Correspondence

Vector Order Statistics Operators as Color Edge Detectors

P. E. Trahanias and A. N. Venetsanopoulos

Abstract—**Color edge detection is approached in this paper using vector order statistics. Based on the R-ordering method, a class of color edge detectors is defined. These detectors function as vector operators as opposed to component-wise operators. Specific edge detectors can be obtained as special cases of this class. Various such detectors are defined and analyzed. Experimental results show the noise robustness of the vector order statistics operators. A quantitative evaluation and comparison to other color edge detectors favors our approach. Edge detection results obtained from real color images demonstrate the effectiveness of the proposed approach in real applications.**

I. Introduction

An edge is characterized in a monochrome image by an intensity discontinuity. This may correspond to object boundaries or to a change in some physical property, such as illumination (shadows) or reflectance (surface orientation). In the case of multispectral (color) images, color plays a significant role in the perception of boundaries between regions as it is indicated by psychological research on the human visual system [1], [2]. It follows that color information should also be considered in addition to the other physical properties in edge detection approaches for color images. However, this is implicitly not considered in color edge detection schemes where the edges are searched separately in different components of a color image (e.g. intensity, normalized red, normalized green [3]) or in a component that results as a combination of some other components (e.g. brightness [4]). More surprisingly, research in the problem of color edge detection seems to have been neglected, especially when compared to the bulk of work done in monochrome edge detection so far [5]–[8].

The early approaches to color edge detection usually comprise extensions of the monochrome edge detectors to color images. Nevatia [3] has proposed the use of the Hueckel edge operator in the luminance, chrominance color space. Robinson [4] has studied the application of compass gradient edge detection method to color images. The gradient operators proposed for grey-level images [9] can be extended to color images by taking the vector sum of the gradients of the individual components [10], [11]. However, this approach may be very unsatisfactory in certain cases [11]. Consider for example, the case of a color image where in a certain direction, the red component is constant while the green and the blue components both show step edges with the same strength but in opposite directions. Then, the vector sum of the gradients would provide a null gradient. Consequently, latest approaches consider the problem of color edge detection in the vector space. In these approaches, color images are treated as vector fields, as has initially been suggested in [12]. Zenzo [11], and Cumani [13] employ vector gradient and second order derivative operators, respectively. In another approach, reported in [10], the entropy operator is used as an edge detector for monochrome as well as for color images.

A different approach to the problem of color edge detection is proposed in this paper which is based on vector order statistics [14]. This approach is inspired by the morphological edge detectors [15], [16] that have been proposed for the case of monochrome images. These detectors essentially operate by detecting local minima and maxima in the image function and combining them in a suitable way in order to produce a positive response for an edge pixel. Since there is no exact equivalent of the min-max scalar operators for multivariate variables, we rely on the vector ordering schemes that have been proposed in the statistics literature [14]. More specifically, the R-ordering (reduced or aggregate ordering) scheme is adopted, and a class of color edge detectors is defined using linear combinations of the sorted vector samples. The minimum over the magnitudes of these linear combinations defines this class of operators. Different coefficients in the linear combinations result in different edge detectors that vary in simplicity and in efficiency. The coefficients are preset and cannot be changed at run time. It is shown by experimental and simulation results that a certain set of coefficients achieves a robust color edge detector. Moreover, our approach is shown superior to previous approaches based on comparative evaluation results using Pratt's figure of merit [17]. The comparison has been made against the well known vector operators (vector gradient, second order derivative, entropy), and the vector sum of the gradients of the three color components. The final operator is also used since it presents one of the best known component-wise operators [18].

In what follows, a brief introduction to vector order statistics is first given in Section II for the self completeness of the paper. Our approach to color edge detection is described in detail in Section III. Evaluation results as well as edge detection results obtained from real images are presented in Section IV, and our conclusions are outlined in Section V.

II. Vector Order Statistics

Scalar order statistics have played an important role in the design of robust signal analysis techniques. This is due to the fact that any outliers will be located in the extreme ranks in the sorted data. Consequently, these outliers can be isolated and filtered out before the signal is further processed. Ordering of univariate data is well defined and has been extensively studied [19]. Let the n random variables X_i, $i = 1, 2, \ldots, n$, be arranged in ascending order of magnitude as

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)} \tag{1}$$

Then the ith random variable $X_{(i)}$ is the so-called ith order statistic. The minimum $X_{(1)}$, maximum $X_{(n)}$, and median $X_{(n/2)}$ are among the most important order statistics, resulting in the min, max, and median filters, respectively.

Manuscript received July 11, 1992; revised April 1, 1994, and January 19, 1995.

P. E. Trahanias was with the Department of Electrical and Computer Engineering University of Toronto, Toronto, Ontario, Canada M5S 1A4. He is now with the Department of Computer Science, University of Crete, GR 714 09 Heraklion, Crete, and the Institute of Computer Science, Foundation for Research and Technology – Hellas (FORTH) GR 711 10 Heraklion, Crete, Greece.

A. N. Venetsanopoulos is with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, Ontario, Canada M5S 1A4.

Publisher Item Identifier S 1083-4419(96)00420-7.

These concepts are, however, not straightforwardly expanded to multivariate data since there isn't any universal way of defining an ordering in multivariate data. There has been a number of ways proposed to perform multivariate data ordering that are generally classified into [14]: marginal ordering (M-ordering), reduced or aggregate ordering (R-ordering), partial ordering (P-ordering), and conditional ordering (C-ordering).

Let $\mathbf{X}$ represent a $p-$dimensional multivariate $\mathbf{X} = [X_1, X_2, \ldots, X_p]^T$ where X_l, $l = 1, 2, \ldots, p$ are random variables and let $\mathbf{X}^i$, $i = 1, 2, \ldots, n$ be an observation of $\mathbf{X}$. Each $\mathbf{X}^i$ is a $p-$dimensional vector $\mathbf{X}^i = [X_1^i, X_2^i, \ldots, X_p^i]^T$.

In M-ordering the multivariate samples are ordered along each one of the $p-$dimensions independently. For color signals this is equivalent to the separable method where each one of the colors is processed independently. The ith marginal order statistic is the vector $\mathbf{X}^{(i)} = \left[\mathbf{X}_1^{(i)}, \mathbf{X}_2^{(i)}, \ldots, \mathbf{X}_p^{(i)}\right]^T$, where $\mathbf{X}_r^{(i)}$ is the ith largest element in the rth channel. The marginal order statistic $\mathbf{X}^{(i)}$ may not correspond to any of the original samples $\mathbf{X}^1, \mathbf{X}^2, \ldots, \mathbf{X}^n$ as it does in one dimension.

In R-ordering, each multivariate observation $\mathbf{X}^i$ is reduced to a scalar value d_i according to a distance criterion. A metric that is often used is the generalized distance to some point $\mathbf{x}$. The samples are then arranged in ascending order of magnitude of the associated metric values d_i.

In P-ordering the objective is to partition the data into groups or sets of samples, such that the groups can be distinguished with respect to order, rank, or extremeness. This type of ordering can be accomplished by using the notion of convex hulls. However, the determination of the convex hull is difficult to do in more than two dimensions. Other ways to achieve P-ordering are ad hoc partitioning procedures and are thus not preferred. Another drawback associated with P-ordering is that there is no ordering within the groups and thus it is not easily expressed in analytical terms. These properties make P-ordering infeasible for implementation in digital image processing.

In C-ordering the multivariate samples are ordered conditional on one of the marginal sets of observations. This has the disadvantage in color image processing that only the information in one component (channel) is used.

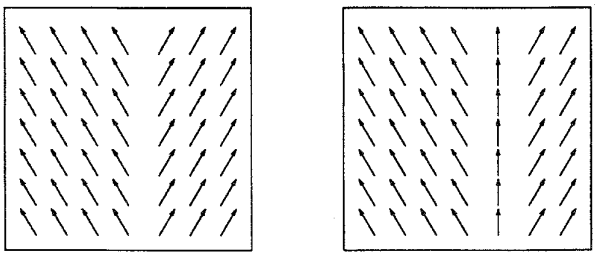
From the above it is evident that R-ordering is more appropriate for color image processing than the other vector ordering methods. If we employ as a distance metric the aggregate distance of $\mathbf{X}^i$ to the set of vectors $\mathbf{X}^1, \mathbf{X}^2, \ldots, \mathbf{X}^n$, then

$$d_i = \sum_{k=1}^{n} \left\| \mathbf{X}^i - \mathbf{X}^k \right\|, \quad i = 1, 2, \ldots, n \tag{2}$$

where $\|\cdot\|$ represents an appropriate vector norm. The arrangement of the d_is in ascending order $\left(d_{(1)} \leq d_{(2)} \leq \cdots \leq d_{(n)}\right)$, associates the same ordering to the multivariate $\mathbf{X}^i$s

$$\mathbf{X}^{(1)} \leq \mathbf{X}^{(2)} \leq \cdots \leq \mathbf{X}^{(n)} \tag{3}$$

In the ordered sequence $\mathbf{X}^{(1)}$ is the vector median of the data samples [20]. It is defined as the vector contained in the given set whose distance to all other vectors is a minimum. Moreover, vectors appearing in low ranks in the ordered sequence are vectors centrally located in the population, whereas vectors appearing in high ranks are vectors that diverge mostly from the data population. These samples are generally called "outliers." It follows that this ordering scheme gives a natural definition of the median of a population and of the outliers of a population.



(a) (b)

Fig. 1. (a) ideal color step edge ($\mathbf{V}_1 \equiv\nwarrow$, $\mathbf{V}_2 \equiv\nearrow$), (b) ideal 3-pixel color ramp edge ($\mathbf{V}_1 \equiv\nwarrow$, $\mathbf{V}_2 \equiv\nearrow$, $\mathbf{V}_3 = \frac{\mathbf{V}_1+\mathbf{V}_2}{2} \equiv\uparrow$).

III. Color Edge Detection

A. Notation and Definitions

In this work a color image is viewed as a vector field, represented by a discrete vector valued function $\mathbf{f}(\mathbf{x}) : Z^2 \rightarrow Z^m$, where Z represents the set of integers[1]. A notation will be used in the following concerning the image function $\mathbf{f}$. For $W \subset Z^2$, $\mathbf{x}_i \in W$, $i = 1, 2, \ldots, n$, n is the size (number of pixels) of W, $\mathbf{f}(\mathbf{x}_i)$ will be denoted as $\mathbf{X}^i$. $\mathbf{X}^{(i)}$ will denote the ith ordered vector in the window W according to the R-ordering method where the aggregate distance is used as a distance metric. Consequently, $\mathbf{X}^{(1)}$ is the vector median in the window W and $\mathbf{X}^{(n)}$ is the outlier in the highest rank of the ordered vectors.

Although definitions of color edges have been given in other works [3], [13] we give here a loose definition of color edges in the context of vector fields. We also want to extend the notion of *ramp* edges that is well understood in monochrome images to color images. It is noted that the following definitions are not intended as formal definitions that can lead to edge detection operators but rather as intuitive descriptions of the notion of color edges in order to facilitate our discussion on edge detectors. We define a color edge as any *significant* discontinuity in the vector field representing the color image function. An abrupt change in the vector field characterizes a color *step* edge, whereas a gradual change characterizes a color *ramp* edge. These points are illustrated in Fig. 1. A vertical color step edge is shown in Fig. 1(a), whereas a vertical color 3-pixel ramp edge is shown in Fig. 1(b).

B. Color Edge Detectors Based on Vector Order Statistics

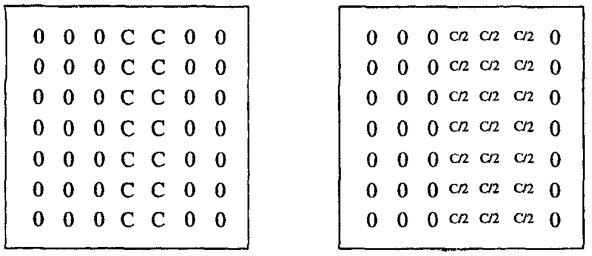
Based on the previous discussion on vector order statistics, we define the basic color edge detector, the *vector range*[2] (VR) edge detector, as

$$VR = \left\| \mathbf{X}^{(n)} - \mathbf{X}^{(1)} \right\| \tag{4}$$

VR expresses in a quantitative way the deviation of the vector outlier in the highest rank from the vector median in W. Consequently, in a uniform area, where all vectors will be close to each other, the output of VR will be small. However, its response on an edge will be large since $\mathbf{X}^{(n)}$ will be selected among the vectors from the one side of the edge (the smaller side, when we assume that W is divided by the edge into two unequal sides) while $\mathbf{X}^{(1)}$ will be selected among the vectors from the other side of the edge (the larger side). By thresholding the output of VR the actual edges can be obtained.

[1] Usually $m = 3$ but the results presented hold for $m \geq 2$.

[2] The name is borrowed from the scalar case where $X^{(n)} - X^{(1)}$ is the range of the ordered random variables.



(a) (b)

Fig. 2. (a) response of VR to the image of Fig. 1(a), $C = \|\mathbf{V}_1 - \mathbf{V}_2\|$, (b) response of VR to the image of Fig. 1b, $C/2 = \|\mathbf{V}_1 - \mathbf{V}_3\| = \|\mathbf{V}_2 - \mathbf{V}_3\|$.

The response of VR to the two images (vector fields) of Fig. 1 is shown in Fig. 2[3]. As can be verified from Fig. 2, VR introduces no bias in the case of a color step edge since it responds equally to both sides of the edge (Fig. 2(a)), but it responds with a 3−pixel wide edge in the case of an ideal color ramp edge. VR is also sensitive to noise, especially to noise modeled with a heavy-tailed distribution (e.g. double-exponential) which, according to [20], [6], is termed as *heavy-tailed* noise. Similarly, noise modeled with a short-tailed distribution (e.g. gaussian) is termed as *short-tailed* noise.

VR will respond with n pixels (the pixels that belong to W) to a single noisy pixel in the center of W. This drawback leads us to consider dispersion measures which are known as more robust estimates in the presence of noise [19]. A class of operators can be defined as a linear combination of the ordered vectors; VR is a special case of this class. This class of operators expresses a measure of the dispersion of the ordered vectors, and hence the name *vector dispersion edge detectors* ($VDED$):

$$VDED = \left\| \sum_{i=1}^{n} \alpha_i \mathbf{X}^{(i)} \right\| \tag{5}$$

where α_i are proper coefficients (weights). VR is obtained from $VDED$ for $\alpha_n = 1$, $\alpha_1 = -1$, and $\alpha_i = 0$, $i = 2, \ldots, n-1$. Equation (5) can be further generalized by considering k sets of coefficients α_i and combining the resulting vector magnitudes in a suitable way. The combination that is proposed employs a minimum operator which attenuates the effect of noise. Such a minimum operator has also been employed in [16] in order to derive a detector insensitive to noise impulses. According to the above, the general class of color edge detectors based on vector order statistics is defined as (6), shown at the bottom of the page, where $MVDED$ stands for *minimum* $VDED$. Specific color edge detectors can be obtained from $MVDED$ by selecting sets of coefficients α_{ij}. Since a strict mathematical approach to this seems very difficult, we select sets of α_{ij}s that satisfy the requirements of (a) noise insensitivity, and (b) proper response to ramp edges.

[3]A 3×3 window is assumed in this and all subsequent examples presented in this section.

The resulting edge detectors, that conform to the above requirements, are considered next. Two cases are presented that result in operators with (a) immunity to heavy-tailed noise, and (b) immunity to short-tailed noise and improved performance in the presence of ramp edges. An operator that combines these characteristics is finally introduced. For the sake of brevity, the coefficients α_{ij} will implicitly be considered as being equal to zero for all the pairs i, j for which they are not defined.

Case 1: $\alpha_{n1} = \alpha_{n-1,2} = \ldots = \alpha_{n-k+1,k} = 1$, $\alpha_{1j} = -1$, $j = 1, 2, \ldots, k$, $k < n$. As explained previously, VR is sensitive to noise (especially to heavy-tailed noise) since the vectors placed in the highest ranks of the ordered data will essentially correspond to noisy samples and consequently VR will erroneously respond in the presence of noise. This problem can be alleviated by employing the magnitudes of the differences of the k highest vectors from $\mathbf{X}^{(1)}$ and obtain an edge detector that is insensitive to heavy-tailed noise (e.g. impulsive or exponential). In terms of (6) this can be formulated as

$$MVR = \min_j \left\{ \left\| \mathbf{X}^{(n-j+1)} - \mathbf{X}^{(1)} \right\| \right\}, \quad j = 1, 2, \ldots, k, \; k < n \tag{7}$$

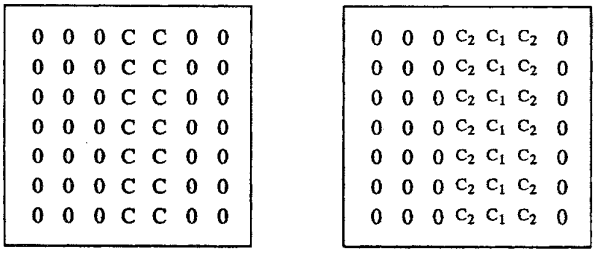
where MVR stands for *minimum* VR. The choice of k depends on n, the size of W. For a 3×3 window, for example, $k \leq 3$ and for a 5×5, $k \leq 10$. Unfortunately, there is no general formula to obtain k and its value should be subjectively estimated. However, this is not a difficult task since k can be interpreted as the number of pixels that belong to the smaller side of an edge, when W is centered on an edge pixel. Moreover, its value is not crucial in the overall performance of MVR. The response of MVR to ideal step and ramp color edges is exactly the same as the response of VR. However, it exhibits improved noise immunity. It will not respond to up to $k-1$ noisy pixels in W.

A final comment on MVR concerns the use of the min function. We note that this can not be simply replaced by taking the magnitude of the vector difference between $\mathbf{X}^{(n-k+1)}$ (the "smallest" among the $\mathbf{X}^{(n-j+1)}$, $j = 1, 2, \ldots, k$, vectors) and $\mathbf{X}^{(1)}$ since it will yield in general different results. In other words, the magnitudes of vector differences in a sorted vector data set, do not necessarily follow the order of the sorted vectors. This is exemplified by a simple example concerning the set of $2-D$ vectors $S_v = \{(5,4), (4,5), (4,7), (4,2), (3,4)\}$. Sorting S_v will result in the ordered set $S_o = \{(4,5), (5,4), (3,4), (4,2), (4,7)\}$. It is noted that $(4,7)$ has been placed in the highest rank in the sorted data set and $(4,2)$ has been placed in the second highest rank. The distance, however, of $(4,2)$ to the vector median $(4,5)$ is larger than the distance of $(4,7)$ to $(4,5)$ for both L_1 and L_2 metrics.

Case 2: $k = 1$, $\alpha_{n1} = 1$, $\alpha_{i1} = -1/l$, $i = 1, 2, \ldots, l$, $l < n$. This choice of coefficients results in a *vector dispersion* (VD) edge detector that employs an averaging scheme

$$VD = \left\| \mathbf{X}^{(n)} - \sum_{i=1}^{l} \frac{\mathbf{X}^{(i)}}{l} \right\|, \quad l < n \tag{8}$$

$$\begin{aligned} MVDED &= \min \left\{ \left\| \sum_{i=1}^{n} \alpha_{i1} \mathbf{X}^{(i)} \right\|, \left\| \sum_{i=1}^{n} \alpha_{i2} \mathbf{X}^{(i)} \right\|, \ldots, \left\| \sum_{i=1}^{n} \alpha_{ik} \mathbf{X}^{(i)} \right\| \right\} \\ &= \min_j \left\{ \left\| \sum_{i=1}^{n} \alpha_{ij} \mathbf{X}^{(i)} \right\| \right\}, \quad j = 1, 2, \ldots, k \end{aligned} \tag{6}$$



(a) (b)

Fig. 3. (a) response of VD to the image of Fig. 1a, $C = \|\mathbf{V}_1 - \mathbf{V}_2\|$, (b) response of VD to the image of Fig. 1b, $C_1 > C_2$.

The response of VD to the two images of Fig. 1 is shown in Fig. 3. Its behavior is exactly the same with the behavior of VR and MVR in the case of the color step edge. However, it has improved performance in the case of a color ramp edge since it assigns a higher value (C_1) to the true edge pixel than the value (C_2) assigned to the neighboring pixels (the fact that $C_1 > C_2$ can be very easily proved by geometrical considerations). Consequently, by thresholding the output of VD with a proper threshold value, the ramp edges can be extracted. For ideal ramp edges of larger spatial extent than the detectors window, more than one central pixels will be given a large value. However, by enlarging the window size the central (true edge) pixel will again be given a higher value than its neighbors and it can be detected by thresholding.

The value of the parameter l used in the computation of VD can not be formally defined as it is the case for the parameter k employed in the computation of MVR. However, a duality exists between these two parameters; l expresses the number of pixels that belong to the larger side of an edge when W is centered on an edge pixel. Therefore, proper values for the parameter l can be subjectively estimated for various sizes of W.

The second term in the vector difference in (8) $\left(\sum_{i=1}^{l} \frac{\mathbf{X}^{(i)}}{l}\right)$ is the *vector α–trimmed mean* ($V\alpha TM$) [21], which is a robust signal estimate in color images when the noise is modeled as short-tailed [22]. It is, therefore, expected that VD will have improved performance in the presence of short-tailed noise due to the smoothing performed.

Case 3: $\alpha_{n1} = \alpha_{n-1,2} = \ldots = \alpha_{n-k+1,k} = 1$, $\alpha_{ij} = \frac{-1}{l}$, $i = 1, 2, \ldots, l$, $j = 1, 2, \ldots, k$, $l,k < n$. The two operators introduced previously (MVR, VD) have some desirable properties that are different in each case. MVR is insensitive in the presence of heavy-tailed noise. VD responds properly to color ramp edges and has improved performance in the presence of short-tailed noise. We wish to combine these two operators in order to exploit the properties of both. This can be achieved by the set of coefficients α_{ij} outlined above. The resulting edge detector (*minimum vector dispersion* – MVD) is given as

$$MVD = \min_{j} \left\{ \left\| \mathbf{X}^{(n-j+1)} - \sum_{i=1}^{l} \frac{\mathbf{X}^{(i)}}{l} \right\| \right\}, \quad j = 1, 2, \ldots, k,\ k, l < n \tag{9}$$

MVD inherits the properties of its ancestors. It is a bias free operator for color step edges whereas it produces a larger response for the true edge of a color ramp edge. Its response to the images of Fig. 1 is exactly the same as the response of VD (Fig. 3). Moreover, it has improved noise performance since it is robust in the presence of heavy-tailed noise (due to the minimum operation) and short-tailed noise (due to the averaging operation). The noise performance of MVD is illustrated in Fig. 4 for the cases of gaussian (Fig. 4(c)) and double-exponential (Fig. 4(e)) noise. The response of MVD, shown in Fig. 4(d) and (f), respectively, is much larger at the true edge which is, therefore, easily detected by thresholding.

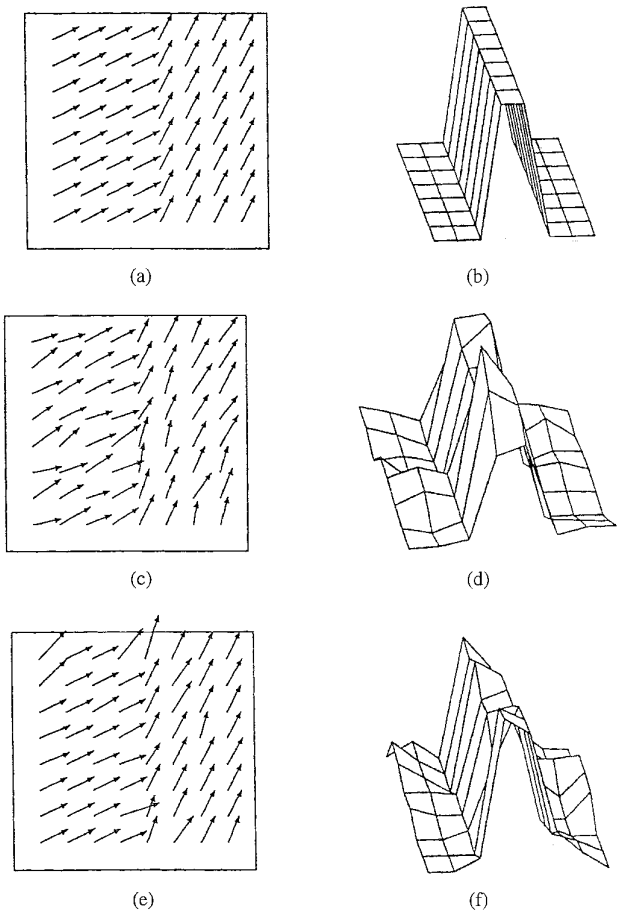


Fig. 4. Response of MVD to noise contaminated edge. (a) Initial edge, (b) Response of MVD to (a), (c) edge (a) corrupted with gaussian noise, (d) response of MVD to (c), (e) edge (a) corrupted with double-exponential noise, (f) response of MVD to (e).

C. Statistical Considerations

A statistical analysis of MVD is attempted here in order to derive the error probability of the edge detector. The analysis is confined to the case of multivariate normal distributions only, since these are the only distributions for which analytical results concerning the distribution functions have been derived in the statistics literature. We consider an ideal edge model with the sample vectors $\mathbf{X}_i$ on the one side of the edge as instances of a random variable $\mathbf{X}$ which follows a multivariate normal distribution $N_m(\mu_x, I_m)$. Similarly, the sample vectors $\mathbf{Y}_i$ on the other side are instances of the random variable $\mathbf{Y}$ which is $N_m(\mu_y, I_m)$. Then, the error probability is given as

$$P_E = P_e P_M + P_n P_F \tag{10}$$

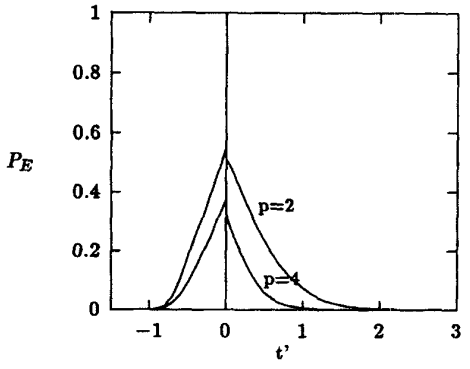


Fig. 5. P_E versus t' (see text for explanation).

where P_e, P_n denote the prior probabilities of "edge" and "no edge," respectively, and P_M, P_F are the probabilities of missing an edge and false edge alarm, respectively.

A detailed analysis of (10) is carried out in the Appendix. Based on the results presented there, a plot of P_E is drawn which is shown in Fig. 5. The parameter p shown on the plot expresses the number of sample distances (i.e. ordered vectors) and is equal to k [see (9)]. It is, therefore, related to the window size n. From (16), (17), and Fig. 5 we can conclude that P_E is decreased with respect to p. Since p increases with the window size n, it is concluded that the probability of error P_E decreases with respect to n. However, n can not be drastically increased since it is well known that a large window has a negative effect in accurate edge localization. Besides, n should be kept reasonably small due to computational requirements as we will see in the next section.

D. Computational Considerations

Computationally the vector ordering task is a $O(n^2)$ operation since $\frac{1}{2}n(n-1)$ vector distances have to be computed. The ordering itself may be performed by a fast sorting algorithm (e.g. quicksort) with a $O(n \log n)$ complexity. Therefore, the whole task of vector ordering is dominated by the $O(n^2)$ term which determines its complexity. For the case, however, of color image processing, a fast algorithm has been proposed for the computation of the vector distances [20]. This algorithm is based on the principle that as the window W (of size n) moves in the image plane, only m ($< n$) new pixels are considered each time whose distances have to be computed. The distances of the rest $n-m$ pixels are simply updated. This algorithm results in a $O\left(n^{3/2}\right)$ complexity for the computation of the vector distances in the case of square windows ($m = \sqrt{n}$). The trade-off is the increased programming complexity in the computation of the distances.

The $V\alpha TM$ can be computed only once since it does not depend on the index j [(9)]. Its computation is a $O(l)$ operation, $l < n$.

The computation of the magnitudes of the k vector differences $\left(\left\| \mathbf{X}^{(n-j+1)} - V\alpha TM \right\| \right)$ is a $O(k)$ operation, $k < n$. This operation can be performed simultaneously with the minimization operation. In summary, the computational complexity of (9) is $O(n^2)$ if all the vector distances in W are computed, which can be reduced to $O\left(n^{3/2}\right)$ if the fast algorithm is employed. For small n, which is the case in edge detection, even the $O(n^2)$ complexity is not very expensive. Furthermore, if the L_1 norm is adopted only integer operations are involved, whereas for the L_2 norm, square root operations have to be employed. However, the square roots can be avoided, which is equivalent to computing the MVD squared.

IV. Experimental Results

Quantitative evaluation of the performance of edge detectors is complicated since different evaluation criteria are employed by the designers which lead to different performance figures. Moreover, quantitative evaluation is in many cases not performed and only qualitative results are presented which are very difficult to be used in assessing the performance of edge detectors. The quantitative performance measures can be grouped into two types, probabilistic measures and distance measures. The first type is based upon the statistics of false edge detection and false edge rejection. The second type is based upon edge deviation or error distance which is the minimum distance between the detected and the truth edge. A distance measure that is often used in edge detector's evaluations is Pratt's figure of merit (FOM) [17]. It is defined as

$$FOM = \frac{1}{\max\{I_D, I_I\}} \sum_{i=1}^{I_D} \frac{1}{1+\alpha(d_i)^2} \tag{11}$$

where I_D, I_I are the number of detected and number of ideal edge points respectively, α (> 0) is a calibration constant, and d_i is the separation distance of the ith detected edge point normal to a line of ideal edge points [17]. In all cases $0 < FOM \leq 1$; for a perfect match between the detected and the ideal edges $FOM = 1$, whereas as the detected edges deviate more and more from the ideal ones FOM goes to zero. FOM has been adopted in this work due to its advantage over the probabilistic measures that it renders a more realistic appraisal of the detected edges [7]. If we consider, for example, the case where all the edges are 1-pixel shifted from the ground truth, a probabilistic measure would give a very poor rating but FOM still gives a performance measure very close to unity (0.9). Moreover, FOM has been used by many authors in the evaluation of edge detectors [7], [17], [25]. The scaling constant $\alpha = 1/9$ proposed by Abdou and Pratt has been adopted in our simulations.

A. Performance Evaluation

An artificial image has been created and used as a benchmark for assessing the performance of the vector order statistics operators and for comparison purposes. It is shown in Fig. 6 along with its red, green and blue components. This image has been intentionally created in order to meet a number of requirements: (a) it contains vertical, horizontal and diagonal edges, (b) there are edges where all three color components, or two color components, or only one color component change, (c) black (all color components are zero), white (all color components are equal to the maximum value, 255) and grey (all color components have the same value) areas are present on the image, and (d) isoluminant (areas with identical luminance) and nonisoluminant areas are present on the image.

A number of edge detection experiments have been conducted using various noise distributions at various noise levels to contaminate the test image. In each case, FOM has been measured and used as the performance criterion. The ground truth (real edges) that is needed for the computation of FOM, is trivially obtained for the noise free (original) artificial image (Fig. 6) with the application of any edge detector. Three noise types were used in the experiments: gaussian, double-exponential and impulsive. For each noise type two sets of experiments have been performed. In the first set, the noise process in each channel has been considered as an independent process. In the second set, the noise process has been considered as a correlated process since there is some indication that this type of correlation may exist in real color images.

The performance measures of the color edge detectors based on vector order statistics are shown graphically in Fig. 7. A number of

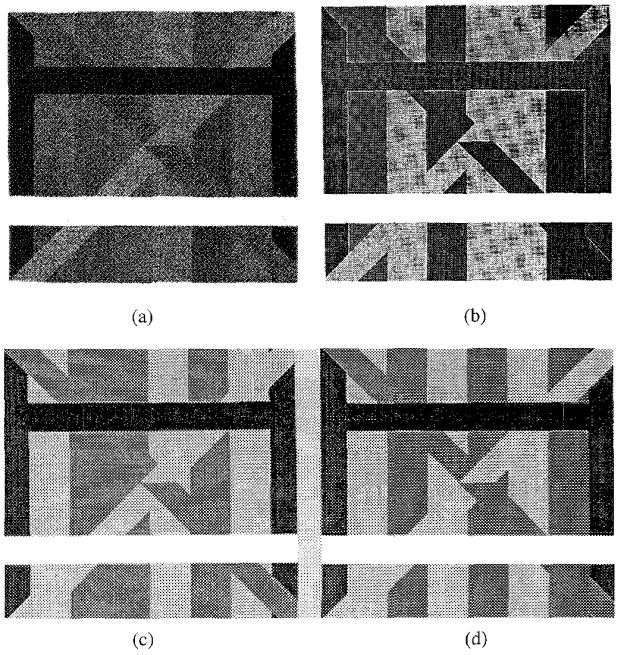


Fig. 6. Artificial image used for testing and comparison purposes (a) image, (b) red component, (c) green component, (d) blue component.

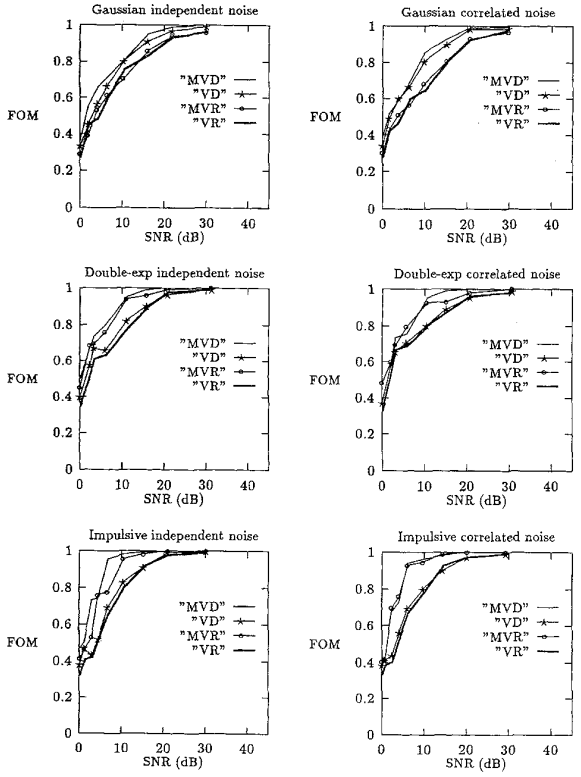


Fig. 7. FOM plots for the color edge detectors based on vector order statistics.

conclusions can be drawn from the graphs of Fig. 7 which are in accordance with the structure of the edge detectors:

- The performance of MVD is superior compared to the other operators for all types of noise. This is quite as expected since MVD has actually been assembled using the desirable properties of the other operators.
- VR is sensitive to noise and its performance is largely deteriorated as noise increases.
- MVR and VD have good performances for heavy-tailed and short-tailed noise, respectively.

From the plots of Fig. 7 we also conclude that the performance deterioration of the vector order statistics edge detectors is "smooth" as noise increases. Only for the case of impulsive noise we observe a more "precipitous" performance degradation in the range 6–8 dB, which, for the test image of Fig. 6, corresponds to 8–10% of noise corruption. This noise level is sufficiently high to justify the performance degradation, especially for the case of color images where the contribution of the three channels causes one in every three or four pixels to be corrupted.

In the experiments described above the L_1 norm has been used. The reason for this is that very similar results have been obtained for both the L_1 and the L_2 norms, but L_1 is cheaper to compute. The values used for the parameters of the operators were: 5×5 window, $k = 8$, $l = 12$. It should be noted here that these values have been experimentally obtained and, more importantly, they are not critical in the overall performance. Practically, their performance has been left unchanged for $7 \leq k \leq 10$ and $10 \leq l \leq 15$. Regarding the window size, even a 3×3 window has given very good results for low noise levels. However, for higher noise levels the 5×5 window performs better since it involves more pixels in the edge detection process.

For comparison purposes, the same test image and the same evaluation procedure have been used. In the experiments conducted MVD has been compared against four other color edge detectors: the vector gradient operator [11], the second-order derivative operator [13], the entropy operator [10], and the vector sum of the gradients of the three color components. The last operator has been selected among the operators that result from combining component-wise edge detectors since it produces results that are generally better [18].

The performance measures obtained are shown in Fig. 8. As can be verified from this figure, all the operators have good performance for low noise levels, although MVD performs slightly better. However, as the SNR decreases, the superiority of the proposed MVD becomes clear for all noise types. The performance of the vector gradient operator and the second-order derivative operator is close enough to the performance of MVD for gaussian noise but is inferior for double-exponential and impulsive noise. The other two operators have lower performance measures. Especially the performance of the entropy operator is largely degraded as noise increases.

In order to impart an intuitive feel of the behavior of MVD and the other four color edge detectors, and complement the quantitative results presented in Figs. 7 and 8, the edge detection results on a noisy version of the test image are shown in Fig. 9. Fig. 9(a) shows the test image corrupted with 5% impulsive noise with a noise channel correlation factor $\rho = 0.5$. Fig. 9(b)–(f) show the edge detection results of the corresponding operators. The superiority of MVD (Fig. 9(b)) is clearly demonstrated in this example. It is robust in the presence of noise while at the same time it is sensitive to the image edges. The vector gradient based operators (Fig. 9(c) and (d)) exhibit a degree of robustness but the entropy operator (Fig. 9(e)) is very noise sensitive. The vector sum of gradients operator (Fig. 9(f)) performs comparably to the vector gradient operator (Fig. 9(c)); it is interesting to note, however, that it fails to detect two vertical edges (red – cyan edge, yellow – pink edge) since the image gradients change in opposite directions in these areas.

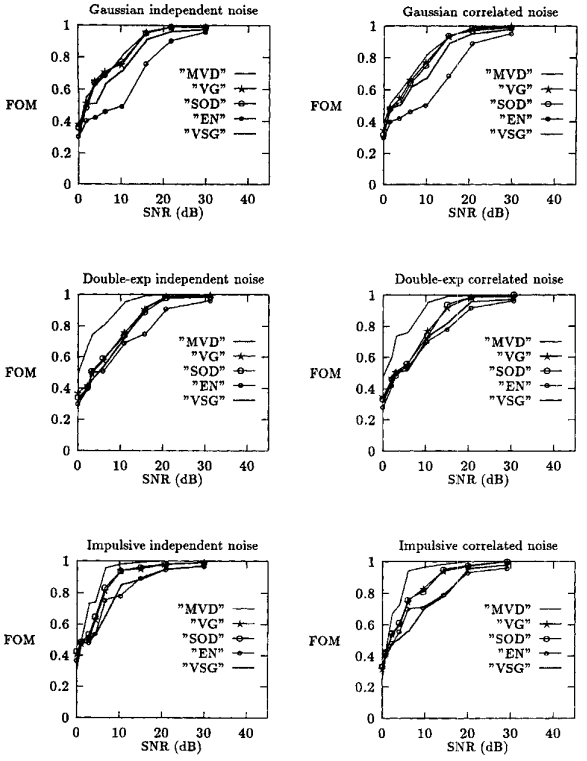


Fig. 8. FOM plots for five color edge detectors: MVD, vector gradient (VG), second-order derivative (SOD), entropy (EN), vector sum of gradients (VSG).

B. Application to Real Images

Since the topic is *image processing*, the subjective evaluation constitutes a very important criterion as of the performance of any operator. Consequently, we have applied the vector order statistics based operators to real color images. Many different kinds of images have been utilized ranging from human faces to detailed outdoor scenes. In all cases, the results obtained were in good agreement with our subjective criteria for color edges. The edge detection results for an "airplane" image are presented here. This image has been selected because of the difficulty associated with the edge detection task due to the low contrast edges. The original color image is shown in Fig. 10(a) and the MVD results are presented in Fig. 10(b). The results of the color edge detectors used for comparison are also shown in Fig. 10(c)-(f), respectively. A visual evaluation gives the impression that the MVD and the vector gradient perform comparably, but still MVD produces thinner edges and is less sensitive to small texture variations. This can be ascribed to the averaging operation which essentially smooths out small variations whereas the gradient-based operators are sensitive even to small changes. The second-order derivative operator is very sensitive to texture variations and the edges produced give the impression of a blurred original image. The reason for this is that a regularizing filter has to be applied before the computation of the derivatives. As suggested in [13], the gaussian filter is used; a standard deviation value $\sigma = 3.5$ has been selected through experimentation. This large value of σ is needed for the computation of the second-order derivatives and is responsible for the blurring produced. The entropy operator is totally insensitive to texture variations but on the other hand it leaves undetected many edges that correspond to fine image details. The results of the vector sum of the gradients operator

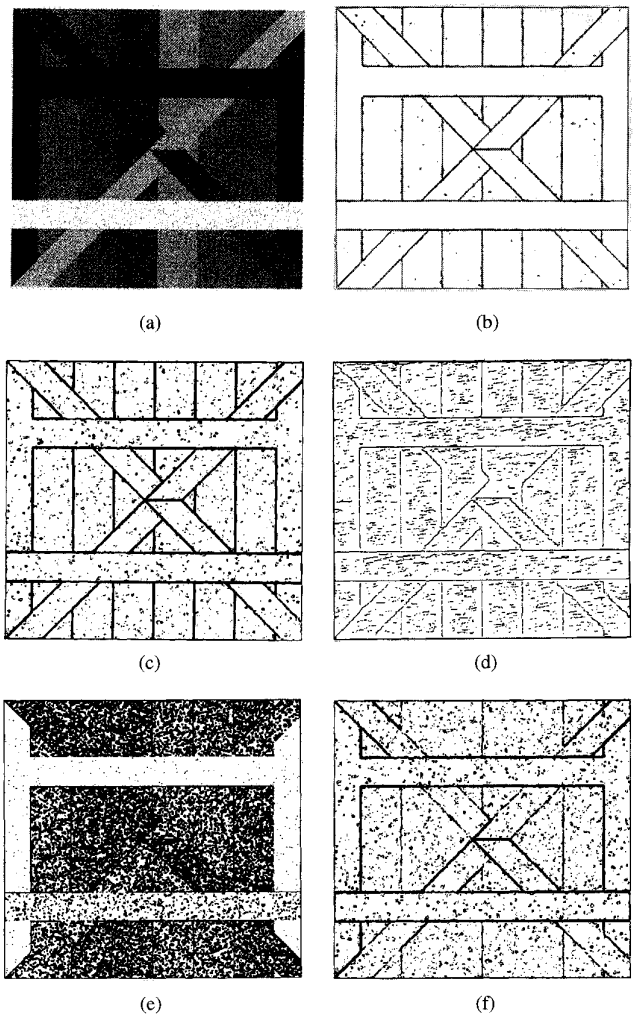


Fig. 9. Noise behavior of color edge detectors. (a) Artificial image corrupted with 5% correlated ($\rho = 0.5$) impulsive noise, (b) MVD, (c) vector gradient, (d) second-order derivative (e) entropy, (f) vector sum of gradients.

resemble the results of the vector gradient operator, but are generally inferior. Moreover, this operator can not detect some simple edges when the gradients of the image components change in the opposite direction (see the back wing for example).

The noise behavior of the color edge detectors is illustrated in Fig. 11. Fig. 11(a) shows the airplane image corrupted with 5% correlated ($\rho = 0.5$) impulsive noise. The results of the MVD and the vector gradient operators are shown in Fig. 11(b) and (c), respectively. As can be verified the performance of the MVD is superior to the vector gradient. The results of the other three operators (second-order derivative, entropy and vector sum gradient) are not shown since they are inferior.

V. Conclusions

The problem of color edge detection has been studied using vector order statistics in this paper. A class of color edge detectors has been proposed and efficient operators from this class have been derived by proper choice of the parameters. The experimental results presented demonstrate the effectiveness of this approach for noiseless as well as for noise contaminated images. The robustness of the proposed

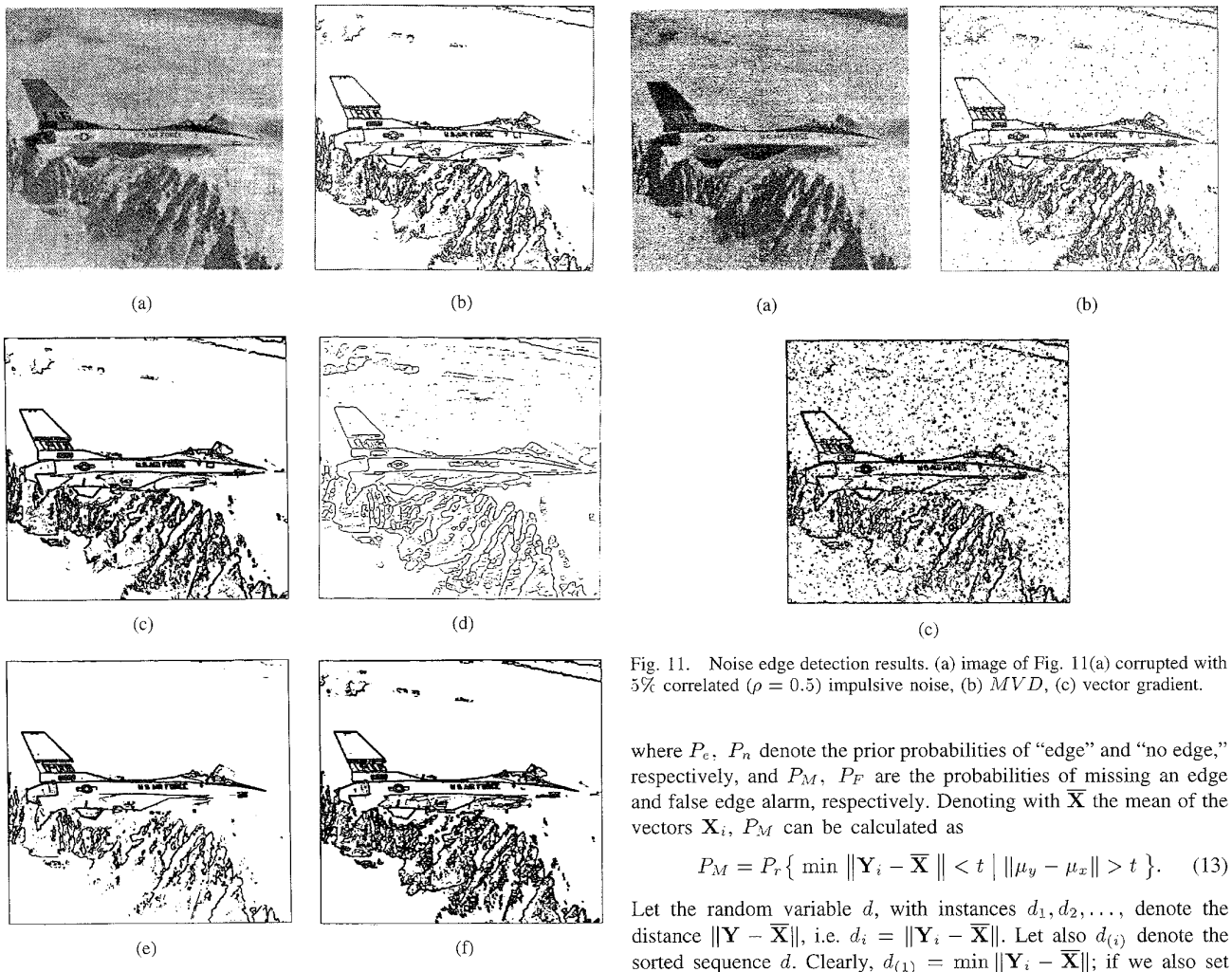


Fig. 10. Edge detection results. (a) Original image, (b) MVD, (c) vector gradient, (d) second-order derivative (e) entropy, (f) vector sum of gradients.

Fig. 11. Noise edge detection results. (a) image of Fig. 11(a) corrupted with 5% correlated ($\rho = 0.5$) impulsive noise, (b) MVD, (c) vector gradient.

operators has been experimentally shown for different types of noise and even for very low SNR's.

Vector order statistics are gaining importance in color image processing [20], [26], [22] since they offer a means of ordering multivariate samples. Until now, however, they have been used only in image filtering tasks by exploiting their property of placing the central vectors in the lower ranks of the sorted data. In this work, vector order statistics have been employed for the first time for a different task, namely color edge detection. Accurate detection of the edges is of primary importance for the later steps in an image analysis system. It has been shown that the proposed approach achieves this goal and consequently it may be very useful in color image analysis.

APPENDIX

Let an ideal edge model with the sample vectors $\mathbf{X}_i$ on the one side of the edge be instances of a random variable $\mathbf{X}$ which follows a multivariate normal distribution $N_m(\mu_x, I_m)$. Similarly, the sample vectors $\mathbf{Y}_i$ on the other side are instances of the random variable $\mathbf{Y}$ which is $N_m(\mu_y, I_m)$. The error probability is given as

$$P_E = P_e P_M + P_n P_F \tag{12}$$

where P_e, P_n denote the prior probabilities of "edge" and "no edge," respectively, and P_M, P_F are the probabilities of missing an edge and false edge alarm, respectively. Denoting with $\overline{\mathbf{X}}$ the mean of the vectors $\mathbf{X}_i$, P_M can be calculated as

$$P_M = P_r\{ \min \|\mathbf{Y}_i - \overline{\mathbf{X}}\| < t \mid \|\mu_y - \mu_x\| > t \}. \tag{13}$$

Let the random variable d, with instances $d_1, d_2, \ldots$, denote the distance $\|\mathbf{Y} - \overline{\mathbf{X}}\|$, i.e. $d_i = \|\mathbf{Y}_i - \overline{\mathbf{X}}\|$. Let also $d_{(i)}$ denote the sorted sequence d. Clearly, $d_{(1)} = \min \|\mathbf{Y}_i - \overline{\mathbf{X}}\|$; if we also set $\|\mu_y - \mu_x\| = \tau$, (13) can be rewritten as

$$\begin{aligned} P_M &= P_r\{ d_{(1)} - \tau < t' \mid t' < 0 \}, \; t' = t - \tau \\ &= \frac{P_r\{ d\tau_{(1)} < t', \; t' < 0 \}}{P_r\{ t' < 0 \}} \\ &= \frac{F_{d\tau_{(1)}}(t')}{P_r\{ t' < 0 \}}, \; t' < 0 \end{aligned} \tag{14}$$

where $d\tau_{(1)} = d_{(1)} - \tau$. Carrying out similar computations, P_F is given as

$$P_F = 1 - \frac{F_{d\tau_{(1)}}(t')}{P_r\{t' \geq 0\}}, t' \geq 0 \tag{15}$$

We also observe that $P_e = P_r\{t' < 0\}$, $P_n = P_r\{t' \geq 0\}$ and, consequently

$$\begin{aligned} P_E = {} & F_{d\tau_{(1)}}(t')u(-t') + P_r\{t' \geq 0\} \\ & - F_{d\tau_{(1)}}(t')u(t') \end{aligned} \tag{16}$$

where $u(x)$ is the unit step function and $F_{d\tau_{(1)}}(t') = F_{d_{(1)}}(t)$ since $F_{d\tau_{(1)}}(t') = P_r\{d\tau_{(1)} \leq t'\} = P_r\{d_{(1)} \leq t\} = F_{d_{(1)}}(t)$. $F_{d_{(1)}}$ can be obtained from F_d, the distribution function of d as [19, p. 8]:

$$F_{d_{(1)}}(x) = 1 - [1 - F_d(x)]^p \tag{17}$$

where p denotes the number of sample distances. In our case $p = k$ [see (9)]. The problem is now reduced to the estimation of F_d. If we

consider Euclidean distances then d^2 follows a noncentral chi-square distribution with m degrees of freedom and noncentrality parameter $s = (\mu_y - \mu_x)^T(\mu_y - \mu_x)$ [23, p. 19]. The cumulative distribution function of the noncentral chi-square distribution, when $z = m/2$ is an integer, can be expressed in terms of the generalized Q function as $F_{d^2}(y) = 1 - Q_z(s, \sqrt{y})$ [24, p. 29]. Since the distances d are nonnegative, F_d can be obtained by a simple change in variables

$$F_d(y) = P_r\{d \le y\} = P_r\{d^2 \le y^2\} = F_{d^2}(y^2) = 1 - Q_z(s, y) \quad (18)$$

From (18), (17) can be computed and (16) can also be computed provided that $P_r\{t' \ge 0\}$ is known. For our model, $t' = t - \tau$, where t is the detector's threshold (deterministic quantity) and $\tau = \|\mu_y - \mu_x\|$ (constant). Therefore, t' is a deterministic quantity and $P_r\{t' \ge 0\}$ is unit or zero for $t' \ge 0$ or $t' < 0$, respectively.

REFERENCES

[1] A. Treisman and G. Gelade. "A feature integration theory of attention," *Cogn. Psych.*, vol. 12, pp. 97–136, 1980.

[2] A. Treisman. "Features and objects in visual processing," *Scientific Amer.*, vol. 255, pp. 114B–125, Nov. 1986.

[3] R. Nevatia. "A color edge detector and its use in scene segmentation," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-7, no. 11, pp. 820–825, Nov. 1977.

[4] G. S. Robinson. "Color edge detection," *Opt. Eng.*, vol. 16, no. 5, pp. 479–484, Sep./Oct. 1977.

[5] J. Canny. "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[6] I. Pitas and A.N. Venetsanopoulos. "Edge detectors based on nonlinear filters," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, no. 4, pp. 538–550, July 1986.

[7] L. J. V. Vliet and J. T. Young. "Nonlinear laplace operator as edge detector in noisy images," *Comp. Vis. Graph. Image Proc.*, vol. 45, pp. 167–195, 1989.

[8] S. Sarkar and K.L. Boyer. "On optimal infinite impulse response edge detection filters," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, no. 11, pp. 1154–1171, Nov. 1991.

[9] A. Rosenfeld and A.C. Kak. *Digital Picture Processing.* New York: Academic, 2d ed., 1982.

[10] A. Shiozaki. "Edge extraction using entropy operator," *Comp. Vis. Graph. Image Proc.*, vol. 36, pp. 1–9, 1986.

[11] S. D. Zenzo. "A note on the gradient of a multiimage," *Comp. Vis. Graph. Image Proc.*, vol. 33, pp. 116–125, 1986.

[12] R. Machuca and K. Phillips. "Applications of vector fields to image processing," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-5, no. 3, pp. 316–329, May 1983.

[13] A. Cumani. "Edge detection in multispectral images," *CVGIP: Graphical Models and Image Processing*, vol. 53, no. 1, pp. 40–51, Jan. 1991.

[14] V. Barnett. "The ordering of multivariate data," *J. Royal Statist. Soc. A*, 139, pt. 3, pp. 318–343, 1976.

[15] R. J. Feechs and G. R. Arce. "Multidimensional morphologic edge detection," In *Proc. SPIE Conf. Visual Comm. and Image Proc.*, vol. 845, pp. 285–292, 1987.

[16] J. S. J. Lee, R. M. Haralick, and L. G. Shapiro. "Morphologic edge detection," *IEEE J.o Robot. Automat.*, vol. RA-3, no. 2, pp. 142–156, Apr. 1987.

[17] I. E. Abdou and W. K. Pratt. "Quantitative design and evaluation of enhancement/ thresholding edge detectors," *Proc. IEEE*, vol. 67, no. 5, pp. 753–763, 1979.

[18] W. K. Pratt. *Digital Image Processing.* New York: Wiley, 1991.

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

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

[21] R. D. Reiss. *Approximate Distributions of Order Statistics.* Berlin: Springer-Verlag, 1989.

[22] S. Sanwalka and A. N. Venetsanopoulos. "Vector order statistics filtering of color images," in *13th GRETSI Symp. on Signal and Image Processing*, pp. 785–788, 1991.

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

[24] J. G. Proakis. *Digital Communications.* New York: McGraw-Hill, 1989.

[25] J. T. Allen and T. Huntsberger. "Comparing color edge detection and segmentation methods," in *Proc. IEEE 1989 Southeastcon*, pp. 722–728, 1989.

[26] R. C. Hardie and G. R. Arce. "Ranking in R^p and its use in multivariate image estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 1, no. 2, pp. 197–209, June 1991.

Enhancing MLP Networks Using a Distributed Data Representation

Sridhar Narayan, Gene A. Tagliarini, and Edward W. Page

***Abstract*—Multilayer perceptron (MLP) networks trained using backpropagation can be slow to converge in many instances. The primary reason for slow learning is the global nature of backpropagation. Another reason is the fact that a neuron in an MLP network functions as a hyperplane separator and is therefore inefficient when applied to classification problems in which decision boundaries are nonlinear. This paper presents a data representational approach that addresses these problems while operating within the framework of the familiar backpropagation model. We examine the use of receptors with overlapping receptive fields as a preprocessing technique for encoding inputs to MLP networks. The proposed data representation scheme, termed ensemble encoding, is shown to promote local learning and to provide enhanced nonlinear separability. Simulation results for well known problems in classification and time-series prediction indicate that the use of ensemble encoding can significantly reduce the time required to train MLP networks. Since the choice of representation for input data is independent of the learning algorithm and the functional form employed in the MLP model, nonlinear preprocessing of network inputs may be an attractive alternative for many MLP network applications.**

I. INTRODUCTION

Multilayer perceptron (MLP) networks can be trained so that particular stimuli evoke desired responses. Algorithms such as backpropagation [1] employ samples of targeted behavior to adjust network connection weights in a manner that causes a network to learn to associate desired output patterns with specific input patterns. Once trained, MLP networks are typically capable of generalizing their response to provide correct outputs for input patterns that were not included in the training set. Neural networks trained using backpropagation have achieved notable success in applications such as sonar signal classification [2], time-series prediction [3], and process control [4].

Manuscript received July 11, 1992; revised January 19, 1995.

S. Narayan is with the Department of Mathematical Sciences, University of North Carolina, Wilmington, NC 28403 USA.

G. Tagliarini and E. Page are with the Department of Computer Science, Clemson University, Clemson, SC 29634-1906 USA.

Publisher Item Identifier S 1083-4419(96)00421-9.