IMAGE RESTORATION USING ORDER-CONSTRAINED LEAST-SQUARES METHODS

A.C. Bovik, T.S. Huang, and D.C. Munson, Jr.

Coordinated Science Laboratory, and Department of Electrical Engineering, University of Illinois
Urbana, Illinois 61801 USA

ABSTRACT

In this paper we consider two new techniques for image restoration, using order-constrained least-squares methods. The first technique consists of a cross-shaped moving window, within which two operations are combined. The first operation consists of simple hypothesis tests for monotonicity in both the horizontal and vertical directions. The second step finds the best least-squares fit of the input variates in both directions, constrained by the results of the hypothesis tests.

The second technique consists of a square moving window, again combining two operations. With the first operation, we introduce a new edge detector with specific edge height δ. Based on detection or non-detection of an edge, we either apply order-constrained least-squares methods to determine the output, or simply average. The techniques described are applied to an actual noise-corrupted image.

I. INTRODUCTION

Information-bearing edges and sharp changes in structure are important features in images. An edge may be defined as a border between two adjacent neighborhoods differing in some property such as (average) grey level or textural structure. The subjective nature of edge definition and characterization has resulted in many edge detection and enhancement schemes. Surveys and comparisons of edge detection schemes can be found in several sources [1-4].

Although many schemes have been proposed for image restoration, most are linear and deal with the spectral content of the signal or the corruptive noise. The techniques described here are novel in that the operation of the restoration algorithms depend directly on the edge or trend structure of the image. While the techniques we will introduce are both non-linear and spatially varying, they are based upon natural assumptions. We consider edge and trend detection methods that impose a simple set of linear constraints on the input variates. We then find the best least-squares fit of the input set with these constraints. First, however, we need some definitions and mathematical preliminaries basic to the later work.

Definition - Consider a finite set X. A binary relation " $\underset{\sim}{\leq}$ " on X defines a simple order on X if

1. it is reflexive: $x \underset{\sim}{\leq} x$ for $x \in X$
2. it is transitive: $x,y,z \in X$, $x \underset{\sim}{\leq} y$, $y \underset{\sim}{\leq} z$ imply $x \underset{\sim}{\leq} z$;
3. it is antisymmetric: $x,y \in X$, $x \underset{\sim}{\leq} y$, $y \underset{\sim}{\leq} x$ imply $x = y$;
4. every two elements are comparable: $x,y \in X$ implies either $x \underset{\sim}{\leq} y$ or $y \underset{\sim}{\leq} x$.

A partial order on X is reflexive, transitive, and antisymmetric, but not every pair of elements need be comparable. It should be noted that every simple order is also a partial order. We will consider a specific type of partial order in our later applications. Suppose we partition X into subsets $A_1, A_2, \ldots, A_k$ which we will call key sets. Then define a semi-partial order in the following way. All elements in each set A_i are non-comparable pairwise, but each element in A_i is identically comparable with all other elements in X, with respect to the semi-partial order, for $i=1,\ldots,k$.

Definition - A function $f:X \to R$ is isotonic with respect to an ordering (simple or partial) " $\underset{\sim}{\leq}$ " on X if $x,y \in X$, $x \underset{\sim}{\leq} y$ imply $f(x) \leqslant f(y)$.

Similarly, a function $g: X \to R$ is antitonic with respect to " $\underset{\sim}{\leq}$ " if $x,y \in X$, $x \underset{\sim}{\leq} y$ imply $g(x) \geqslant g(y)$.

Definition - Suppose that g is a real-valued function on X. An isotonic function g^* on X is an isotonic regression of g if and only if it minimizes the sum

$$\sum_{x \in X} [g(x) - f(x)]^2 \qquad (1)$$

over the class of isotonic functions f on X.

We can similarly define the antitonic regression of g. It can be shown that an isotonic (antitonic) regression of g must always exist, and that it is unique. A more general form of the above definitions can be found in [5], as well as proofs of the existence and uniqueness of the isotonic and antitonic regressions.

Suppose that we are endowed with or assume some a priori knowledge about the local ordering, or edges in the image. If we assume accurately, we should be able to improve on our estimate. The following theorem quantifies this notion [5].

Theorem - Let μ be an unknown function on a finite set X, known only to be isotonic with respect to an order (simple or partial) on X. Let g be an estimate of μ, and let g^* be the isotonic regres-

18A.3

CH1841-6/83/0000-0828 $1.00 © 1983 IEEE

sion of g. Then

$$\sum_{x \in X} [\mu(x)-g^*(x)]^2 \le \sum_{x \in X} [\mu(x)-g(x)]^2.$$

It should be noted that this theorem presupposes absolutely no knowledge of the statistical nature of g. Thus, given any estimate or restoration of a corrupted image, we can always improve on the estimate by using the isotonic regression, provided that we are supplied with accurate information regarding the relative values of the grey levels in the picture.

For later use consider order-constrained maximum likelihood (ML) parameter estimation. Suppose, for example, that we have a set of independent observations $Y=(y_1,\dots,y_n)$ and we would like to estimate the means $\tilde{M}=(\mu_1,\dots,\mu_n)$. Let us define an ordering " $\underset{\sim}{<}$ " on $\tilde{X}$, and denote the class of functions that are isotonic with respect to " $\underset{\sim}{<}$ " on X as I. If we then assume that the means $\tilde{M}$ are isotonic with respect to " $\underset{\sim}{<}$ ", we may define an order-constrained ML estimate $\hat{M}=(\hat{\mu}_1,\dots,\hat{\mu}_n)$ as

$$\hat{M} = \arg\{\max_{M:\mu \in I} f(Y|M)\}$$

where $f(Y|M)$ is the conditional density or likelihood function of Y given M. In general, this is a difficult problem to solve, depending on the nature of f. Suppose, however, that the observations are Gaussian with (equal) variances σ^2. Then,

$$\hat{M} = \arg\{\max_{M:\mu \in I} (2\Pi\sigma^2)^{-n}\exp[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu_i)^2]\}$$

which easily reduces to

$$\hat{M} = \arg\{\min_{M:\mu \in I} \sum_{i=1}^{n} (y_i-\mu_i)^2\}.$$

This is simply the definition of the isotonic regression given earlier, so that

$$\hat{M} = (\hat{\mu}_1,\dots,\hat{\mu}_n) = (\hat{\mu}_1^*,\dots,\hat{\mu}_n^*)$$

where μ^* is the isotonic regression of Y with respect to " $\underset{\sim}{<}$ ". We will use this result in Section III, where a hypothesis testing scheme is used for trend detection. We will now consider the computation of isotonic regressions for simple and semi-partial orders.

II. ALGORITHMS FOR CALCULATION

The isotonic regression with respect to a particular ordering can be found using the "Pool-Adjacent Violators" algorithm from [5]. For simple orders we proceed as follows.

Consider a function g on a finite set $X=\{x_1,\dots,x_n\}$. We wish to compute the isotonic regression g^* of g so that $g^*(x_i) \le g^*(x_j)$ when $i \le j$. The function g^* can be shown to partition X into subsets of consecutive elements of X on which it is constant; these will be referred to as solution blocks. On each of the solution blocks the value of g^* is the average of the values of g over the block. Associated with each solution block is a weight, equal to the number of elements in the block. The object of the algorithm is to find a set of solution blocks such that elements of adjacent solution blocks satisfy the simple order. An arbitrary set of consecutive elements of X will be referred to as a block, also associated with a weight. The algorithm is begun with the finest possible partition into blocks, that is, the individual elements of X. Two or more adjacent blocks whose elements do not satisfy the simple order between blocks are known as violators; these are "pooled" into a single block by taking the weighted average of the violators and assigning this value to the elements of the new block. This process is continued until the final partition of solution blocks is obtained.

A version of this algorithm due to Kruskal [6] is easily implemented numerically and was used for our applications. These algorithms can also be applied for partial and semi-partial orders, but due to space limitations we will not discuss this.

III. ISOTONIC MEDIAN FILTERS (IMF's)

We will first introduce a one-dimensional version of this scheme. The two-dimensional version is composed of a combination of one-dimensional filters.

Consider a moving filter window containing an odd number of input signal values $X=x_1,\dots,x_n$, where the filter is centered at x_m, with $m=(n+1)/2$. The filtering process consists of two steps. The first step is a trend detection scheme based on a simple hypothesis test. It is a trend detector, rather than an edge detector, since only simple orders will be considered. We will assume for simplicity that the input variates are independent and Gaussian with equal variances σ^2, and unknown means $M=\mu_1,\dots,\mu_n$ respectively. Consider the hypotheses

$$H_0: \mu_1 \le \mu_2 \le \dots \le \mu_n$$

versus

$$H_1: \mu_1 \ge \mu_2 \ge \dots \ge \mu_n$$

based on observations $x_1,\dots,x_n$. The hypotheses H_0 and H_1 assume that the μ_i are isotonic and antitonic, respectively, with respect to simple (algebraic) order.

The decision rule for this test follows, where we assume that either hypothesis is equally likely:

$$\text{choose} \begin{cases} H_0 \\ H_1 \end{cases} \text{if } \max_{M \in H_0} \log f(X|M) \underset{<}{>} \max_{M \in H_1} \log f(X|M).$$

The log of the likelihood function is maximized by the maximum likelihood estimate, which we showed in section II to be given in the order-constrained case by the isotonic (antitonic) regression of the x_i under H_0 (H_1). Hence, if we denote the isotonic (antitonic) regression of x_i to be $\overset{<}{\mu}_i$ ($\overset{>}{\mu}_i$) for $i=1,\dots,n$, then simple algebra yields the test

$$\text{choose} \begin{cases} H_0 \\ H_1 \end{cases} \text{if } \sum_{i=1}^{n} \{(x_i-\overset{>}{\mu}_i)^2 - (x_i-\overset{<}{\mu}_i)^2\} \underset{<}{>} 0.$$

18A.3

Suppose that $M^*=\mu_1^*,\ldots,\mu_n^*$ is the result of the LRT, that is, the better least-squares fit of X (either $\overrightarrow{M}=\overrightarrow{\mu}_1,\ldots,\overrightarrow{\mu}_n$ or $\overleftarrow{M}=\overleftarrow{\mu}_1,\ldots,\overleftarrow{\mu}_n$). The second step of the algorithm is quite simple: replace the center input value x_m with the median of the μ_i^*, that is, the middle value in the algebraic sense. With the median filter, the input variates must be ordered first, but here, the μ_i^* are already ordered, either by increasing or decreasing value. Hence, the filter output is simply μ_m^*.

The two-dimensional filter is defined simply: at each image pixel, we center two one-dimensional filters, one oriented horizontally, one vertically (giving a cross-shaped window). The average of the two 1-D filter outputs is then considered to be the 2-D output. Consider the image in Fig. 1, corrupted with zero-mean Gaussian noise with variance 100, as shown in Fig. 2. The images are 240 x 240 with values ranging from 0 to 256 (8 bits of resolution). Fig. 3 shows a filtered version using a 5 x 5 IMF. The noise is suppressed considerably while the edge and trend structure is not disrupted.

IV. δ-HEIGHT ISOTONIC EDGE FILTERS (δ-IEF's)

Consider an n x n square moving filter window, where n is odd. As before, we first apply a process to detect the presence or absence of an edge at the center window pixel. We separately test for vertically and horizontally oriented edge components, but we will describe the process for vertical edges only, since the horizontal case is completely analogous. We begin by partitioning the window values into two key sets $A_L=\{x_1^v,\ldots,x_m^v\}$, $A_R=\{x_{m+1}^v,\ldots,x_N^v\}$, with $m=n(n+1)/2$ and $N=n^2$. A_L contains the leftmost m window values, and A_R contains the N-m rightmost, as illustrated in Fig. 4 for n=5. So that we may specify the "height" of an edge, we introduce the δ-height edge model. Let

$$\alpha_i^v = \begin{cases} x_i^v+\delta \ ; & x_i^v \in A_L \\ x_i^v \ ; & x_i^v \in A_R \end{cases}$$

and

$$\beta_i^v = \begin{cases} x_i^v-\delta \ ; & x_i^v \in A_L \\ x_i^v \ ; & x_i^v \in A_R \end{cases}$$

Consider the hypotheses

$$H_o^v:\ \alpha_1^v,\ldots,\alpha_m^v \le \alpha_{m+1}^v,\ldots,\alpha_N^v$$

$$H_1^v:\ \beta_1^v,\ldots,\beta_m^v \ge \beta_{m+1}^v,\ldots,\beta_N^v$$

$$H_2^v:\ x_1^v = \ldots = x_N^v$$

where we again assume additive white background noise. Further, let $f_0^v(i)$ and $f_1^v(i)$ denote the isotonic and antitonic regressions of the α_i^v and β_i^v respectively, with respect to the semi-partial ordering " $\lesssim$ " defined by $x_i^v \in A_L$, $x_j^v \in A_R$ imply $x_i^v \lesssim x_j^v$. With hypothesis H_2^v we associate the ML estimate $f_2^v(i) = \frac{1}{N}\Sigma_{j=1}^N x_j^v$, i=1,...,N. The test is based on the errors

$$e_0^v = \frac{1}{N}\ \Sigma_{i=1}^N\ \{f_0^v(i) - \alpha_i^v\}^2,$$

$$e_1^v = \frac{1}{N}\ \Sigma_{i=1}^N\ \{f_1^v(i) - \beta_i^v\}^2,$$

$$e_2^v = \frac{1}{N}\ \Sigma_{i=1}^N\ \{f_2^v(i) - x_i^v\}^2.$$

The decision rule is then

$$\text{choose } H_j^v \text{ if } j=\arg\{\min_{\ell\in\{0,1,2\}} e_\ell^v\},$$

and we associate with our choice the minimizing function

$$f_*^v(i) = \begin{cases} f_j^v(i) + c, & i: x_i^v \in A_L \\ f_j^v(i), & i: x_i^v \in A_R \end{cases}$$

where c=-δ if j=0, c=δ if j=1, and c=0 if j=2.

We define a completely analogous operation based on upper and lower key sets A_T and A_B to obtain a horizontally oriented minimizing function f_*^h. We complete the filtering operation by defining $f^*=(f_*^v+f_*^h)/2$ and by replacing the center window value with the corresponding value of f* in the filtered version. Fig. 5 illustrates a filtered version of the image using a 3 x 3 δ-IEF with δ=20, while Fig. 6 shows the effect of a 5 x 5 δ-IEF with δ=30. The noise is effectively removed while preserving the edges and considerable detail, but most notably the selectivity of the filter allows averaging at appropriate pixels, producing very smooth flat regions. In Table I, below, the empirical mean-squared error

$$e = \frac{1}{K^2}\ \Sigma_{i=1}^K\ \Sigma_{j=1}^K\ \{F(i,j)-O(i,j)\}^2$$

is given, where O and F are the original and filtered image pixel values respectively, and K = 240, for two window sizes and several values of δ. Values of e for square median and averaging filters are given for comparison. Some of the new δ-IEF's obviously perform well.

TABLE I

Filter	Window Size	δ	e
δ-IEF	3x3	0	72.6
δ-IEF	3x3	10	60.6
δ-IEF	3x3	20	37.1
δ-IEF	3x3	30	26.7
δ-IEF	5x5	0	57.3
δ-IEF	5x5	15	34.3
δ-IEF	5x5	30	22.3
δ-IEF	5x5	40	22.1
Averager	3x3	--	74.0
Averager	5x5	--	97.8
Median	3x3	--	72.7
Median	5x5	--	71.8

REFERENCES

[1] W.K. Pratt, Digital Image Processing. New York: John Wiley and Sons, 1978.

[2] A. Rosenfeld and A.C. Kak, Digital Picture Processing. New York: Academic Press, 1976.

18A.3

[3] L.S. Davis, "A Survey of Edge Detection Techniques," Comput. Graphics Image Processing, vol. 4, pp. 248-270, 1975.

[4] T. Peli and D. Malah, "A Study of Edge Detector Algorithms," Comput. Graphics Image Processing, vol. 20, pp. 1-21, 1982.

Fig. 1 Original Image.

Fig. 2 Image with additive white N(0,100) noise.

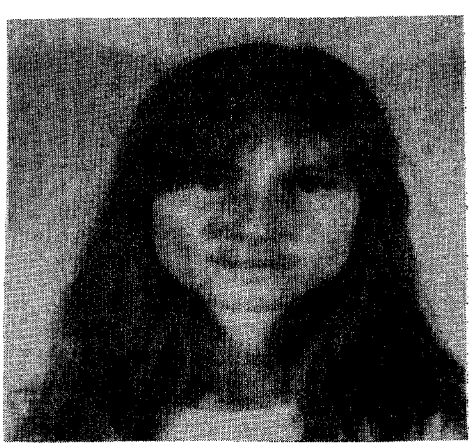

Fig. 3 Filtered image using 3x3 IMF.

[5] R.E. Barlow, D.J. Bartholomew, J.M. Bremner, and H.D. Brunk, Statistical Inference Under Order Restrictions. New York: John Wiley and Sons, 1972.

[6] J.B. Kruskal, "Nonmetric Multidimensional Scaling: A Numerical Method," Psychometrika, vol. 29, pp. 115-129, 1964.

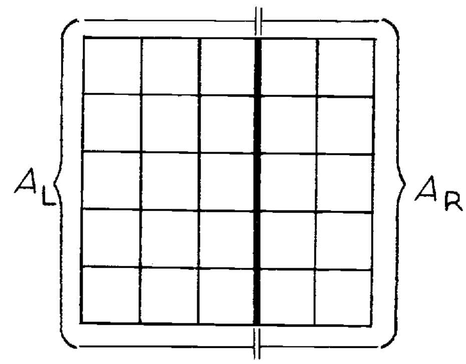

Fig. 4 Configuration for 5x5 δ-IEF for detecting vertical edges.

Fig. 5 Filtered image using 3x3 δ-IEF with δ=20.

Filtered image using 5x5 δ-IEF with δ=30.