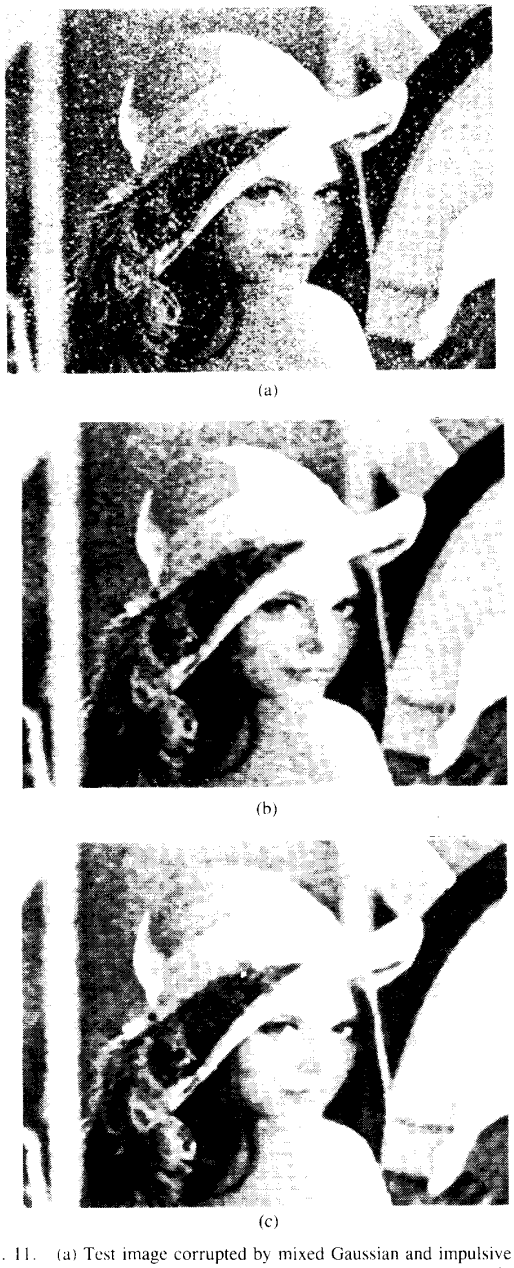


Fig. 11. (a) Test image corrupted by mixed Gaussian and impulsive noise. (b) Output of a 3×3 marginal median filter. (c) Output of a 3×3, 5×5 DW-MTM filter.(For color supplement see p. 296.)

VII. Conclusions

Multichannel nonlinear filters based on order statistics have been discussed in this paper. Two major ordering schemes have been used: marginal ordering and reduced ordering. The properties of a class of multichannel filters based on these ordering techniques has been studied. The marginal median filter, the vector median filter, the multichannel α-trimmed mean filter, and the multichannel DW-MTM filters have been analyzed. It has been found that vector median filters and marginal median filters have similar performances despite the fact that they have different definitions. If we take into account the fact that the vector median filter has higher computational complexity, we can conclude that marginal median filter is preferable in most practical systems. Both vector and marginal median filters are inferior to the α-trimmed mean filter, the MTM filter, and the DW-MTM filter in additive Gaussian noise removal and in contaminated Gaussian noise removal. The DW-MTM filter has clearly the best performance in the simulations both for one-dimensional signals and for color images. Its superiority can be explained partially by the fact that it uses a larger filter window. However, the price paid for its performance is its increased computational complexity.

APPENDIX A

Each point $(x_1, \ldots, x_p)$ defines p hyperplanes, each one parallel to one of the p axes. These hyperplanes divide the p-dimensional space in 2^p subspaces. Let us denote by $F_i(x_1, \ldots, x_p)$ the probability masses of the input pdf in those subspaces. The p-dimensional cdf of the marginal-order statistics $[X_{1(r1)}, \ldots, X_{p(rp)}]^T$ is given by (7). Let n_i, $i = 0, \ldots, 2^p - 1$, denote the number of data points belonging to each of the 2^p subspaces. In this case:

$$P\{i_1 \text{ of } X_{1i} \le x_1, \ldots, i_p \text{ of } x_{pi} \le x_p\} = \sum_{n_0} \cdots \sum_{n_{2^p-1}} \frac{n!}{\prod_{i=0}^{2^p-1} n_i!} \prod_{i=0}^{2^p-1} F_i^{n_i}(x_1, \ldots, x_p). \quad \text{(A1)}$$

The number of data points must be nonnegative. Therefore,

$$n_i \ge 0 \qquad i = 0, \ldots, 2^p - 1. \quad \text{(A2)}$$

The total number of data points is n. Thus, the sum of the data points on every subspace satisfies

$$\sum_{i=0}^{2^p-1} n_i = n. \quad \text{(A3)}$$

Let $(j_{p-1}, \ldots, j_0)_2$ be the binary representation of a number j, $0 \le j \le 2^p - 1$. The label of each region R_j, $j = 0, \ldots, 2^p - 1$, can be chosen in such a way so that all regions R_j having indices $j_0 = 0$ correspond to variables $\mathbf{X}_i = [X_{1i}, \ldots, X_{pi}]^T$ satisfying $X_{1i} \le x_1$, all regions R_j having $j_1 = 0$ correspond to variables $\mathbf{x}$ satisfying X_2 x_2, etc. The total number of variables in all regions having indices $\mathbf{j}$ such that $j_k = 0$, $1 \le k \le p$ is i_k. Therefore, the following conditions must be imposed on the summation indices in (8):

$$\sum_{j_0=0} n_j = i_1, \ldots, \sum_{j_{p-1}=0} n_j = i_p. \quad \text{(A4)}$$

Thus, the cdf for the p-dimensional case is given by

$$F_{(r_1, \ldots, r_p)}(x_1, \ldots, x_p) = \sum_{i_1=r_1}^{n} \cdots \sum_{i_p=r_p}^{n} \sum_{n_0} \cdots \sum_{n_{2^p-1}} \frac{n!}{\prod_{i=0}^{2^p-1} n_i!} \prod_{i=0}^{2^p-1} F_i^{n_i}(x_1, \ldots, x_p) \quad \text{(A5)}$$

subject to the constraints (A2)–(A4).

APPENDIX B

Let us suppose that one data point exists on (x_1, x_2) and $n_0, \ldots, n_3$ are the data points in $R_0 \cup R_0' \cup R_2'$, $R_1' \cup R_3'$, and $R_2' \cup R_1'$, R_3, respectively. The total number of data is n. The probability of such a layout is given by

$$f_1 = f(x, y) \sum_{n_0} \cdots \sum_{n_3} \frac{n!}{\prod_{i=0}^{3} n_i!} \prod_{i=0}^{3} F_i^{n_i}(x_1, x_2).$$

Since the total number of data points is n, the following constraint must be satisfied:

$$\sum_{i=0}^{3} n_i = n - 1.$$

The total number of points in the regions $R_0 \cup R_0' \cup R_2'$ and $R_2 \cup R_1'$ must be $r_1 - 1$. Therefore,

$$n_0 + n_2 = r_1 - 1.$$

Similarly,

$$n_0 + n_1 = r_2 - 1.$$

Thus, we have only one free variable and (14) and (15) have been proven.

Let us suppose that the marginal-order statistic comes from one data point in R_0' and one data point in R_2'. The probability of such a configuration is given by (16). Since the total number of data points is n, the first constraint in (17) must be valid. The total number of points in the regions $R_0 \cup R_0' \cup R_2' \cup R_2 \cup R_1'$ and $R_0 \cup R_0' \cup R_2' \cup R_1 \cup R_3'$ must be $r_2 - 2$ and $r_1 - 2$, respectively. Thus, the rest two constraints in (17) must also be valid. Similar proofs can be found for (18)–(23).

APPENDIX C

The probability that the random variable x_1 is equal or opposite to the random variable x_2 can be expressed as

$$Pr\{x_1 = x_2 \text{ or } x_1 = -x_2\} = Pr\{x_1 = x_2\} + Pr\{x_1 = x_2\} - Pr\{x_1 = x_2 = 0\}. \quad \text{(C1)}$$

Assuming that the random variables follow the Gaussian distribution (32) we get:

$$\Pr\{x_1 = x_2 = x\} = \lim_{\Delta x \to 0} \left[\int f(x, x) \cdot dx\right] \Delta x^2 = \lim_{\Delta x \to 0} \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{(\sigma_1^2 + \sigma_2^2 - 2r\sigma_1\sigma_2)}} \Delta x^2$$

$$\Pr\{x_1 = x_2 = -x\} = \lim_{\Delta x \to 0} \left[\int f(x, -x) \cdot dx\right] \Delta x^2 = \lim_{\Delta x \to 0} \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{(\sigma_1^2 + \sigma_2^2 + 2r\sigma_1\sigma_2)}} \Delta x^2$$

and

$$\Pr\{x_1 = x_2 = 0\} = \lim_{\Delta x \to 0} f(0, 0) \Delta x^2 = \lim_{\Delta x \to 0} \frac{1}{2\pi\sigma_1\sigma_2\sqrt{(1 - r^2)}} \Delta x^2.$$

By substituting the above probabilities into (C1) and setting $\sigma_1 = \sigma_2 = \sigma$ we obtain (33).

APPENDIX D

In the following we give the favorable input vector combinations and their probability of occurrence for filter length $n = 5$. We take into account that $d \gg s_1, s_2$ and thus $d - s_1 \cong d - s_2$. We begin with the vector median filter:

1) Five vectors of the form $[s_1, s_2]$ with probability of occurrence

$$P_{g_1} = (1 - p)^5.$$

2) Four vectors of the form $[s_1, s_2]$ and one of the form $[s, d]$ or $[d, s_2]$ or $[d, d]$ with probability

$$P_{g_2} = 5(1 - p)^4 \cdot p.$$

3) Three vectors of the form $[s_1, s_2]$ and two of the form $[s, d]$ or $[d, s_2]$ or $[d, d]$, with probability

$$P_{g_3} = 10(1 - p)^3 \cdot p^2.$$

4) Two vectors $[s_1, s_2]$, two vectors $[s_1, d]$ ($[d, s_2]$), one vector $[d, s_2]$ ($[s_1, d]$) with probability

$$P_{g_4} = \frac{5!}{2!2!}(1 - p)^2\left((p_2 p)^2 p_1 p + p_2 p(p_1 p)^2\right).$$

5) Two vectors $[s_1, s_2]$, one vector $[s_1, d]$, one vector $[d, s_2]$, and one vector $[d, d]$ with probability

$$P_{g_5} = \frac{5!}{2!}(1 - p)^2 p_1 p p_2 p p_3 p.$$

6) Two vectors $[s_1, s_2]$, one vector $[s_1, d]$ ($[d, s_2]$), and two vectors $[d, d]$, with probability

$$P_{g_6} = \frac{1}{2}\frac{5!}{2!2!}(1 - p)^2(p_2 p + p_1 p)p^2 p_3^2.$$

By adding the probabilities P_{g_1} through P_{g_6} we get (41a) for the vector median filter. For the marginal median filter, all the above combinations except 6) are valid. Also, another favorable combination is the following: one vector $[s_1, s_2]$, two vectors $[s_1, d]$, and two vectors $[d, s_2]$ with probability

$$P_g = \frac{5!}{2!2!}(1 - p)(p_2 p)^2(p_1 p)^2.$$

By adding the probabilities P_{g_1} through P_{g_5} and P_{g_6} we get (41b) for the marginal median filter.

REFERENCES

[1] A. K. Jain, *Fundamentals of Digital Image Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.

[2] R. C. Gonzalez and P. Wintz, *Digital Image Processing*. Reading, MA: Addison-Wesley, 1987.

[3] N. Magnenat-Thalmann and D. Thalmann, *Image Synthesis*. New York: Springer-Verlag, 1987.

[4] L. Chariglione, Ed., *Signal Processing for HDTV*. New York: Elsevier, 1989.

[5] G. Wyszecki and W. S. Stiles, *Color Science*. New York: Wiley, 1967.

[6] B. R. Hunt, "Karhunen-Loeve multispectral image restoration, part I: theory," *IEEE Trans. Acoust. Speech Signal Processing*, vol. ASSP-32, no. 3, pp. 592-599, June 1984.

[7] N. P. Galatsanos and R. T. Chin "Digital restoration of multichannel images," *IEEE Trans. Acoust. Speech Signal Processing*, vol. 37, no. 3, pp. 415-421, Mar. 1989.

[8] G. Aggelopoulos and I. Pitas "Least-squares multichannel filters in color image restoration," in *Proc. European Conf. Circuit Theory Design ECCTD89* (Brighton, England, Sept. 1989).

[9] C. W. Therrien, "Multichannel filtering methods for color image segmentation," in *Proc. 1985 IEEE Conf. Comput. Vision Pattern Recognition*, 1985, pp. 637-639.

[10] I. Pitas and A. N. Venetsanopoulos, *Nonlinear Digital Filters: Principles and Applications*. Norwell, MA: Kluwer Academic Publ., 1990.

[11] I. Pitas, "Marginal order statistics in color image filtering," *Optical Eng.*, vol. 29, no. 5, pp. 495-503, 1990.

[12] J. Astola, P. Haavisto, and Y. Neuvo, "Vector median filters," *Proc. IEEE*, vol. 78, no. 4, pp. 678-689, Apr. 1990.

[13] G. A. F. Seber, *Multivariate Observations*. New York: Wiley, 1984.

[14] V. Barnett, "The ordering of multivariate data," *J. Statist. Soc. America A*, vol. 139, pt. 3, pp. 318-354, 1976.

[15] H. A. David, *Order Statistics*. New York: Wiley, 1980.

[16] A. M. Mood, "On the joint distribution of the medians in samples from a multivariate population," *Ann. Mathemat. Statist.*, vol. 12, pp. 268-279, 1941.

[17] C. M. Mustafi, "A recurrence relation for distribution functions of order statistics from bivariate distributions," *J. Amer. Statist. Assn.*, vol. 64, pp. 600-601, 1969.

[18] J. Galambos, "Order statistics of samples from multivariate distributions," *J. Amer. Statist. Assn.*, vol. 70, pp. 674-680, 1975.

[19] P. S. Huber, *Robust Statistics*. New York: Wiley, 1981.

[20] F. Hampel, E. Ronchetti, P. Rousseeuw and W. Stahel, *Robust Statistics*. New York: Wiley, 1986.

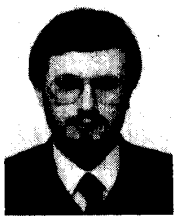
[21] J. Astola, P. Haavisto, P. Heinonen, and Y. Neuvo, "Median type filters for color signals," in *Proc. IEEE Int. Symp. Circuits Syst.* (Helsinki, Finland, 1988), pp. 1753-1756.

[22] J. B. Bednar and T. L. Watt, "Alpha-trimmed means and their relationship to the median filters," *IEEE Trans. Acoust., Speech Signal Processing*, vol. ASSP-32, no. 1, pp. 145-153, Feb. 1984.

[23] Y. H. Lee, and S. A. Kassam, "Generalized median filtering and related nonlinear filtering techniques," *IEEE Trans. Acoust., Speech Signal Processing*, vol. ASSP-33, no. 3, pp. 672-683, June 1985.

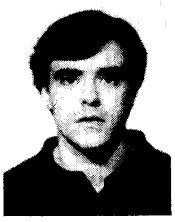
[24] ——, "Nonlinear edge preserving filtering techniques for image enhancement," in *Proc. 27th Midwest Symp. Circuits Syst.*, (1984), pp. 554-557.

[25] B. I. Justusson, "Median filters: Statistical properties," in *Two-dimensional digital signal processing II*, T. S. Huang, Ed. New York: Springer-Verlag, 1981.



Ioannis Pitas received the Diploma of electrical engineering and the Ph.D. degree in electrical engineering from the University of Thessaloniki, Thessaloniki, Greece in 1980 and 1985, respectively.

From 1980 to 1989 he served as a Scientific Assistant and Lecturer at the university. He has also served as a Visiting Research Associate and a Visiting Assistant Professor in the Department of Electrical Engineering, University of Toronto, Toronto, Canada, in 1983-1984 and 1988-1989, respectively. Since 1989 he has been an Assistant Professor at the University of Thessaloniki. His current interests are in the areas of multidimensional digital signal processing, digital image processing, and computer vision.



Panagiotis Tsakalides received the Diploma in electrical engineering from the University of Thessaloniki, Thessaloniki, Greece, in 1990. Since the fall of 1990 he has been working toward the master's degree in computer engineering at the University of Southern California, Los Angeles. He is a Fulbright scholar.

Color Supplement

Multivariate Ordering in Color Image Filtering
(pp. 247-259)

Ioannis Pitas and Panagiotis Tsakalides

Fig. 9. (a) Test image. (b) Image corrupted by multichannel Gaussian noise. (c) Output of a 3 × 3 moving average filter. (d) Output of a 3 × 3 marginal median filter. (e) Output of a 3 × 3 vector median filter. (f) Output of a 3 × 3.5 × 5 DW-MTM filter. (Color supplement continued on next page.)

Pitas and Tsakalides, continued

(a)

(b)

(c)

Fig. 10. (a) Test image corrupted by impulsive noise. (b) Output of a 3 × 3 marginal median filter. (c) Output of a 3 × 3 vector median filter.

(a)

(b)

(c)

Fig. 11. (a) Test image corrupted by mixed Gaussian and impulsive noise. (b) Output of a 3 × 3 marginal median filter. (c) Output of a 3 × 3 vector median filter.