Reprinted from

Ninth International Symposium

Machine Processing of

Remotely Sensed Data

with special emphasis on

Natural Resources Evaluation

June 21-23, 1983

Proceedings

Purdue University
The Laboratory for Applications of Remote Sensing
West Lafayette, Indiana 47907 USA

EDGE AND LINEAR FEATURE ENHANCEMENT BY KRIGING FILTERING

K. CONRADSEN, G. NILSSON

The Technical University of Denmark/
Department of Mathematical Statistics
and Operations Research
Lyngby, Denmark

ABSTRACT

A two dimensional filter based on the difference between a pixel value and a predicted value for the pixel is proposed. The prediction is based on values from different subsets of the image. The prediction is based on a random field model for the image, and it is of kriging type, i.e. the predicted value is linear in the remaining pixel values and has minimum variance among such estimates. By choosing the predictor subset in a suitable way one can obtain reasonably good enhancements of edges and linear features.

I. INTRODUCTION

A major problem in analyses of Landsat imagery is the enhancement of linear features, especially in geological applications where the interesting linear features very often are subtle or fuzzy. Therefore edge enhancement techniques based on e.g. gradients or on "usual" high pass filtering will often give unsatisfactory results.

The basic idea behind the filters proposed in this paper is the following. A linear feature in an image represents some kind of irregularity with respect to the surrounding pixels. A measure of the magnitude of irregularity at a specific point is the difference between the actual value for that point minus an interpolated value based on the remaining pixel values. In order to obtain an interpolation that takes the random nature of a Landsat scene into account one can model the scene as the outcome of a stationary random field and then determine the MMSE (Minimum Mean Squared Error) solution to the problem. If the interpolation is based on all the surrounding pixels there will very often be a close connection between the actual and the interpolated value, especially in cases where edges and troughs are "developing" gradually. In such cases it will be an advantage only to use pixels that had a distance greater than a threshold value in the prediction. This is illustrated in Table 1, where three signals with troughs of different widths have been filtered with two filters. The first filter corresponds to the Laplacian

TABLE 1. Two Simple Edge Enhancement Filters. Three signals (~ s) showing troughs of different widths, and the resulting signals after applying central filters with coefficients (-1, 2, -1) (~ A) and (-1, 0, 2, 0, -1) (~B).

s	3	3	3	3	3	0	3	3	3	3	3
As	0	0	0	0	3	-6	3	0	0	0	0
Bs	0	0	0	3	0	-6	0	3	0	0	0
s	3	3	3	3	1½	0	1½	3	3	3	3
As	0	0	0	1½	0	-3	0	1½	0	0	0
Bs	0	0	1½	3	0	-6	0	3	1½	0	0
s	3	3	3	2	1	0	1	2	3	3	3
As	0	0	1	0	0	-2	0	0	1	0	0
Bs	0	1	2	1	-2	-4	-2	1	2	1	0

and estimates a second derivative. It is given by

$$As_i = -s_{i-1} + 2s_i - s_{i+1} .$$

The second filter has weights with the same numerical values, but does not use the immediate neighbors. Instead values with lags 2 are used, i.e.

$$Bs_i = -s_{i-2} + 2s_i - s_{i+2} .$$

We see that the advantage of the second filter is that the wider troughs are detected better. The disadvantage is that there is a stronger ringing effect. In the sequel we shall construct the interpolations based on "remote" pixels taking the spatial correlation into account. In order to distinguish this method, where the interpolation constants use the randomness present, from other schemes with fixed coefficients (e.g. the Laplacian) we shall use the term prediction instead of interpolation.

II. PREDICTION

This section contains the necessary results on MMSE prediction of correlated variables. Some of the results can be found in the literature on multivariate statistical analysis, cf. Anderson[1], others correspond to results on kriging given in the geostatistical literature, cf. Journel and Huijbregts[2].

We consider a random variable Y that is correlated with the n-dimensional random variable $\underline{X}$ (in the sequel a $\_$ corresponds to a vector and a $=$ to a matrix). We assume that the mean and dispersion matrix (variance-covariance matrix) of the combined variable $(Y,\underline{X}')'$ (a $'$ corresponds to the transpose) is

$$E\begin{pmatrix}\begin{pmatrix} Y \\ \underline{X} \end{pmatrix}\end{pmatrix} = \begin{pmatrix} \mu \\ \mu \; \underline{1} \end{pmatrix}$$

$$D\begin{pmatrix}\begin{pmatrix} Y \\ \underline{X} \end{pmatrix}\end{pmatrix} = \begin{pmatrix} \delta^2 & \underline{\sigma}' \\ \underline{\sigma} & \underline{\underline{\Sigma}} \end{pmatrix} .$$

Here $\underline{1}$ denotes an n-vector of ones. With this setup we have the following

<u>Theorem 1</u>. The MMSE <u>affine</u> prediction function for Y based on $\underline{X}$ is

$$\hat{Y}_1 = \mu(1-c) + \underline{\sigma}' \underline{\underline{\Sigma}}^{-1} \underline{X} ,$$

where

$$c = \underline{\sigma}' \underline{\underline{\Sigma}}^{-1} \underline{1}$$

is the sum of the elements in $\underline{\sigma}'\underline{\underline{\Sigma}}^{-1}$. $\hat{Y}_1$ is unbiased. The unbiased linear MMSE prediction function is

$$\hat{Y}_2 = \left(\underline{\sigma}' + \frac{1-c}{d}\underline{1}'\right) \underline{\underline{\Sigma}}^{-1} \underline{X} ,$$

where

$$d = \underline{1}' \underline{\underline{\Sigma}}^{-1} \underline{1}$$

is the sum of the elements in $\underline{\underline{\Sigma}}^{-1}$.

<u>Proof</u>. For the sake of clarity we shall sketch the procedure for $\hat{Y}_2$. The mean of a predictor $\underline{a}'\underline{X}$ is $\mu\underline{a}'\underline{1}$, and unbiasedness thus requires

$$\underline{a}'\underline{1} = 1 ,$$

i.e. the sum of the coefficients a_i must equal 1. Using the unbiasedness the mean squared error of $\underline{a}'\underline{X}$ is

$$E(Y-\underline{a}'\underline{X})^2 = V(Y-\underline{a}'\underline{X})$$

$$= (1 \;\; -\underline{a}') \begin{pmatrix} \delta^2 & \underline{\sigma}' \\ \underline{\sigma} & \underline{\underline{\Sigma}} \end{pmatrix} \begin{pmatrix} 1 \\ -\underline{a} \end{pmatrix}$$

$$= \delta^2 - 2\underline{a}'\underline{\sigma} + \underline{a}'\underline{\underline{\Sigma}}\underline{a} .$$

In order to minimize this expression - still under the constraint $\underline{a}'\underline{1} = 1$ - we introduce a Lagrange multipler λ and obtain

$$F(\underline{a},\lambda) = \delta^2 - 2\underline{a}'\underline{\sigma} + \underline{a}'\underline{\underline{\Sigma}}\underline{a} + \lambda(\underline{a}'\underline{1} - 1) .$$

Differentiation gives the equations

$$\frac{\partial F}{\partial \underline{a}} = -2\underline{\sigma} + 2\underline{\underline{\Sigma}}\underline{a} + \lambda\,\underline{1} = \underline{0}$$

$$\frac{\partial F}{\partial \lambda} = \underline{a}'\underline{1} - 1 = 0 .$$

Solving these gives the desired result.

A simple corollary to the theorem is

<u>Corollary</u>. For c close to 1 the unbiased predictor

$$\hat{Y}_3 = \frac{1}{c}\underline{\sigma}' \underline{\underline{\Sigma}}^{-1} \underline{X} .$$

is an approximation to the MMSE linear predictor.

Proof. Trivial.

The properties of the predictors can be stated somewhat stronger if we assume normality of the random variables. E.g. Y_1 will be MMSE among all prediction functions, not only among all affine.

III. OI-FILTER

In this section we shall define the new filters. They are using a prediction of a single pixel value $X_{0,0}$ based on $(2p+1)^2 - (2q+1)^2$ pixel values with a "distance" of at least q from $X_{0,0}$, cf. Table 2. We call such a filter an OI(p,q) filter as a mnemonic abbreviation for outer box of size $2p+1$ and inner box of size $2q+1$. The boxes defining the predictor sets are squares, but this has only been done for notational reasons.

$$Y = X_{0,0}$$

$$X' = (X_{p,-p}, \ldots, X_{p,-p}, \ldots, X_{p,-q}, \ldots, X_{q+1,-q}, X_{-q-1,-q}, \ldots, X_{-p,p})' ,$$

i.e. Y is the central value in Table 2 and $\underline{X}$ consists of all the elements between the two frames organized in a $(2p+1)^2 - (2q+1)^2$ vector.

With these definitions it is now possible to use the predictors in Theorem 1 and the corollary.

For small values of p and q one can easily write down the dispersion matrix $\underline{\underline{\Sigma}}$ for $\underline{X}$ that is used in the computation of the predictor coefficients. For larger values, however, the following remarks can be useful in the design of a computer program for calculating the values. More specifically we shall write $\underline{\underline{\Sigma}}$ as a block matrix $(\underline{\underline{\Sigma}}_{ij})$, $i=1, \ldots, 2p+1$, $j=1, \ldots, 2p+1$.

TABLE 2. Predictor Variables. The observations (pixel values) between the two frames are used in predicting the observation $X_{0,0}$ in the OI(p,q)-filter.

$X_{p,-p}$	...	$X_{p,-q}$	...	$X_{p,0}$	...	$X_{p,q}$	...	$X_{p,p}$
⋮		⋮		⋮		⋮		⋮
$X_{q,-p}$	...	$X_{q,-q}$	...	$X_{q,0}$	...	$X_{q,q}$	...	$X_{q,p}$
⋮		⋮		⋮		⋮		⋮
$X_{0,-p}$	...	$X_{0,-q}$	...	$X_{0,0}$	...	$X_{0,q}$	...	$X_{0,p}$
⋮		⋮		⋮		⋮		⋮
$X_{q,-p}$	...	$X_{-q,-q}$	...	$X_{-q,0}$	...	$X_{-q,q}$	...	$X_{-q,p}$
⋮		⋮		⋮		⋮		⋮
$X_{-p,-p}$	...	$X_{-p,-q}$	...	$X_{-p,0}$	...	$X_{-p,q}$	...	$X_{-p,p}$

We now consider the image as the outcome of a stationary random field with covariance function

$$\gamma(i,j) = E(X_{s,t} - \mu)(X_{s+i,t+j} - \mu) ,$$

where

$$\mu = E(X_{s,t})$$

is the common mean. We define

For $s = 0, \ldots, 2p$ we define the $(2p+1) \times (2p+1)$ Toeplitz matrix $\underline{\underline{C}}_s$ by putting the elements equal to

$$c_{\mu\nu} = \gamma(-|\mu-\nu|, s), \quad \mu,\nu = 1, \ldots, 2p+1 .$$

We partition the integers

$$I = \{1, \ldots, 2p+1\}$$

into the sets

$$I_1 = \{1, \ldots, p-q\}$$

$$I_2 = \{p-q+1, \ldots, p+q+1\}$$

$$I_3 = \{p+q+2, \ldots, 2p+1\} .$$

For $i \in I_1 \cup I_3$, $j \in I_1 \cup I_3$, $i \leq j$, we simply have that the $(2p+1) \times (2p+1)$ matrix $\underline{\underline{\Sigma}}_{ij}$ is given by

$$\underline{\underline{\Sigma}}_{ij} = C_{j-i} .$$

For $i \in I_1$, $j \in I_2$, we obtain the $(2p+1) \times (2p-2q)$ matrix $\underline{\underline{\Sigma}}_{ij}$ by deleting columns no. ν, $\nu \in I_2$, from $\underline{\underline{C}}_{j-i}$. For $i \in I_2$, $j \in I_2$, $i \leq j$, we obtain the $(2p-2q) \times (2p-2q)$ matrix $\underline{\underline{\Sigma}}_{ij}$ by deleting columns no. ν, $\nu \in I_2$, and rows no. μ, $\mu \in I_2$, from $\underline{\underline{C}}_{j-i}$. For $i \in I_2$, $j \in I_3$, we obtain the $(2p-2q) \times (2p+1)$ matrix $\underline{\underline{\Sigma}}_{ij}$ by deleting rows no. μ, $\mu \in I_2$, from $\underline{\underline{C}}_{j-i}$. Finally we have

$$\underline{\underline{\Sigma}}_{ij} = \underline{\underline{\Sigma}}'_{ji} \quad \text{for} \quad i \geq j .$$

We write the vector $\underline{\sigma}$ as

$$\underline{\sigma}' = (\underline{\sigma}'_1, \ldots, \underline{\sigma}'_{2p+1}) .$$

We define

$$\underline{c}'_i = (d_1, \ldots, d_{2p+1})$$

by putting

$$d_\mu = \gamma(p+1-\mu, i-p-1), \mu=1, \ldots, 2p+1 .$$

Now, for $i \in I_1 \cup I_3$ we have $\underline{\sigma}'_i = \underline{c}'_i$. For $i \in I_2$ we obtain the $(2p-2q)$ row vector $\underline{\sigma}'_i$ by deleting the elements d_μ, $\mu \in I_2$.

With the notation given above we now have the following

Definition 1. An $OI(p,q)$ filter is given by

$$X_{0,0} - \hat{Y}_i ,$$

where $\hat{Y}_i$ is one of the predictions of $X_{0,0}$ obtained by applying Theorem 1 or the corollary.

IV. RESULTS

The OI-filters have been applied on several Landsat scenes in order to enhance linear features. As the true correlations are not known, it has been necessary to estimate these. This has been done by ordinary mean product deviations, i.e.

$$\hat{\gamma}(i,j) = \frac{1}{N} \sum_r \sum_s (x_{r,s}-\bar{x})(x_{r+i,s+j}-\bar{x}) ,$$

where

$$\bar{x} = \frac{1}{N} \sum_r \sum_s x_{r,s} ,$$

and N is the number of observations used in the estimations.

In Table 3 are given estimated correlations based on 400×400 sample values for Landsat MSS channel 7. The area, where the measurements are taken, is the southwestern part of Jameson Land in East Greenland. The scene has been corrected geometrically and the pixel size is 50×50 meters (app.). It is seen that the correlations do not tail off very rapidly. Therefore the neighboring pixels contain much information on the individual pixels.

In Table 4 are given the resulting filter weights for two OI-filters based on the predictor given in the corollary. It is seen that there are major differences between the weights in the two cases (which of course is not surprising).

It has turned out that filters of the form $OI(5,1)$ - $OI(5,3)$ give rather good results. In Figure 1 is compared a standard Laplacian filtering and an $OI(5,I)$ filtering of an area around Igaliko in South Greenland. In both cases the upper 15% of the distributions of the filtered images have been marked black, the rest white. There are substantial differences between the two, and several structures are more easily recognized in the OI-filtered image. We therefore conclude that these filters represent a useful supplement to other line- and edge-enhancement filters.

FIGURE 1. Laplacian and OI(5, 1) Filters. The area analyzed is around Igaliko in South Greenland.

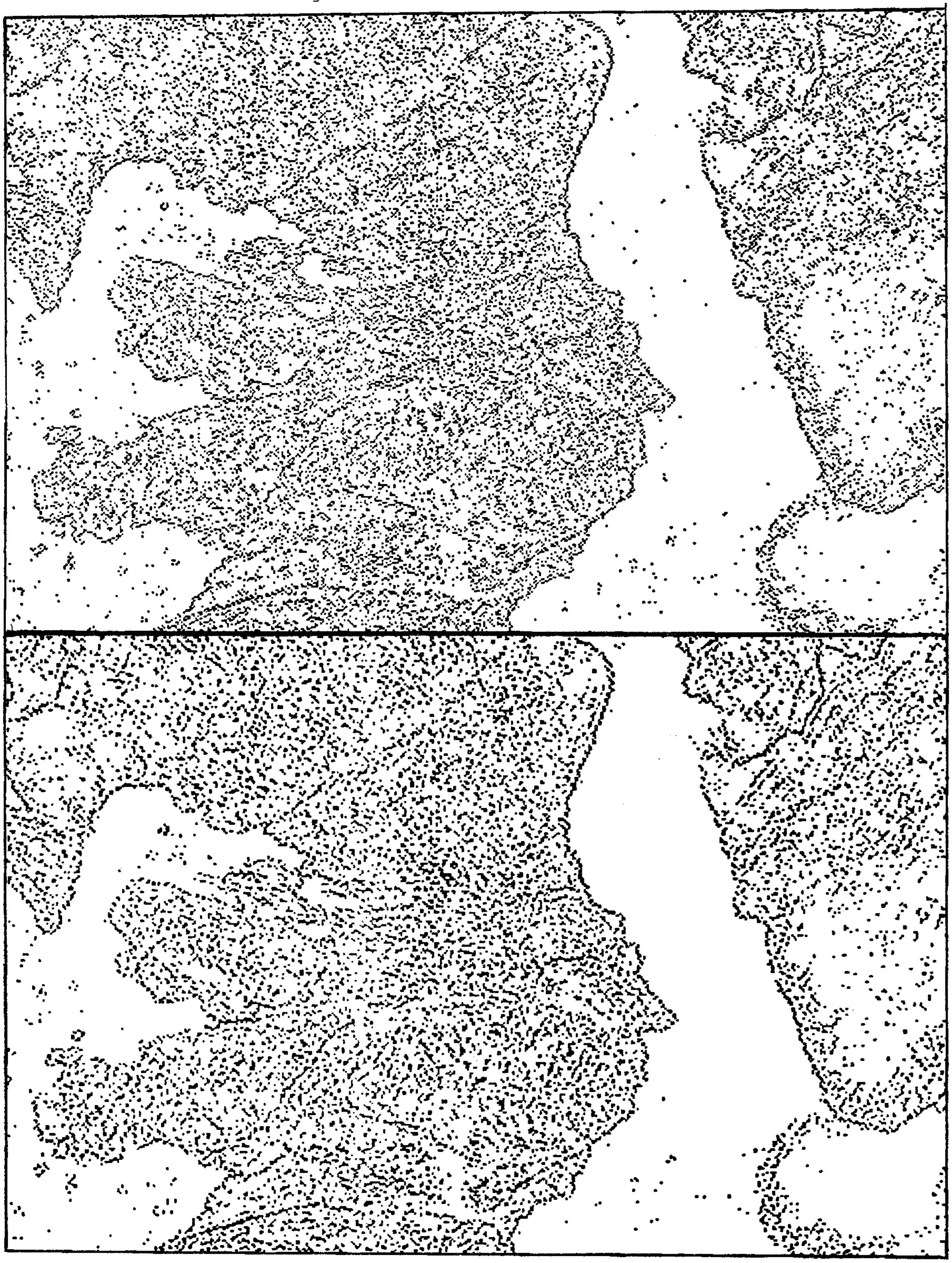

TABLE 3. Correlation Function. Part of the estimated correlation function for Landsat MSS channel 7 based on 400 × 400 pixels from Jameson Land, East Greenland. The scene has been corrected geometrically, and the pixel size is (app.) 50 × 50 meters.

i \ j	0	1	2	3	4	5	6
6	.15	.18	.21	.24	.24	.24	.23
5	.20	.24	.27	.29	.28	.27	.24
4	.28	.32	.35	.35	.33	.29	.26
3	.39	.42	.44	.42	.37	.32	.27
2	.55	.60	.55	.48	.41	.34	.28
1	.77	.73	.65	.52	.42	.35	.28
0	1.00	.81	.64	.53	.41	.32	.26
-1	.77	.69	.55	.43	.35	.29	.24
-2	.55	.48	.40	.33	.28	.23	.20
-3	.39	.33	.28	.24	.21	.18	.15
-4	.28	.23	.19	.17	.14	.12	.11
-5	.20	.16	.13	.11	.09	.08	.07
-6	.15	.11	.09	.07	.06	.05	.05

TABLE 4. Filter Weights. The filter weights (× 100) for an OI(2,0)- and an OI(2,1)-filter based on the correlation function given in Table 3.

i \ j	-2	-1	0	1	2	-2	-1	0	1	2
-2	2	3	10	-11	7	8	-1	-4	-18	7
-1	1	-14	-19	-1	-6	-10	0	0	0	-17
0	9	-31	100	-31	9	-16	0	100	0	-16
1	-6	-1	-19	-14	1	-17	0	0	0	-10
2	7	-11	10	3	2	7	-18	-4	-1	8

V. ACKNOWLEDGEMENTS

The present work has been supported by grants from the Commission of the European Communities, The Danish Science Research Foundation, and Otto Monsteds Foundation.

VI. REFERENCES

1. Anderson, T. W. (1958): An Introduction to Multivariate Statistical Analysis. John Wiley & Sons, 374 pp.

2. Journel, A. and Huijbregts, C. (1978): Mining Geostatistics, Academic Press, 600 pp.

Knut Conradsen received his cand. scient. degree in mathematics and statistics from the University of Copenhagen in 1970. Since 1970 he has been at the Department of Statistics and Operations Research (IMSOR) at the Technical University of Denmark, from 1974 as Associate Professor. He visited Stanford University in the academic year 1982-83.

Gert Nilsson received his fil. dr. degree in theoretical physics from the University of Lund in 1980. Since 1981 he has worked as a computer scientist at the Department of Statistics and Operations Research (IMSOR) at the Technical University of Denmark.