Communications

Formulation of Forward–Backward Method Using Novel Spectral Acceleration for the Modeling of Scattering from Impedance Rough Surfaces

Hsi-Tseng Chou and Joel T. Johnson

***Abstract*—F-BM/NSA [3] with computational complexity of $O(N)$ is very efficient in method of moment (MoM) modeling of large-scale scattering problems from rough surfaces. The previous formulation for PEC surfaces is here extended to treat impedance surfaces. Similarly, numerical experiment shows F-BM/NSA is far more efficient than the competitive BMIA/CAG [4] in the order of magnitude.**

***Index Terms*—Forward-backward method, method of moment (MoM), rough surfaces, scattering.**

I. Introduction

Iterative forward–backward method (F-BM) [1], [2] with novel spectral acceleration (F-BM/NSA) [3] has been shown to be very efficient in solving the method of moment (MoM) matrix equations for the large-scale scattering problems from the one-dimensional (1-D) rough surface. F-BM requires only three or four iterations to obtain a very accurate solution in many cases of practical applications, and the NSA algorithm reduces the computational complexity from $O(N^2)$ in the original F-BM approach to $O(N)$, where N is the number of unknowns in MoM modeling. This is true in the operational count of both the matrix-vector multiplication and memory storage as the surface length increases for a fixed frequency. In the previous paper [3], the algorithm was formulated for perfectly electrical conducting (PEC) rough surfaces. The formulations of an NSA algorithm for impedance rough surfaces are presented in this paper. The numerical results will be compared with a competitive method of the banded-matrix iterative approach/canonical grid (BMIA/CAG) [4]. Similar to the PEC cases, it is found that the CPU time required is also a very small portion of that obtained from BMIA/CAG.

II. Formulation

A large-scale problem of scattering from a 1-D rough surface such as an ocean-like surface is illustrated in Fig. 1, in which $z = f(x)$ denotes the impedance surface profile. The MoM matrix equation in terms of electrical field E for a tapered TM_z plane wave incidence (E^i) on the surface is given by [5]

$$\overline{\overline{Z}} \cdot \overline{I} = \overline{V}. \tag{1}$$

Approximated expressions for the elements of $\overline{\overline{Z}}$ are [5] shown in (2) (shown at the bottom of the next page), where γ is the Euler number 0.577 216, ϵ_0 and ϵ_1 are the permittivities of the nonmagnetic regions above and below the surface profile respectively, $k = \omega\sqrt{\mu_0\epsilon_0}$ is the propagation constant in free space, and $\vec{\rho}_n$ denotes the position vector

Manuscript received January 27, 1999; revised May 4, 1999.

H.-T. Chou is with the Department of Electrical Engineering,Yuan-Ze University, Chung-Li 320, Taiwan, R.O.C., (e-mail: hchou@saturn.yzu.edu.tw).

J. T. Johnson is with the Department of Electrical Engineering, Electro-Science Laboratory, The Ohio State University, Columbus, OH 43210 USA.

Publisher Item Identifier S 0196-2892(00)00404-6.

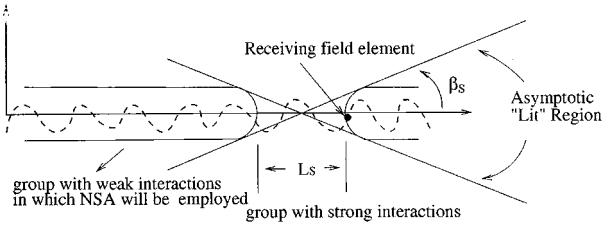

Fig. 1. Impedance rough-surface profile $z = f(x)$ and geometric interpretation of asymptotic "lit" region along forward sweeps. L_s is the distance within which the source elements have strong interaction with the receiving element.

at the center of the n^{th} element. In (2), $g(\vec{\rho}_n, \vec{\rho}_m)$ is the Green's function given by

$$g(\vec{\rho}_n, \vec{\rho}_m) = \frac{j}{4}H_0^{(1)}(k|\vec{\rho}_n - \vec{\rho}_m|) \tag{3}$$

where $H_0^{(1)}$ is the zeroth order Hankel function of the first kind. The elements of $\overline{V}$ and $\overline{I}$ are given by $V_n = E^i(\vec{\rho}_n)$ and $I_n = (\partial E(\vec{\rho}_n)/\partial n)$. In (1), $\overline{I} = [I_n]$ is the unknown matrix to be found. The iterative forward-backward sweep procedure described in [3] can be implemented directly here to solve the induced current in (1). Like the PEC case, numerical experiments have shown that usually only three or four iterations are sufficient to obtain accurate results. However, F-BM requires repeated computation of

$$E_f(\vec{\rho_n}) = \sum_{m=1}^{n-1} Z_{nm}I_m \tag{4}$$

and

$$E_b(\vec{\rho_n}) = \sum_{m=n+1}^{N} Z_{nm}I_m \tag{5}$$

which denote the forward and backward radiation by the source current elements in the front and in the rear of the receiving nth element, respectively, and result in an $O(N^2)$ operational count in each iteration. In this paper, we take E_f as an example and extend the NSA algorithm previously formulated for the PEC cases in [3] to fast compute (4) and thus reduce the operational count and memory storage to $O(N)$. The acceleration on the computation of E_b in (5) can be treated in the same fashion.

Similar to the development in [3], the acceleration algorithm starts with the decomposition of E_f in terms of strong interaction contribution E_s and weak interaction contribution E_w, respectively. E_s is the field radiated from the strong interaction source group, which is selected within neighborhood distance L_s of receiving nth element. It is noted that in most of the practical cases, L_s is a very small portion of the surface length and remains fairly fixed for a given roughness of surfaces regardless of the surface length, and therefore, E_s is found via the conventional exact computation. On the other hand, E_w, which is radiated from the source group selected outside the strong interaction source group, becomes important when the angle of wave incidence is near grazing or if one is interested in backscattered field. The computation of E_w becomes most time consuming for a CPU if it is performed

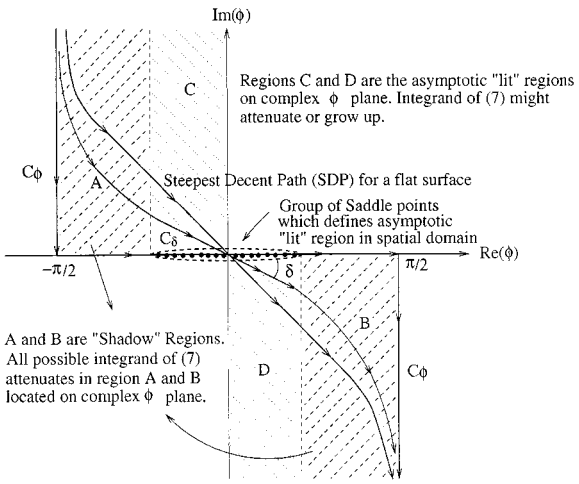

Fig. 2. Integration contour of $H_0^{(1)}$ on the complex ϕ plane. C_ϕ is the original contour, and C_δ is the deformed contour.

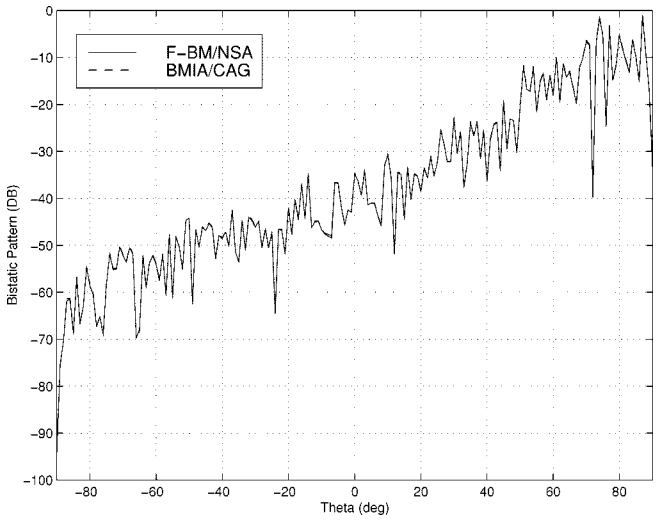

Fig. 3. Bistatic scattering pattern of a 4096λ-long rough surface (rms height 2.27λ) illuminated by a tapered field with incident angle of 85°.

via the exact computation, and makes F-BM very inefficient. The computation of E_w can now be accelerated via the novel spectral-expansion technique. One first expresses E_w in (4) as

$$E_w(\vec{\rho_n}) = \sum_{m=1}^{n-N_s-1} I_m \left\{ -\frac{j}{k}\sqrt{\frac{\epsilon_0}{\epsilon_1}}\frac{\partial g(\vec{\rho}_n, \vec{\rho}_m)}{\partial n} + g(\vec{\rho}_n, \vec{\rho}_m) \right\} \sqrt{1+(\frac{\partial f(x_m)}{\partial x_m})^2}\,\Delta x \tag{6}$$

where N_s ($=L_s/\Delta x$) denotes the number of elements that have strong interaction with the nth element. Employing the spectral representation of Green's function $g(\vec{\rho}_n, \vec{\rho}_m)$ in (3) gives

$$g(\vec{\rho}_n, \vec{\rho}_m) = \frac{j}{4\pi}\int_{C_\phi} e^{jk[(x_n-x_m)\cos\phi+(z_n-z_m)\sin\phi]}\,d\phi \tag{7}$$

where in (7), the contour of integration C_ϕ is shown in Fig. 2. Substituting (7) into (6) gives

$$E_w(\vec{\rho_n}) = \frac{j\Delta x}{4\pi}\int_{C_\phi} F_n(\phi)e^{jkz_n\sin\phi}\,d\phi \tag{8}$$

where $F_n(\phi)$ is related to far-field pattern (or plane-wave spectrum) of the weak interaction group with the reference point selected at $(x_n, 0)$, and it can be found by a recursive procedure

$$F_n(\phi) = F_{n-1}(\phi)e^{jk\Delta x\cos\phi} + I_{(n-1-N_s)}\left\{\sqrt{\frac{\epsilon_0}{\epsilon_1}}\left[-\sin\phi + \left(\frac{\partial f(x_m)}{\partial x_m}\right)\cos\phi\right] + \sqrt{1+\left(\frac{\partial f(x_m)}{\partial x_m}\right)^2}\right\} \cdot e^{jk(N_s+1)\Delta x\cos\phi}e^{-jkz_{n-1-N_s}\sin\phi}. \tag{9}$$

It is noted that the far-field pattern in (9) is independent of the receiving element except the reference point is selected at x_n in order to avoid numerical error. The plane-wave expansion relates the contribution of source elements' radiation to the receiving element. It is noted that far-field pattern $F_n(\phi)$ will continuously build up via (9), while a new source element is born in the weak interaction group as the receiving element sweeps in the forward direction. On the other hand, the radiation contribution to the receiving element will only rely on the far field pattern of the direct radiation contribution from the source elements as described in (8). In practice, it is desirable to employ a new integration contour C_δ instead of C_ϕ in (8) because the far field pattern $F_n(\phi)$ in real ϕ space tends to have a narrow main lobe and many narrow side lobes for a large surface, in contrast to a slowly varying pattern along C_δ. The high efficiency of this algorithm relies on the fact that for a slowly varying far-field pattern, one may integrate (8) numerically with a constant sampling rate according to the roughness of the surface and regardless of the surface length. The criterion in selecting C_δ is described in [3] and is indicated in Fig. 2, where $\delta = \tan^{-1}(1/b)$ with $b = \max[\sqrt{kR_s/20}\cdot\phi_s - 1, 1]$, $\phi_s = \tan^{-1}[(z_{\max}-z_{\min})/R_s]$, and $R_s = \sqrt{L_s^2+(z_{\max}-z_{\min})^2}$. The integral of (8) over ϕ is discretized into $2Q+1$ plane-wave directions and mapped to the real axis according to the complexity of the far-field pattern in (9). The mapping in the "lit" region, where $|\mathrm{Re}(\phi)| \le \beta_s$ as indicated in Fig. 1, is given by

$$d\phi \to \Delta\phi e^{-j\delta} \tag{10}$$

and

$$\phi \to \phi_p = p\Delta\phi e^{-j\delta}. \tag{11}$$

However, it is noted that along C_δ, outside the asymptotic lit region, the plane waves are highly evanescent and cancel their contribution via fast oscillation. Therefore, the integrand in (8) may be windowed with

$$Z_{nm} = \begin{cases} \dfrac{j}{2k}\sqrt{\epsilon_0}+\dfrac{i\Delta x}{4}\left\{1+\dfrac{2j}{\pi}\left[\ln\left(\dfrac{k\Delta x}{4e}\sqrt{1+(df(x_n)/dx_n)^2}\right)+\gamma\right]\right\}\sqrt{1+(df(x_n)/dx_n)^2}, & n=m \\ \left\{-\dfrac{i}{k}\sqrt{\dfrac{\epsilon_0}{\epsilon_1}}\dfrac{\partial g(\vec{\rho_n},\vec{\rho_m})}{\partial n}+g(\vec{\rho_n},\vec{\rho_m})\right\}\sqrt{1+\left(\dfrac{\partial f(x_m)}{\partial x_m}\right)^2}\Delta x, & n\neq m \end{cases} \tag{2}$$

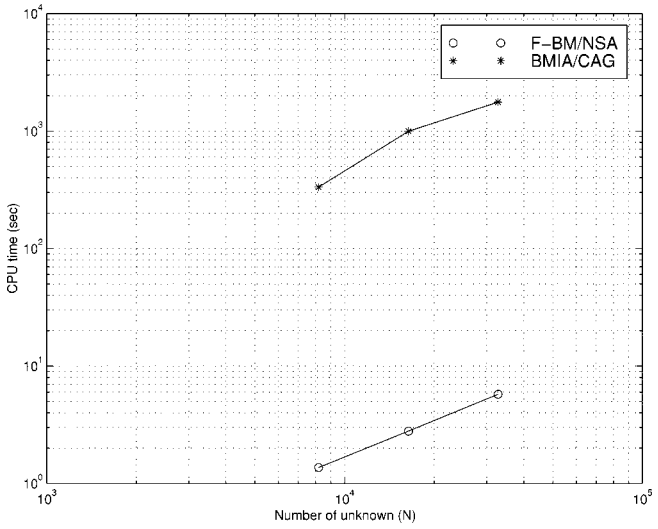

Fig. 4. CPU time requirement to solve a matrix equation for the induced current over rough surfaces (rms height 2.27λ and $\lambda/8$ sampling segment) illuminated by a tapered field with incident angle of 85°.

a filter that is flat over the lit region and tapers to zero in the "shadow" region ($|\mathrm{Re}(\phi_p)| > \beta_s$) to filter out plane waves that would otherwise be attenuated via cancellation. The windowed form of (8) is given by

$$E_w(\vec{\rho_n}) = \frac{j\Delta x e^{-j\delta}}{4\pi} \sum_{p=-Q}^{Q} W(\phi_p) F_n(\phi_p) e^{jkz_n \sin\ \phi_p} \Delta\phi \quad (12)$$

where the window function $W(\phi_p)$ is unity for $|\mathrm{Re}(\phi_p)| \leq \beta_s$ and tapers smoothly to zero over a fixed number (usually five on each side) of oscillations for $|\mathrm{Re}(\phi_p)| > \beta_s$. This reduces the number of plane waves required to be included in the current approach. Therefore, Q and $\Delta\phi$ can be found [3] by

$$Q = \frac{\beta_s}{\Delta\phi} + 5 \quad (13)$$

and

$$\Delta\phi = \sqrt{\frac{5}{kR_s}}/22. \quad (14)$$

It is noted that $\Delta\phi$ depends only on the roughness of the surface and the size of segment selected for the strong interaction source group. Also β_s is selected according to the maximum roughness of the surface so that the asymptotic lit region remains fixed for a given surface when the receiving element sweeps forward and backward. However, in many cases, one can simply employ $\beta_s = \sqrt{(10/kL_s)}$ and select $L_s \geq (z_{\max} - z_{\min})/4$, as suggested in [3].

Using (12) to compute the weak interaction term and direct numerical summation to compute the strong interaction term, the total operational count involved in this procedure is $O(N)$ [3] in both vector-matrix multiplication and memory storage as the surface length increases for a fixed frequency.

III. Numerical Results and Discussion

In this section, F-BM/NSA algorithm is validated by considering the scattering from a 4096λ 1-D impedance rough surface illuminated by tapered plane wave with a near-grazing incident angle of 85°. The surface considered is one realization of a 3 m/s windspeed Pierson-Moskowitz ocean spectrum at 14 GHz (rms height at 2.27λ) with a dielectric constant $38 + i40$ and is sampled with eight unknowns per wavelength. The numerical result is compared with a reference solution obtained from BMIA/CAG. The accuracy of the F-BM is demonstrated in Fig. 3, where two curves are almost overlapping and cannot be distinguished from each other. It is noted that both F-BM/NSA and BMIA/CAG require only three iterations to obtain the converged result. It is also noted that in some applications, such as if one is interested in the emittivity from dielectric surfaces, eight unknowns per wavelength may not be sufficient, and more unknowns need to be employed in MoM modeling. However, this does not affect the validity of the current approach, because F-BM/NSA tends to accelerate the computation in solving the unknown current matrix in MoM instead of trying to modify MoM procedure itself. Furthermore, the number of total unknowns employed in F-BM/NSA is not restricted to be a power of 2, as usually employed in BMIA/CAG. In BMIA/CAG, the number of unknowns, including the pending zeros, usually increases in the power of 2 in order to efficiently use the fast fourier transform (FFT) for the acceleration of computation, while surface length increases. The CPU time is also compared and shown in Fig. 4, which shows $O(N)$ behavior for F-BM/NSA, while surface length increases for a fixed frequency. It is found that F-BM/NSA is far more efficient than BMIA/CAG in the order of magnitude.

References

[1] D. Holliday, L. L. DeRaad, Jr., and G. J. St-Cyr, "Forward-backward: A new method for computing low-grazing angle scattering," *IEEE Trans. Antennas Propagat.*, vol. 44, pp. 722–729, May 1996.

[2] D. A. Kapp and G. S. Brown, "A new numerical method for rough-surface scattering calculations," *IEEE Trans. Antennas Propagat.*, vol. 44, pp. 711–721, May 1996.

[3] H.-T. Chou and J. T. Johnson, "A novel acceleration algorithm for the computation of scattering from rough surfaces with the forward-backward method," *Radio Sci.*, vol. 33, pp. 1277–1287, Sept./Oct.

[4] L. Tsang, C. H. Chan, K. Pak, and H. Sangani, "Monte-Carlo simulation of large-scale problems of random rough surface scattering and applications to grazing incident with the BMIA/Canonical grid method," *IEEE Trans. Antennas Propagat.*, vol. 43, pp. 851–859, Aug. 1995.

[5] J. T. Johnson, "An extension of the canonical grid method for two dimensional scattering problems," *IEEE Trans. Antennas Propagat.*, vol. 46, Jan. 1998.

A Generalization of the Maximum Noise Fraction Transform

Christopher Gordon

***Abstract*—A generalization of the maximum noise fraction (MNF) transform is proposed. Powers of each band are included as new bands before the MNF transform is performed. The generalized MNF (GMNF) is shown to perform better than the MNF on a time dependent airborne electromagnetic (AEM) data filtering problem.**

***Index Terms*—Maximum noise fraction, noise filtering, time dependent airborne electromagnetic data.**

I. Introduction

The maximum noise fraction (MNF) transform was introduced by Green *et al.* [1]. It is similar to the principle component transform [2] in that it consists of a linear transform of the original data. However, the MNF transform orders the bands in terms of noise fraction.

One application of the MNF transform is noise filtering of multivariate data [1]. The data is MNF transformed, the high noise fraction bands are filtered and then the reverse transform is performed.

We show an example where the MNF noise removal adds artificial features due to the nonlinear relationship between the different variables of the data. A polynomial generalization of the MNF is introduced which removes this problem.

In Section II we summarize the MNF procedure. The problem data set is introduced in Section III and the MNF is applied to it. In Section IV, the generalized MNF transform is explained and applied. The conclusions are given in Section V.

II. The Maximum Noise Fraction (MNF) Transform

In this section, we define the MNF transform and list some of its properties. For further details the reader is referred to Green *et al.* [1] and Switzer and Green [3]. A good review is also given by Nielsen [4]. A reformulation of the MNF transform as the noise-adjusted principle component (NAPC) transform was given by Lee *et al.* [5]. An efficient method of computing the MNF transform is given by Roger [6].

Let

$$Z_i(x), \qquad i = 1, \cdots, p$$

be a multivariate data set with p bands and with x giving the position of the sample. The means of $Z_i(x)$ are assumed to be zero. The data can always be made to approximately satisfy this assumption by subtracting the sample means. An additive noise model is assumed

$$Z(x) = S(x) + N(x)$$

where $Z^T(x) = \{Z_1(x), \cdots, Z_p(x)\}$ is the corrupted signal and $S(x)$ and $N(x)$ are the uncorrelated signal and noise components of $Z(x)$. The covariance matrices are related by

$$\text{Cov}\{Z(x)\} = \Sigma = \Sigma_S + \Sigma_N$$

where Σ_N and Σ_S are the noise and signal covariance matrices.

Manuscript received March 30, 1998; revised February 28, 1999.

The author is with the School of Computer Science and Mathematics, University of Portsmouth, Portsmouth PO1 2EG, U.K. (e-mail: christopher.gordon@port.ac.uk).

Publisher Item Identifier S 0196-2892(00)00417-4.

The noise fraction of the ith band is defined as

$$\text{Var}\{N_i(x)\}/\text{Var}\{Z_i(x)\}.$$

The maximum noise fraction transform (MNF) results in a new p band uncorrelated data set which is a linear transform of the original data

$$Y(x) = A^T Z(x).$$

The linear transform coefficients, A, are found by solving the eigenvalue equation

$$A\Sigma_N\Sigma^{-1} = \Lambda A \tag{1}$$

where Λ is a diagonal matrix of the eigenvalues, λ_i. The noise fraction in $Y_i(x)$ is given by λ_i. By convention the λ_i are ordered so that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$. Thus the MNF transformed data will be arranged in bands of *decreasing* noise fraction. The proportion of the noise variance described by the first r MNF bands is given by

$$\frac{\sum_{i=1}^{r}\lambda_i}{\sum_{i=1}^{p}\lambda_i}.$$

The eigenvectors are normed so that $A^T\Sigma A$ is equal to an identity matrix.

The advantages of the MNF transform over the PC transform are that it is invariant to linear transforms on the data and the MNF transformed bands are ordered by noise fraction.

The high noise fraction bands can be filtered and then the transform reversed. This can lead to an improvement in the filtering results because the high noise fraction bands should contain less signal that might be distorted by the filtering. Examples of this approach have been given by Green *et al.* [1], Nielsen and Larsen [7], and Lee *et al.* [5].

An extreme version of MNF filtering is based on excluding the effects of the first r components. That is r is chosen so as to include only bands with high enough noise ratios. This can be achieved by

$$Z^*(x) = (A^{-1})^T R A^T Z(x) \tag{2}$$

where $Z^*(x)$ is the filtered data and R is an identity matrix with the first r diagonal elements set to zero. Thus eliminating the effect of one or more of the MNF bands produces a filtered data set which is a linear transform of the original data. This MNF-based filter uses interband correlation to remove noise.

In order to use (1) to compute A, Σ_N has to be known. Nielsen and Larsen [7] have given four different ways of estimating $N(x)$. They all rely on the data being spatially correlated. A simple method for computing $N(x)$ is by

$$N(x) = Z(x) - Z(x+\delta) \tag{3}$$

where δ is an appropriately determined step length. We are effectively assuming

$$S(x) = S(x+\delta).$$

To the extent that this is not true, the estimate of $N(x)$ is in error.

When this method of noise estimation is used, the MNF transform is equivalent to the min/max autocorrelation factor transform [3].

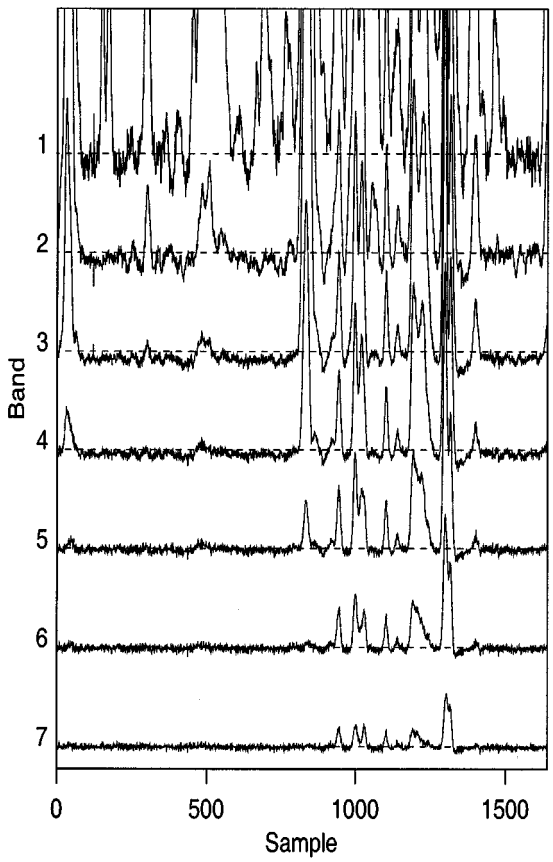

Fig. 1. Unfiltered AEM data. Bands 1 to 7 are shown. The band number of each spectrum is labeled to the left of the spectrum. The dotted line of each spectrum marks the zero amplitude for that spectrum.

III. Airborne Electromagnetic Data

We test the MNF filtering methodology on a flight line produced by SPECTREM's time dependent airborne electromagnetic (AEM) system. Background information on this AEM system has been explained by Leggatt [8]. A multiband image can be formed by consecutive flight lines but usually each flight line is examined separately.

Fig. 1 shows a flight line of data, consisting of the seven windowed AEM X band spectra. All seven bands are displayed stacked above each other. The amplitude of a band at a particular point is proportional to the vertical distance of the spectrum from its corresponding zero amplitude reference (dotted) line. Neighboring points along a line are responses from neighboring points on the ground. The higher band numbers are associated with greater underground depths.

Ore bodies are often associated with small features in the higher bands. Analysis can be made easier by filtering the spectra. Because this data set has substantial interband correlation, the MNF filtering methodology can be used.

Fig. 2(b) shows the MNF filtering of the spectra in Fig. 1. Only the last three bands (i.e., 5, 6, and 7) and a portion of the flight line are shown. The noise was estimated by taking the difference in neighboring pixels, as in (3). The data were filtered by excluding the first two MNF bands which accounted for approximately 86% of the noise fraction. Although the noise has been reduced, spurious features have been added, indicated by "S." Excluding only the last MNF component does not significantly reduce the magnitude of the spurious features and does almost no noise reduction.

As seen in (2), the MNF filtered data is composed from a linear function of the original data. Fig. 3 shows a plot of $Z_1(x)$ against $R_{Z_1}(x)$,

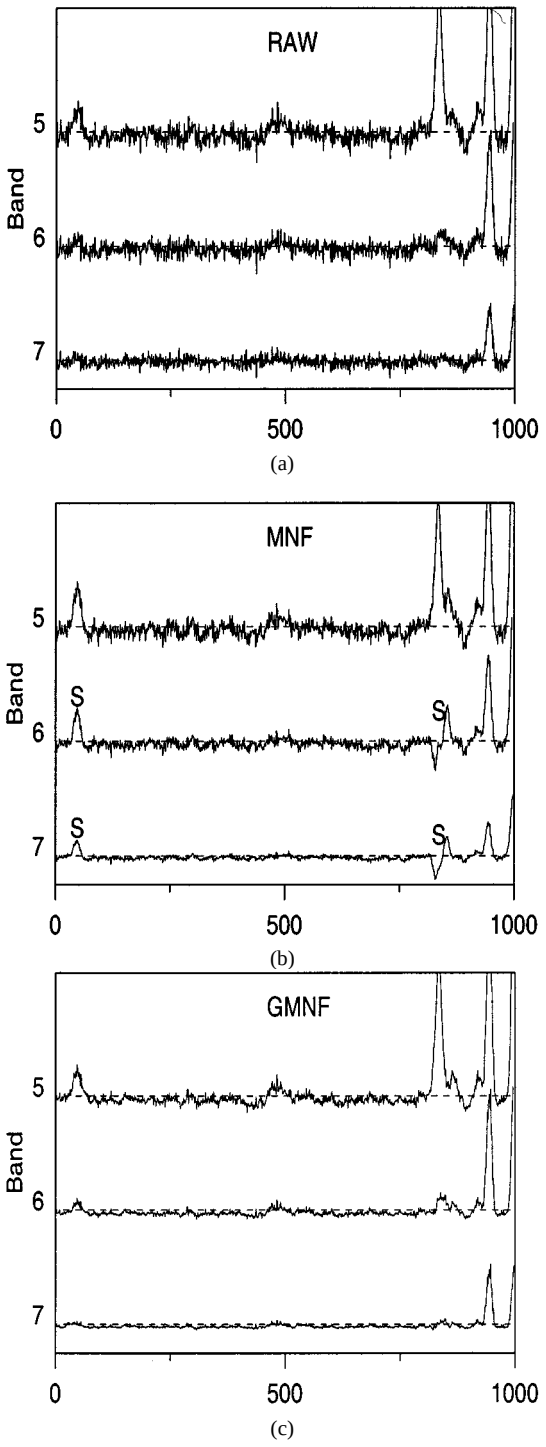

Fig. 2. A comparison of the MNF and GMNF filtering methods. Only a portion of the flight line for bands 5, 6, and 7 is shown for each figure. The sample number is displayed on the horizontal axis of each subplot: (a) unfiltered AEM data and (b) MNF filtered AEM data. The "S" symbols mark parts of the data where spurious features have been introduced by the MNF filtering, and (c) GMNF filtered AEM data.

where $R_{Z_1}(x)$ is the difference between $Z_1(x)$ and a least squares regression of $Z_1(x)$ based on all the other bands. The clear pattern of the residuals plotted in Fig. 3 is evidence that the relationship between $Z_1(x)$ and the other bands is not linear. Similar patterned residuals were found for residual plots based on the other bands. In the next section we show how the linear assumption can be relaxed.

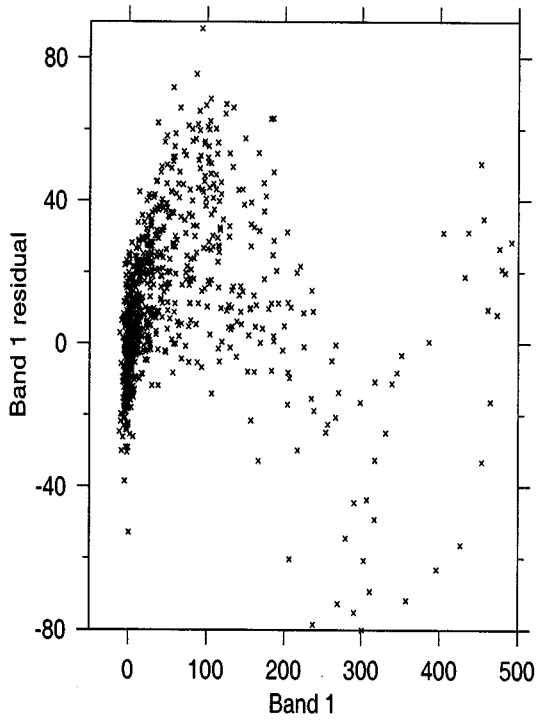

Fig. 3. Plot of the residual of a linear regression of band 1 based on bands 2 to 7, versus band 1 values.

IV. The Generalized Maximum Noise Fraction Transform

From the discussion in the previous section it appears that using a linear filter is too restrictive for this data set. Gnanadesikan and Wilk [9] proposed a generalization of the principle component transform. Powers of the original bands were appended to the data set as new bands. For example, p new bands can be created by appending the square of each band to the original data set. Thus each generalized principle component would be a polynomial, as opposed to linear, function of all the bands in the original data set.

The same procedure can be applied to generalize the MNF transform. More formally, a new data set, $Z'(x)$, can be created by appending up to q powers of the original data set

$$Z'(x) = \{Z_1(x),\ Z_2(x),\ \cdots,\ Z_p(x),\ Z_1^2(x),\ Z_2^2(x) \cdots,\ Z_p^2(x),\ \cdots,\ Z_1(x)^q,\ \cdots,\ Z_p^q(x)\}.$$

We are assuming that the $Z_i(x)$ have zero means. Cross terms, such as $Z_1(x)Z_2(x)$ can also be appended. The rest of the methodology remains unchanged.

From (2), each band of the generalized MNF (GMNF) filtered data can be seen to be

$$Z_i^*(x) = \sum_{j=1}^{p}\sum_{k=1}^{q} F_{i,\,j+(k-1)p} Z_j^k(x)$$

where $F_{i,\,j+(k-1)p}$ is the element in row i and column $j+(k-1)p$ of the filter matrix

$$F = (A^{-1})^T R A^T.$$

Thus, the GMNF transform leads to a polynomial filter.

To apply the GMNF filter to the data in Fig. 1, the GMNF transform was applied with powers of up to order 6 for each band appended to the original data. Cross terms were found to make little difference to the result and so were not included. The first 15 of the 42 GMNF components, contributing approximately 80% of the noise fraction, were eliminated.

Fig. 2(c) shows the GMNF filtered AEM data. A comparison with the MNF filtered data [Fig. 2(b)] shows that for GMNF filtered data, the noise reduction is greater and spurious features are much less evident.

V. Conclusion

We have proposed a generalized maximum noise fraction transform (GMNF) that is a polynomial as opposed to linear transform. The GMNF was applied to filtering a test AEM data set. It was found to remove more noise while adding less artificial features than the MNF-based filter.

Implementing the GMNF is a simple extension of the MNF implementation. Software written for the MNF transform can be be used for the GMNF transform without any modification.

References

[1] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Trans. Geosci. Remote Sensing*, vol. 26, pp. 65–74, Jan. 1988.

[2] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Reading, MA: Addison-Wesley, 1992.

[3] P. Switzer and A. Green, "Min/max autocorrelation factors for multivariate spatial imagery," Dept. Statistics, Stanford University, Stanford, CA, Tech. Rep. 6, 1984.

[4] A. A. Nielsen, "Analysis of regularly and irregularly sampled spatial, multivariate, and multi-temporal data," Ph.D. dissertation, Institute of Mathematical Modeling, Univ. Denmark, 1994.

[5] J. B. Lee, A. S. Woodyatt, and M. Berman, "Enhancement of high spectral resolution remote-sensing data by a noise-adjusted principal components transform," *IEEE Trans. Geosci. Remote Sensing*, vol. 28, pp. 295–304, Mar. 1990.

[6] R. E. Roger, "A faster way to compute the noise-adjusted principal components transform matrix," *IEEE Trans. Geosci. Remote Sensing*, vol. 32, June 1994.

[7] A. A. Nielsen and R. Larsen, "Restoration of GERIS data using the maximum noise fractions transform," in *Proc. 1st Int. Airborne Remote Sensing Conf. and Exhib.*, vol. III, Strasbourg, France, 1994, pp. 557–568.

[8] P. B. Leggatt, "Some algorithms and code for the computation of the step response secondary EMF signal for the SPECTREM AEM System," Ph.D. diss., Univ. Witwatersrand, Johannesburg, South Africa, 1996.

[9] R. Gnanadesikan and M. B. Wilk, "Data analytic methods in multivariate statistical analysis," in *Multivariate Analysis II*, P. R. Krishnaiah, Ed. New York: Academic, 1969, pp. 593–638.

Detection and Extraction of Buildings from Interferometric SAR Data

Paolo Gamba, Bijan Houshmand, and Matteo Saccani

***Abstract*—In this paper, we present a complete procedure for the extraction and characterization of building structures starting from the three-dimensional (3-D) terrain elevation data provided by interferometric SAR measurements. Each building is detected and isolated from the surroundings by means of a suitably modified machine vision approach, originally developed for range image segmentation. The procedure is based on a local approximation of the 3-D data by means of best-fitting planes. In this way, a building footprint, height and position, as well as its description with a simple 3-D model, are recovered by a self-consistent partitioning of the topographic surface reconstructed from interferometric radar data.**

***Index Terms*—SAR urban analysis, 3-D building extraction.**

I. Introduction

The urban environments, with their complex structure composed of buildings of different kinds and shapes, small and/or large green areas, infrastructures (roads, railroads, bridges, . . .) and continuously changing suburbs have constantly been a challenge for remote sensing analysts. Notwithstanding the large number of works on the interpretation of urban images acquired by different sensors, from the *classic photocameras* to synthetic aperture radars (SAR) [1], [2] from multispectral [3] to hyperspectral sensors (like AVIRIS [4]), a large amount of information is still hidden in the raw data.

On the other hand, with the largest part of the population in the world already settled in towns and cities, it is increasingly important to develop a set of flexible tools for the analysis, monitoring and planning of urban environments. Even the study of geological and hydrological risks in urban areas can give useful hints to prevent and alleviate hazards like earthquakes and floods, whose costs (in terms of lives more than dollars) have been steadily increasing in the past years [5].

To this aim, the continuous trend in research is to merge measurements and data from different sensors [6]–[8] to refine, by means of this interaction, the quality of the information extracted. Contemporarily, very interesting analyses in recent years have been dedicated to investigate how all-weather sensors, like the SAR, can be exploited to evaluate bio- and geophysical parameters in urban areas [9]. In particular, many papers have been presented aiming at determining which radar data (in terms of polarization [10], [11], wavelength [12], [13] or viewing angle [14], [15]) are more useful for urban image analysis.

However, very few papers are devoted to the use of interferometric SAR (IFSAR) measurements [16], [17] for urban image analysis: one of them is [4], where IFSAR and AVIRIS data are merged to better distinguish buildings from green areas. Indeed, the three-dimensional (3-D) measurements obtained by this system may be extremely useful for extracting the complete topography of a urban environment (for instance, for hydrological purposes) as well as for gathering more insight on particular structures or infrastructures (like the road network).

Analysis of the IFSAR terrain elevation data in urban areas are difficult due to the insufficient spatial resolution (with respect to urban features), multiple scattering due to the building geometries, and layover effects, in addition to the intrinsic IFSAR system level noise. Therefore, it is clear that there is still a strong need to evaluate which type of information is available from these data and to what extent it is possible to extract them.

The resolution problem is being increasingly resolved by the new generation of radar sensors that are currently operational or will be operational in the near future, like the NASA/JPL AIRSAR system [18]–[20] (currently a 40 MHz system, but to get upgraded to 80 MHz) and the DLR E-SAR system [21]. The goal of these systems is to provide a 1-m level spatial resolution, which therefore can resolve many of the objects present in an urban environment. As for the second problem, instead, we found very interesting to apply to the original remote sensing images some suitable machine vision approaches. Indeed, even if developed for very different situations, these procedures are of invaluable utility when used in this context.

In this work, we focus on the task to extract information on urban structures of interest from high-resolution IFSAR data. Specifically, we want to automate the detection (and subsequent analysis) of the height and shape of the buildings present in a given area. To this aim, we apply to the original data a segmentation algorithm able to exploit their resolution, while maintaining at the same time a high robustness to noise.

The paper is organized as follows: Section II presents the complete building detection strategy used in the 3-D image analysis procedure. Section III, instead, shows the results obtained on actual images and discusses their significance for urban area analysis as well as their limit for a more refined model-based extraction of the urban profile. Finally, in Section IV some conclusions and lines of thought for future improvements are introduced.

II. Building Detection Strategy

Our goal is to extract the significant buildings from interferometric SAR images, that is to locate some *special regions* inside them. Therefore, we must face a segmentation of the image, since segmenting an image means to divide it into meaningful objects according to a given criterion. The task is analyzed in [22], where some heuristic criteria for the correction of the shape of isolated buildings are applied to interferometric data, and in [23], where the author explored, in a more general context, how to find the parameters of a given building model that best fit the measured 3-D data. Both papers, however, do not address the problem of building extraction in a crowded, complex urban environment.

On the other hand, we may find useful to rely on consolidated approaches studied in machine vision. In this field, when considering 3-D (usually called *range*) images, generally the criteria applied to segment the data are geometric ones (see, for instance, [24] or [25]), often involving the principle of plane-fitting (i.e., to find the plane which better approximates a given surface). In our situation this approach can be useful when looking for the regions corresponding to the building roofs. The idea is therefore not only to partition the image (each pixel must belong to one region) but also to discard the data that do not carry useful information during image segmentation.

To this aim, the simplest possible algorithm could be an iterative region growing approach: we start from randomly chosen pixels (*seeds*) and examine all the adjacent ones. If one of them is *sufficiently* near to the seed in the 3-D space (where *sufficiently* must be defined by a suitable threshold), it is added. However, this is only a first step of the segmentation, since the data is now divided into regions whose geometric characteristics are still to be determined (for instance, are they

Manuscript received June 9, 1998; revised November 23, 1998. This work was supported by a contract with the National Aeronautics and Space Administration, Washington, DC.

P. Gamba and M. Saccani are with the Dipartimento di Elettronica, Università di Pavia, I-27100 Pavia, Italy.

B. Houshmand is with the Jet Propulsion Laboratory, Pasadena, CA 91109 USA.

Publisher Item Identifier S 0196-2892(00)00408-3.

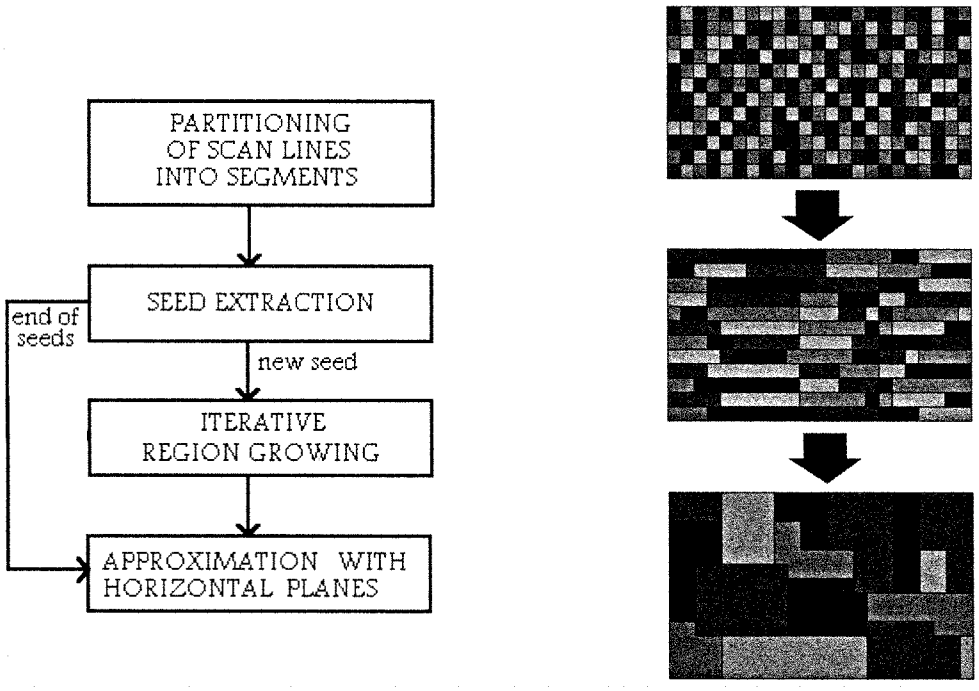

Fig. 1. Block diagram of the segmentation algorithm used to analyze the IFSAR images.

Fig. 2. On the right, an AIRSAR C-band image of Santa Monica, West Los Angeles (VV polarization). For a visual comparison, on the left an aerial photo of the same area is displayed. The blocks of black pixels correspond to large building shadows in the photograph, and to pixels labeled as "erroneous acquisition" in the SAR data.

planes or not?). However, since we expect that almost all of the structures in a urban environments can be roughly described by polyhedra with plain faces, we can try to approximate each of these regions by a plane.

This idea can be further improved by the algorithm outlined in [26]. In this approach the primitives of segmentation are not pixels, but scan lines (the lines of the image), in order to save cpu time. Grouping lines, it's faster to find consistent planes hidden in noisy data. We applied this procedure, suitably changed, following three processing steps (see also Fig. 1).

First Step—Scan Line Segmentation: The pixels belonging to the same scan line are grouped into segments according to a simple geometric criterion [27]: a curve is iteratively broken in two parts until no point of the original curve is far from the resulting segment chain more than a given threshold (θ_1). Since one scan line of the IFSAR image can be viewed as a curve in the third (*range*) dimension, this step represent an approximation by segments of the 3-D topographic data along each line of the image. Moreover, since this 3-D curve actually presents some discontinuities (the building edges, for instance), we follow [26] in using edge pixels (pixels with value very different from their left

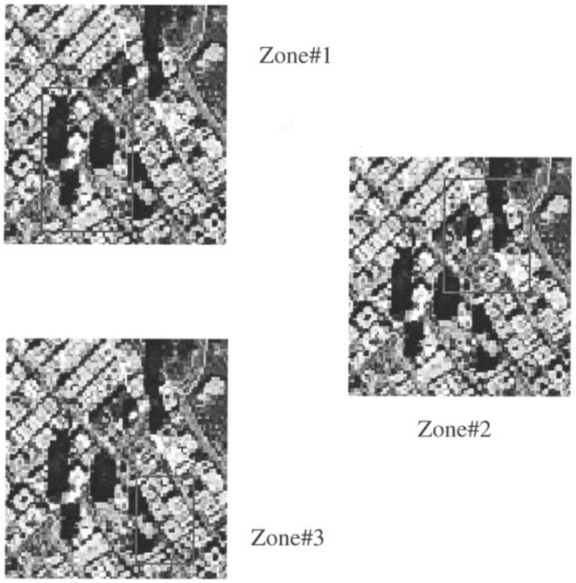

Fig. 3. Zoomed views of the aerial image in Fig. 2 showing the three zones around Wilshire Boulevard that were analyzed in this paper.

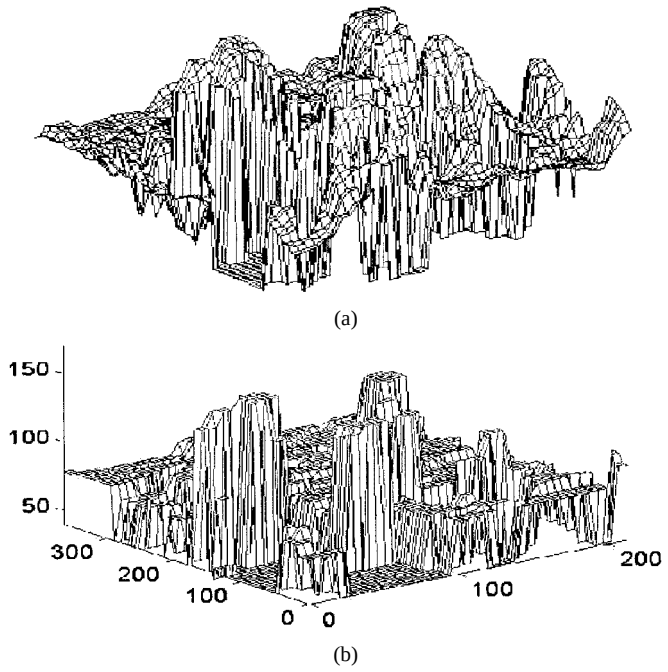

Fig. 4. Raw 3-D interferometric data provided by the TOPSAR system above the first zone of Fig. 3 and the reconstructed building profiles after the data analysis.

or right neighbor) as further breakpoints. Finally, each segment found is recorded in a list, with pointers to its neighbors (i.e., adjacent segments).

Second Step—Planar Region Aggregation: It consists of finding first the *seeds* for the aggregation and then to perform a region growing procedure to get the final, segmented image. Each seed is constituted by three adjacent segments (longer than a given threshold θ_2) belonging to different scan lines. The seeds are ordered and used for segment aggregation into planar surfaces starting from the one nearest to the ideal condition of three segments aligned on a plane.

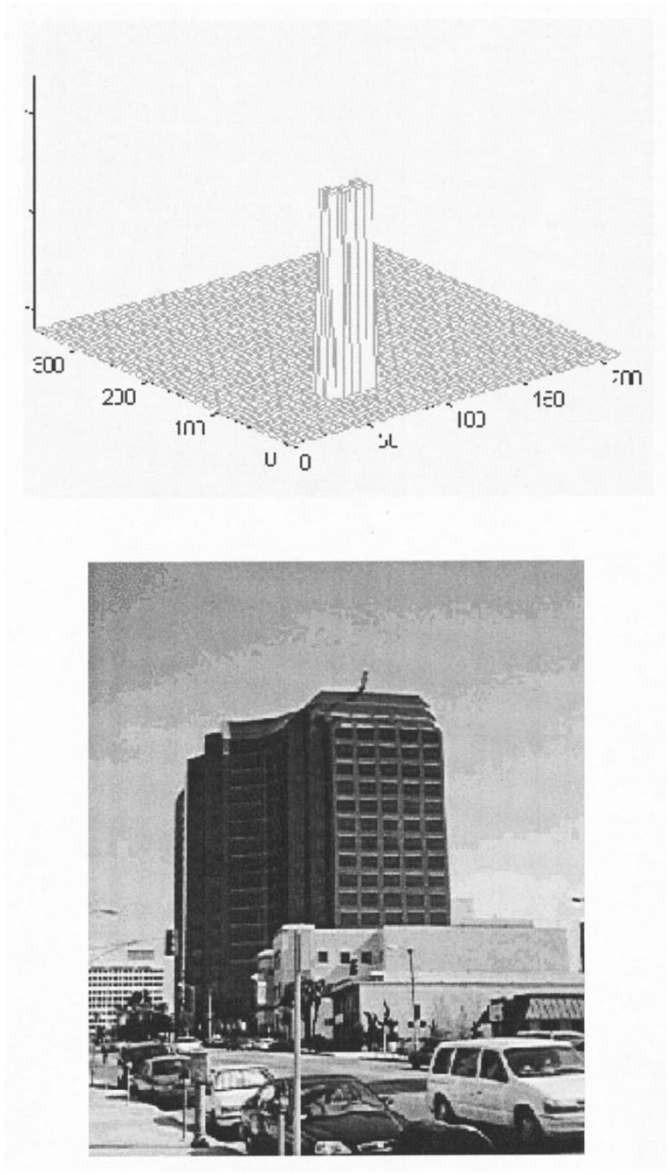

Fig. 5. The Coastal Federal Bank in Wilshire Boulevard: the building profile as reconstructed from IFSAR data and a photo from the ground.

This condition is measured by controlling that the directions and the intercepts of the seed segments coincide as much as possible (see also [26, Eq. (11)]). The index used is

$$i_p = 0.5 + \frac{1}{12}\left(\sum_{i \neq j} \frac{m_i \cdot m_j}{|m_i||m_j|} + \frac{n_i \cdot n_j}{|n_i||n_j|}\right) \tag{1}$$

where $m_i = (a_i, -1)$, $n_i = (b_i^*, -1)$, and $s_i = a_i x + b_i$, $i = 1, 2, 3$ are the algebraic expressions of the segments of a seed.

Next, the iterative region growing is performed. All the segments adjacent to the best seed are examined: if a segment is close enough (again, by a threshold θ_3) with respect to both its ends to the plane that approximates the seeds, it is added to the region. This process is iterated (considering the new region as an enlarged seed), until no more expansion is possible. Successively, less optimal seeds are used for the same process until the image is divided into planes and only segments that could not be aggregated are left.

Third Step—The Final Refinement: The previous segmentation may be improved by means of heuristic algorithms or more refined edge detection scheme to adjust the boundaries of the regions. In [26] the

simple method to reassign boundary pixels to the nearest plane is suggested.

As already stated above, to this procedure we must add a last step requiring that the best-fitting plane for each region is approximated with an horizontal one, for a first, imprecise simulation of building roofs.

Moreover, we must note that not all the pixels of the original image belong to a plane, when the algorithm stops: points affected by large noise, or regions where no actual planar surface is observable (for instance, trees in a park) are not aggregated. These points could be due to an error to be corrected or could carry important information not to be missed. At this point of the procedure it is difficult to say which is the case, until no further information is gained.

A. Critical Points of the Algorithm for Interferometric Data Segmentation

It is clear that the method described in the preceding paragraphs was developed originally as a machine vision approach to range image segmentation. Therefore, several problems arise when we try to obtain significant results from the application of this algorithm to the topographic data computed by SAR interferometry. We discuss here first the choice of the thresholds in the above outlined procedure (θ_1–θ_3), and then the point of prefiltering or not the original data. Finally, a few words will be also dedicated to the choice to approximate each plane with an horizontal one in the final resulting 3-D topography.

We found that, as far as the first step (scan line partitioning) is concerned, the breakpoints based on the original algorithm were not always the optimal ones. In low-resolution SAR images, it is necessary to overpartition the lines because of the small number of points defining each structure, while a different choice causes the merging of separate buildings into a unified object. It can be argued that this method introduces some kind of error in the segmentation procedure. However, the problem is later corrected by the segment grouping carried out in the growing process (while an underpartitioning would have been impossible to adjust). In a few words, this corresponds to a choice of θ_1 (see the previous paragraph) as low as $0.8\sigma_{img}$ where σ_{img} is the mean local image variance of the original data.

For the same reason also the definition of *edge pixels* may be changed according to the type of building that we want to extract. For instance, for high structures, edges correspond to large height steps, while when looking for residential objects, we have lower values. In this research we adopted a sort of conceptually pyramidal approach, and started by first extracting the large buildings. When working instead on residential structures we think that it would be better to lower the value of what is to be considered an edge accordingly.

The second threshold in the algorithm (θ_2) is set to evaluate only significant segments when looking for plane *seeds*. It is clear that, to discard possible error, it must be chosen as small as possible with respect to the physical characteristics of the searched objects and the resolution of the image. As for our IFSAR data are concerned (10 m resolution, 5 m posting), even segments of two of three pixels are meaningful since there exist buildings with these dimensions. This requires to consider a large number of seeds, leaving to a successive interpretation step the task to choose if all the planar regions are meaningful or not. Therefore a value smaller than the original value of 10 is used.

Finally, the last parameter to be tuned (θ_3) defines somehow which is the largest difference between a plane and a given segment to allow its aggregation. This parameter refers to the distance between the plane and the segment ends. Therefore, even if in [26] it is suggested a unique value for any situation, it is intuitive that longer segments require lower values (they can be aggregated only if they are significantly consistent with those already grouped), while smaller ones can be considered also in worse cases. We found, however, that the overpartitioning rule in the first step of the procedure usually provide only small segments, and the choice of a varying θ_3 does not change the quality of the results.

Another important point regards the prefiltering of the data. We decided not to make any prefiltering of the image (even if this could be extremely critical for the image analysis) because classical approaches to SAR filtering (see [28] for a recent review) usually provide also a smoothing of the image, that would have caused a loss of resolution, making the identification of the correct borders of each building difficult. Instead, the range segmentation offers a self-consistent smoothing of the 3-D data driven by the simple building models (essentially, parallelepipeds) that we use. In other words, the best-fitting plane procedure above delineated can be seen as the application of the *optimum filter* to the problem to recover a *signal* corrupted by noise knowing its shape (again, it's a plane). For this reason there is no need to perform any prefiltering. It is true, however, that more recent papers have developed pyramidal [29] or filter bank approaches [30] to avoid as much as possible the degradation of SAR image details after denoising. These methods allow a more precise edge location in radar measurements affected by noise, and we plan to add some sort of edge analysis as a further information source useful for our task.

Finally, we want to add a practical note: the approximation of the range regions with horizontal planes is made by using the mean value between the points belonging to each slanting plane. This introduces a further approximation which affects the values of the heights of the buildings. On the other hand it minimizes evaluation errors due to noise and/or false reflections.

III. Experimental Results and Discussion

The interferometric SAR range image used to show the results of this research covers a portion of Santa Monica, in the metropolitan area of Los Angeles (see Fig. 2). It is a range image, that is to say an array of numbers representing the surface elevation with respect to a reference plane. So, this image already gives us the 3-D profile of the urban surface.

The data were obtained with an interferometric SAR, the AIRSAR system, operated by NASA/JPL and mounted on a DC8 plane. The system is operated at C-band (5.6689 cm wavelength) with a 40 MHz pulse bandwidth, and has a nominal height accuracy in the order of ±2.5 m. The spatial resolution of the SAR system is therefore 3.75 m in range direction but, after the interferometric processing by phase unwrapping procedures [31], [32], this range is reduced to 7.5 m, since two pixels are averaged. The averaging in the azimuth direction is also performed to yield a square resolution cell. So, the ground range resolution for the mid-swath area (nominally 45 radar incident direction) is about 10 m, even if the images of the AIRSAR interferometric elevation data are provided after sampling at 5 m postings, geocoded, and rectified.

This resolution makes the building detection an extremely difficult task; nevertheless, these data were used in this research just to show how well the procedure behaves even when the spatial coarseness of the measurements is relatively low (10 m are comparable with most of the building footprints' dimensions).

The images we show here come from a larger data series recorded on August 5, 1994, from the height of 11 000 m. The flight path was from 33.97 N latitude, –118.47 longitude to 33.97 N latitude, –118.41 longitude. The radar look angle for the proposed area is nominally 45° and shadow/layover effects are observable as it can be seen by looking at the black pixels in Fig. 3 left, corresponding to incorrect measurements that were discarded by the TOPSAR data processor.

We stress that we present here only a very small part of the data recorded.

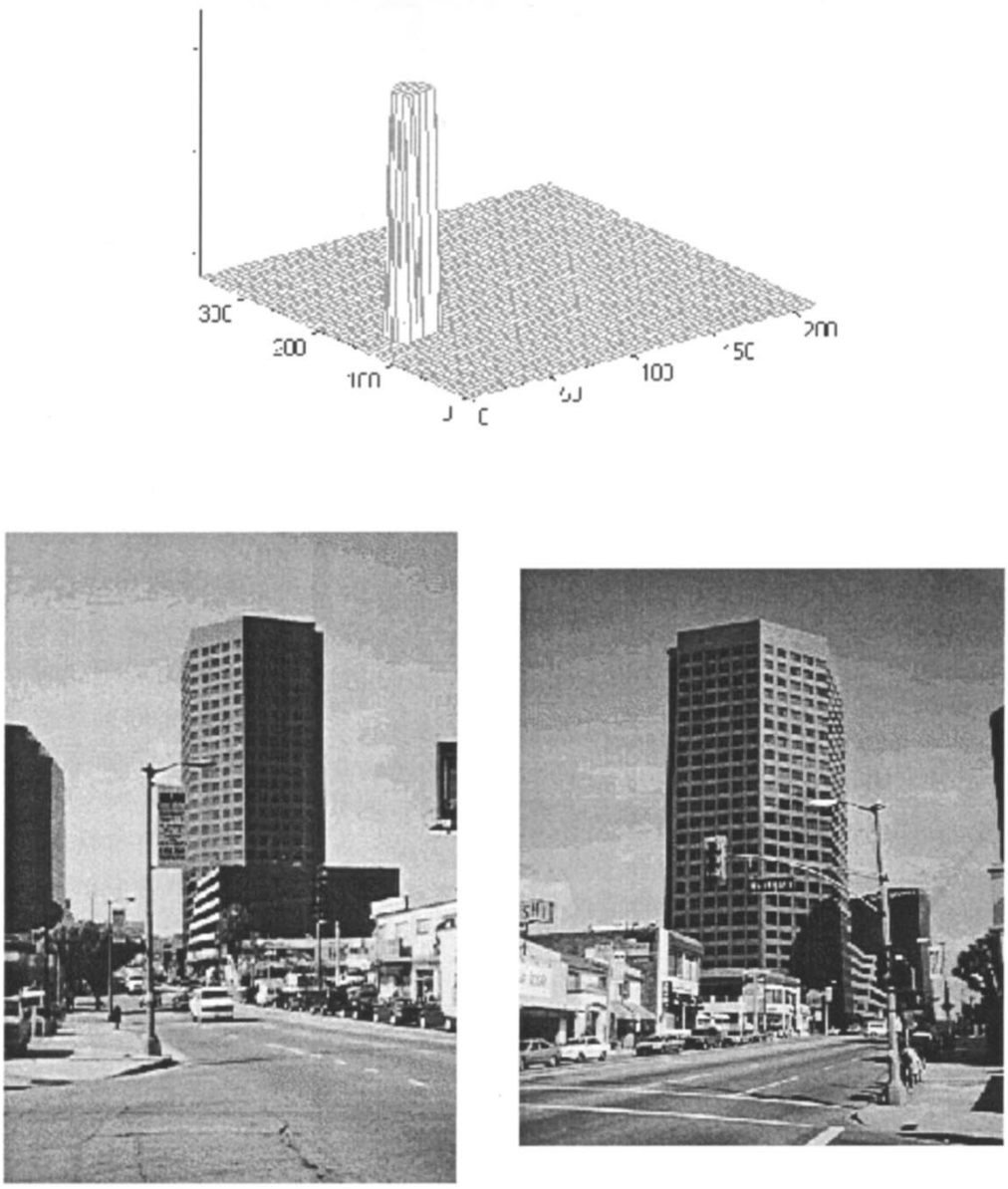

Fig. 6. Building profile of 11755 Wilshire Boulevard as reconstructed from IFSAR data and two photos from the ground.

A. Wilshire Boulevard

Within the study area, we applied the above outlined procedure to a subimage covering part of Wilshire Boulevard (East Santa Monica). The image was in turn divided into three parts (see Fig. 3) for a better analysis, and the above presented algorithm was applied separately to each of them. This was done in order to handle the data easily and to facilitate the identification of the buildings.

However, it should be mentioned that the original 3-D data lack the definition of a suitable ground level, because multiple reflections at the building edges produce responses that alter this value. To overcome this problem, we used the same procedure discussed in [4] and compute the overall height distribution of the data. Ground level is taken as the highest peak in the histogram, due to the presence in the area of large flat green areas, and points that have lower values are discarded. By means of this technique, only a very small number of pixels (in our area, less than 2%, and mainly around building edges) are not considered, but the successive detection procedure is considerably improved.

The results of the complete analysis for zone #1 of Fig. 3 are shown in Fig. 4: the upper picture represents the raw 3-D surface extracted from the interferometric measurements, while the other one shows the output of our program. It is immediately clear that the raw data are confusing, with disturbing noise and blurred building edges, while in the 3-D graph obtained from our algorithm the profiles of the most relevant buildings are now evident. Moreover, each of them is now a separate object. In other words, we operate simultaneously on Fig. 4(a) three operations:

1) a denoising procedure, as it is clear for instance looking at the roofs of the building in Fig. 4(b);
2) a structure recognition, because now we are able to distinguish the buildings from their (probably less interesting) surroundings;
3) an image segmentation.

As a final comment, we should note that there are also some artifacts still visible for the lower objects present in the scene, that we neglect for the moment.

More in detail, in Figs. 5 and 6 two of the buildings situated in Wilshire Boulevard are shown by simply retaining only those 3-D values that we found belonging to each of them. After that, their shape was compared either with maps or to the information that we

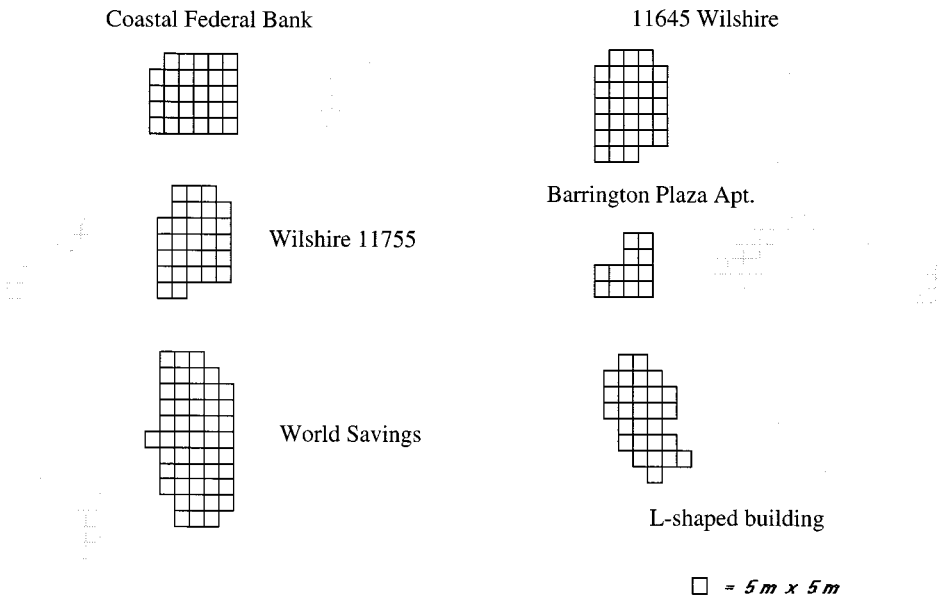

Fig. 7. Footprints of all the buildings extracted from the three subimages in Fig. 3 compared with their actual sections (shadowed): a square corresponds to 25 mq.

TABLE I
PERCENTAGE ERROR IN DETERMINING THE BUILDING FOOTPRINTS BEFORE AND AFTER THE REFINEMENT STEP IN SECTION VI-A

building	*before refinement*	*after refinement*
Coastal Federal Bank	-39%	-20%
World Savings	-48%	-25%
11755 Wilshire Boulevard	-67%	-37%
Barrington Plaza Apt.	-88%	-88%
11645 Wilshire Boulevard	-11%	-9%
L-shaped building	-50%	-1%

TABLE II
ACTUAL AND MEASURED HEIGHTS OF THE BUILDINGS EXTRACTED (MEAN ERROR = 2.2 m, σ = 4.9 m)

building	*actual height*	*measured height*	*absolute error*
Coastal Federal Bank	81	86	-5
World Savings	110	99	-11
11755 Wilshire Boulevard	98	99	+1
Barrington Plaza Apt.	74	71	-3
11645 Wilshire Boulevard	45	49	+4
L-shaped building	51	52	+1

extracted from a color image of the analyzed zone coregistered with the IFSAR original data. Furthermore, we had a number of color photographs taken on the ground depicting the main buildings of Wilshire Boulevard. The photographs allowed us to identify each building and to have a first idea of the validity of our results by means of a rough comparison of its shape and height. These figures show, respectively, the reconstructed shapes of the Coastal Federal Bank and 11755 Wilshire Boulevard, together with some photographs from the ground.

B. Some Comments on the Results

The first important thing to observe is that all the large buildings portrayed in the photographs had been extracted by our algorithm without any exception, and this is the minimum results we expected for. In addition, we observe that all the extracted structures correspond to commercial, financial and directional sites, that is buildings characterized by relevant heights and large dimensions. Smaller houses on the contrary are much more difficult to distinguish because of their small dimension. Moreover, we have not yet collected ground truth in a sufficient detail for that area to help in tuning the parameters of the extraction procedure. Finally, IFSAR data with higher spatial resolution than the one we used are needed to extract and recognize them; to this aim, the 2.5 m posting data, that the new TOPSAR system will be able to provide, may give sufficient information for a more detailed extraction of the less evident structures.

Each building carries at least three types of information, namely its position, shape and height. As far as the positions of the buildings are concerned, comparing our output graph with the color image, we have noticed that they are substantially correct (within the range of spatial precision of the data).

Instead, Fig. 7 shows the footprints of all the largest buildings in the three zones of Fig. 3, and a comparison with their actual shape is performed. It is clear that while the reconstructed shape of the objects is generally a sufficiently good representation of the real one, the area of the buildings is heavily underestimated. The main reason is the shadowing/overlay effect due to the relative position of the airplane carrying the radar and the illuminated large bodies on the ground. It affects the accuracy of the 3-D data related to the transition between a building and the road (or another building, the grass, etc.). As a result, it is hard

for the algorithm to find good borders, and this was the reason to obtain a building footprint mask using another sensor in [4]. Table I presents the errors computed as a percentage of the original area for the same examples of Fig. 7. Worse results are related to buildings "badly" oriented with reference to the flight direction (this is the case, for instance, of the Barrington Plaza Apt.)

To overcome the problem, we introduced a further refinement step to our procedure, devoted to two tasks: reduce the unclassified pixels, and merge horizontal planes that have very similar heights. The first task is due to the fact that in our final images there are many pixels not belonging to any plane; they are added to the nearest classified set if it is sufficiently near (less than 3 times the SAR precision, i.e., 7.5 m). Moreover, since the previous results show that probably each building roof is detected as a set of differently oriented planes (due to SAR processing errors, or to spurious reflections) we merged the final detected surfaces that are adjacent and very similar in height (the threshold is the same than above). Heights are changed taking into account the weight (in terms of pixels) of each merged set. The results obtained with the aid of this technique are presented in Table I, and show a significant improvement in the footprint estimate in almost all the cases.

The last comment regards the building heights, that happen to be necessarily approximated because we model each structure by using only planar surfaces; moreover, all the roofs are taken as flat ones. Nevertheless, the resulting values seem to be in very good accordance with those determined from field measurements. Table II shows the differences between the extracted heights and the actual values, together with the mean error; again, the minus sign represents an underestimate. This result (±4.9 m) must be compared with the mean error that we expect (according to [20]) from TOPSAR measurements, i.e., ±2.5 m. We may say that the loss in resolution from the original data to the classified one is limited, especially considering that this result is partially due to the error in the location of the ground level. Moreover, we think that it is better in an urban environment to have less precise information on each built structure as a single entity than to know the exact 3-D position of each measured point, without knowing to what it belongs. In other words, we believe that the limited loss in precision of our results with respect to the original data is more than compensated by the recognition of interesting urban structures.

IV. Conclusions

This work presents the application of a modified machine vision approach to 3-D data extracted from interferometric SAR measurements. The proposed approach has proved to be useful in reconstructing the 3-D structure of large commercial structures from a 10 m resolution data. Their shape is sufficiently well reconstructed and their height is found with an absolute mean precision around 2 m and standard deviation of ±4.9 m. However, building footprints are largely underestimated.

Therefore, the proposed approach exploits almost completely the vertical (±2.5 m) resolution of the original data and enables us to recognize and isolate those buildings that raise well over their surroundings, but lacks a suitable system to overcome layover/shadowing effects. A further refinement of this work is also needed for the recognition and analysis of smaller buildings such as residential houses by using both more detailed data and improved extraction algorithms. However, even these preliminary results show that there is a strong possibility to extract from IFASR data building models characterized with a precision only slightly worse than the original topographic data. Such analysis could be extremely useful for research as well as civic applications such as urban growth, and change monitoring.

References

[1] F. M. Henderson and Z. G. Xia, "SAR applications in human settlement detection, population estimation and urban land use pattern analysis: A status report," *IEEE Trans. Geosci. Remote Sensing*, vol. 35, pp. 79–85, Jan. 1997.

[2] Z. G. Xia, "Applications of multi-frequency, multi-polarization and multi-incident angle SAR systems in urban land use and land cover mapping," in *Proc. IGARSS'96*, vol. 4, Lincoln, NE, pp. 2310–2314.

[3] J. Hrikkonen, I. Kanellopoulos, A. Varfis, A. Steel, and K. Fullerton, "Urban land-use mapping with multi-spectral and SAR satellite data using neural networks," in *Proc. IGARSS'97*, vol. 4, Singapore, pp. 1660–1662.

[4] G. F. Hepner, B. Houshmand, I. Kulikov, and N. Bryant, "Investigation of the potential for the integration of AVIRIS and IFSAR for urban analysis," *Photogramm. Eng. Remote Sensing*, vol. 64, no. 8, pp. 512–520, 1998.

[5] G. Bertz and A. Smolka, "Urban earthquake loss potential: Economical insurance aspects," in *Proc. 10th Eur. Conf. Earthquake Engineering*.

[6] B. N. Haack, "Multisensor data analysis of urban environments," *Photogrammetric Eng. Remote Sensing*, vol. 50, no. 10, pp. 1471–1477, 1984.

[7] D. J. Weydahl, X. Becquey, and T. Tollefsen, "Combining ERS-1 SAR with optical satellite data over urban areas," in *Proc. IGARSS'95*, vol. 4, Florence, Italy, pp. 2161–2163.

[8] M. Caetano, J. Santos, and A. Navarro, "A multi-strategic approach for land use mappings of urban areas by integrating satellite and ancillary data," in *Proc. IGARSS'97*, vol. I, Singapore, pp. 240–242.

[9] Z. G. Xia and F. M. Henderson, "Understanding the relationships between radar response patterns and the bio- and geophysical parameters of urban areas," *IEEE Trans. Geosci. Remote Sensing*, vol. 35, pp. 93–101, Jan. 1997.

[10] M. L. Bryan, "Analysis of two Seasat synthetic aperture radar images of an urban scene," *Photogrammetric Eng. Remote Sensing*, vol. 48, no. 3, pp. 393–398, 1982.

[11] Y. Dong, B. C. Foster, and C. Ticehurst, "Decomposition of radar signatures from built and natural targets," *Int. Arch. Photo. Remote Sensing*, vol. XXXI, no. B7, pp. 196–203, 1996.

[12] F. M. Henderson, "An evaluation of Seasat SAR imagery for urban analysis," *Remote Sens. Environ.*, vol. 12, pp. 439–461, 1982.

[13] B. N. Haack, "L- and X-band like- and cross-polarized synthetic aperture radar for investigating urban environments," *Photogramm. Eng. Remote Sensing*, vol. 50, no. 3, pp. 331–340, 1984.

[14] B. Dousset, "Synthetic aperture radar imaging of urban surfaces: A case study," in *Proc. IGARSS'95*, vol. 4, Florence, Italy, pp. 2092–2096.

[15] D. J. Weydahl, "Identifying urban features using RADARSAT images taken at multiple incidence angles," in *Proc. IGARSS'97*, vol. 1, Singapore, pp. 287–289.

[16] E. Rodriguez and J. M. Martin, "Theory and design of interferometric synthetic aperture radars," *IEE Proc. Radar, Sonar, and Navigation*, vol. 139, no. 2, pp. 147–159, 1992.

[17] R. Gens and J. L. Van Genderen, "SAR interferometry: Issues, techniques, applications," *Int. J. Remote Sensing*, vol. 17, pp. 1803–1835, 1996.

[18] H. A. Zebker, S. N. Madsen, J. M. Martin, K. B. Wheeler, T. Miller, Y. Lou, G. Alberti, S. Vetrella, and A. Cucci, "The TOPSAR interferometric radar orographic mapping instrument," in *IEEE Trans. Geosci. Remote Sensing*, vol. 30, 1992, pp. 933–940.

[19] S. N. Madsen, H. A. Zebker, and J. M. Martin, "Topographic mapping using radar interferometry: Processing techniques," in *IEEE Trans. Geosci. Remote Sensing*, vol. 31, 1993, pp. 246–256.

[20] S. N. Madsen, J. M. Martin, and H. A. Zebker, "Analysis and evaluation of the NASA/JPL TOPSAR across-track interferometric SAR system," *IEEE Trans. Geosci. Remote Sensing*, vol. 33, pp. 383–391, 1995.

[21] M. Coltelli, G. Fornaro, G. Franceschetti, R. Lanari, A. Moreira, E. Sansoti, R. Scheiber, M. Tesauro, and T. I. Stein, "Results of the Mt. Etna interferometric E-SAR campaign," in *Proc. IGARSS'97*, vol. 4, Singapore, pp. 1554–1556.

[22] G. R. Burkhart, Z. Bergen, R. Carande, W. Hensley, D. Bickel, and J. R. Fellerhoff, "Elevation correction and building extraction from interferometric SAR imagery," in *Proc. IGARSS'96*, vol. 1, Lincoln, NE, pp. 659–661.

[23] U. Weidener, "Building extraction from digital elevation models," Institut fur Photogrammetrie, Bonn, Germany, 1995.
[24] R. W. Taylor, M. Savini, and A. P. Reeves, "Fast segmentation of range imagery into planar regions," *Comput. Vis. Graph. Image Processing*, vol. 45, pp. 42–60, 1989.
[25] J. Mukherjee, P. P. Das, and B. N. Chatterji, "Segmentation of range images," *Pattern Recognit.*, vol. 25, no. 10, pp. 1141–1156, 1992.
[26] X. Jiang and H. Bunke, "Fast segmentation of range images into planar regions by scan line grouping," *Mach. Vis. Applicat.*, no. 7, pp. 115–122, 1994.
[27] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1972.
[28] Y. Sheng and Z.-G. Xia, "A comprehensive evaluation of filters for radar speckle suppression," in *Proc. IGARSS'96*, vol. 2, Lincoln, NB, pp. 1559–1561.
[29] *Proc. IGARSS'96*, vol. 1, Lincoln, NE, pp. 411–413.
[30] C. H. Fosgate, H. Krim, W. W. Irving, W. C. Karl, and A. S. Willsky, "Multiscale segmentation and anomaly enhancement of SAR images," *IEEE Trans. Image Processing*, vol. 6, pp. 7–20, Jan. 1997.
[31] R. M. Goldstein, H. A. Zebker, and C. L. Werner, "Satellite radar interferometry: Two-dimensional phase unwrapping," *Radio Sci.*, vol. 23, no. 4, pp. 713–720, 1988.
[32] Q. Lin, J. F. Vesecky, and H. A. Zebker, "New approaches in interferometric SAR data processing," *IEEE Trans. Geosci. Remote Sensing*, vol. 30, Mar. 1992.

Effects of Atmospheric Boundary Layer Moisture on Friction Velocity with Implications for SAR Imagery

Steven M. Babin and Donald R. Thompson

***Abstract*—Using computer simulations, it is shown that failure to consider water vapor effects may lead to anomalous friction velocity changes with static stability. When interpreting synthetic aperture radar (SAR) imagery showing sea surface features induced by marine boundary layer spanning eddies, consideration should be given to the effects of water vapor.**

***Index Terms*—Atmospheric boundary layer, friction velocity, synthetic aperture radar (SAR), water vapor.**

I. Introduction

Synthetic aperture radar (SAR) images of the ocean surface have not only revealed features of the ocean but also of the atmosphere. Many previous studies have discussed atmospheric effects on the ocean surface and their corresponding signatures observed in SAR imagery (e.g., [1]–[4]). In particular, the presence of atmospheric boundary layer spanning eddies affects the roughness of the sea surface. Boundary layer spanning eddies enhance sea surface roughness beneath and down wind of convective downdrafts. Similarly, these eddies decrease sea surface roughness beneath and down wind of convective updrafts [3]. This roughness variation is often detectable with SAR. Under certain conditions, these atmospheric boundary layer spanning eddies form longitudinal rolls roughly parallel to the mean wind [5]–[7]. The role of atmospheric stability in the development of boundary layer rolls has been discussed by many authors [3], [8]–[10]. While the meteorological community is familiar with the effects of atmospheric moisture on stability, many in the engineering community may not be. The routine use of empirical relations in deriving atmospheric quantities from SAR imagery may lead one to forget the important impacts of water vapor physics on SAR observations. These impacts may lead to discrepancies in comparisons of SAR-derived quantities with buoy data if variations in humidity are ignored. Therefore, this paper is intended to explain and demonstrate the effects of humidity on atmospheric stability and its possible consequences for SAR imagery of the ocean.

To study the effects of boundary layer humidity on SAR imagery, we use an algorithm to calculate the friction velocity (u_*) based on bulk atmospheric measurements. Monin and Obukhov [11] defined the friction velocity as a characteristic velocity scale in the surface layer

$$u_* = \sqrt{\frac{\tau}{\rho}} \tag{1}$$

where τ is the surface stress and ρ is the air density. By nondimensionalizing the wind shear using this friction velocity, the following expression can be obtained:

$$\frac{d\overline{U}}{dz} = \frac{u_* \phi}{\kappa z} \tag{2}$$

where

U mean wind velocity in the surface layer;
z altitude;
κ von Karman's constant.

The von Karman's constant κ is nondimensional and has been determined from various experiments to be about 0.4 (e.g., [12]). The quantity ϕ is a universal similarity function that is often called a nondimensional stability function. This function is represented by different expressions for stable and unstable conditions but reduces to unity for neutral stability. When (2) is integrated over altitude using the boundary condition that the velocity at the surface is zero, we obtain

$$\overline{U} = \frac{u_*}{\kappa}(\ln(z/z_0) - \psi) \tag{3}$$

where z_0 is an integration constant called the surface roughness length and ψ is an integrated form of ϕ. The surface roughness z_0 is proportional to but smaller than the mean height of the rough elements of the surface. This surface roughness is related to the wind-induced surface stress. Therefore, the SAR backscatter cross section is related to the friction velocity.

We ran several computer simulations to calculate the friction velocity from a variety of input air–sea temperature differences, wind speeds, and humidities. Plots were then made comparing friction velocities calculated from various wind speeds and air–sea temperature differences. This paper will demonstrate that atmospheric boundary layer water vapor has a significant impact on the stability and therefore on ocean surface roughness and the resulting SAR imagery.

II. Algorithm

The algorithm for calculating the friction velocity (u_*) was adapted from Babin *et al.* [13] and Babin [14]. The foundation for this algorithm is the atmospheric surface layer theory of Liu *et al.* [15] and modifications to this theory made by Fairall *et al.* [16]. Fairall *et al.* [16] found that certain areas of the Liu *et al.* [15] theory could be improved. For example, because Monin–Obukhov similarity theory is not strictly applicable to low wind speeds, Godfrey and Beljaars [17] proposed changes that would extend the validity of this theory to the low wind speed regime. These modifications were employed by Fairall *et al.* [16]

Manuscript received January 5, 1999; revised April 23, 1999. This work was supported by the Office of Naval Research.

The authors are with Applied Physics Laboratory, Johns Hopkins University, Laurel, MD 20723-6099 USA.

Publisher Item Identifier S 0196-2892(00)00425-3.

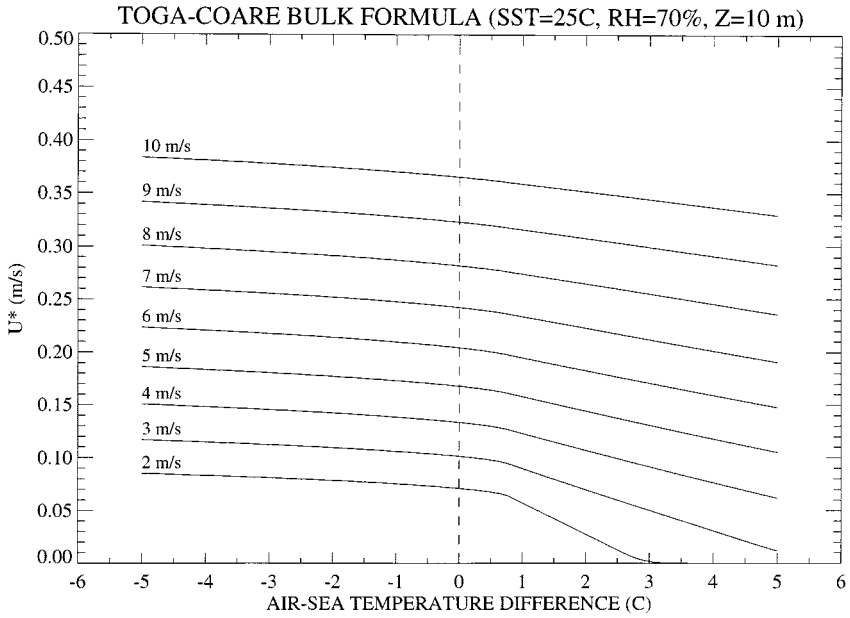

Fig. 1. Plots of friction velocity u_* versus air–sea temperature difference where this difference is determined using ordinary temperatures. Each line represents a different wind speed. This plot is determined from calculations assuming a 10 m measurement height, a sea surface temperature of 25 °C, and a relative humidity of 70%.

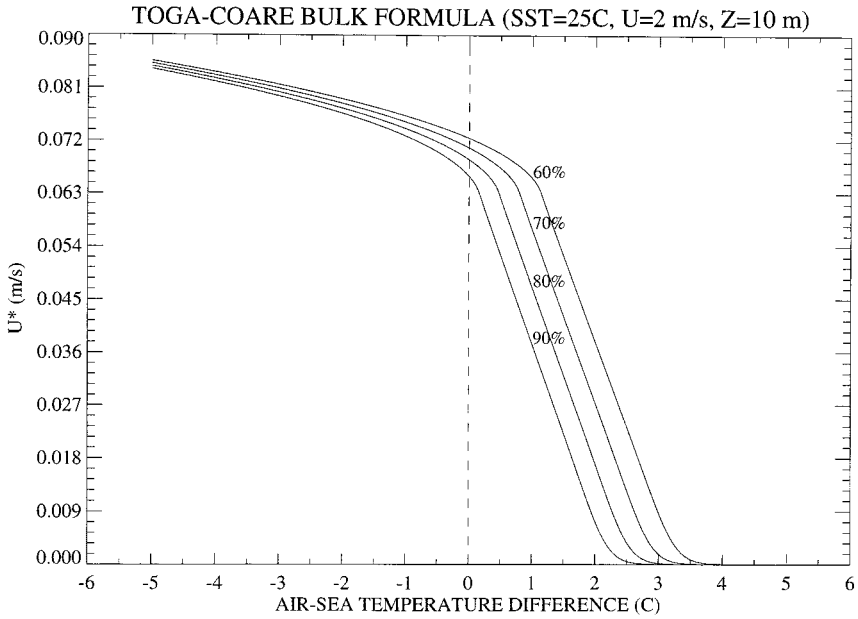

Fig. 2. Same as Fig. 1 except that the 10 m wind speed is fixed at 2 m/s and each line represents a different relative humidity (60, 70, 80, and 90%).

in the Tropical Ocean Global Atmosphere Coupled Ocean Atmosphere Response Experiment (TOGA-COARE) and were verified by simultaneous flux and bulk measurements made from numerous ships, buoys, and aircraft [18].

The equations of Buck [19] and a salinity correction [20] are used to calculate saturation water vapor pressures accurately. Using the Fairall *et al.* [16] technique, initial estimates of the Monin–Obukhov scaling parameters are made assuming neutral stability. Then, an iterative procedure is used to obtain their nonneutral values. Whereas Fairall *et al.* [16] fixed the number of these iterations, Babin *et al.* [13] continued the iterations until the new Monin–Obukhov parameters were within 0.001% of their previous values. This convergence typically takes less than 20 iterations.

For the calculations used in this paper, the measurement height was fixed at 10 m. The sea surface temperature was assumed to be fixed at 25 °C. The atmospheric pressure was fixed at 1015 hPa. Calculations were performed for wind speeds from 2 to 10 m s^{-}1. The relative humidity at the measurement height was an input variable. Two cases were examined. In the first case, the influence of variations in the 10 m relative humidity on plots of u_* versus air–sea temperature difference were examined. In the second case, the effects of this relative humidity on plots of u_* versus air–sea virtual potential temperature difference were examined.

Virtual potential temperature is a more accurate determinant of atmospheric stability over the ocean than ordinary temperature [21], [22]. Virtual temperature is defined as the temperature that dry air must have

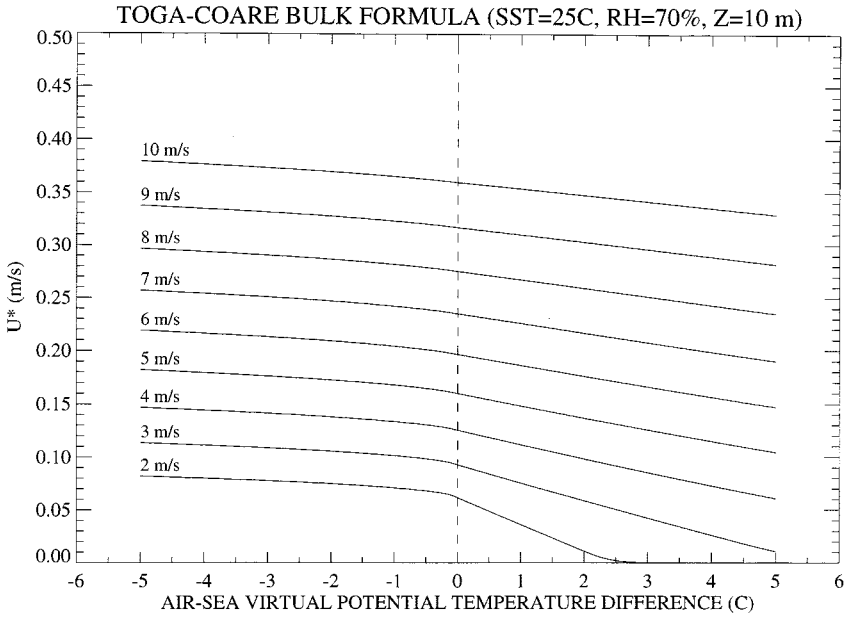

Fig. 3. The same as Fig. 1 except that the abscissa is virtual potential temperature difference.

to equal the density of moist air at the same pressure. Because moist air is less dense than dry air, it is more buoyant. Therefore, virtual temperature effectively accounts for the effects of moisture on stability. The virtual potential temperature (θ_v) is commonly used in the atmospheric boundary layer because it also removes the temperature variation caused by altitude changes (i.e., adiabatic lapse rate). Therefore, the buoyancy or stability of a moist air mass can be assessed by the vertical profile or gradient of virtual potential temperature. When θ_v is constant with altitude, the air is neutrally buoyant and therefore has neutral stability. When θ_v decreases with altitude, the air is positively buoyant and is considered stable. When θ_v increases with altitude, the air is negatively buoyant and is considered unstable.

III. Results

To illustrate the effects of atmospheric water vapor on the friction velocity, plots of friction velocity versus air–sea temperature difference (Fig. 1) were made for a 10 m relative humidity of 70% and wind speeds between 2 and 10 m/s. Fig. 2 shows a similar plot but for a fixed wind speed of 2 m/s and relative humidities between 60 and 90%. Note how the friction velocity curves change with relative humidity. The location of the maximum slope change in friction velocity is farther from neutral at the lower relative humidity.

Fig. 3 show the same plots as in Fig. 1 except that the abscissa is now virtual potential temperature. Because virtual potential temperature takes into account the buoyancy effects of water vapor, these curves will be identical for different relative humidities. If the criterion for stability were based instead on ordinary temperatures, then the effects of humidity on stability and hence on friction velocity will be overlooked. Therefore, it is important that virtual potential temperature be used as a stability criterion when examining friction velocities over the ocean. It is also interesting to note that, as would be expected, the change in u_* with stability is less at higher wind speeds.

IV. Summary

We have briefly reviewed the contribution of atmospheric water vapor to stability and demonstrated how this affects the friction velocity. When atmospheric water vapor is not considered as part of the stability criterion (Figs. 1 and 2), the u_* versus stability curves change with relative humidity. However, when atmospheric water vapor is considered, these curves do not change with relative humidity (Fig. 3). Because the friction velocity u_* is related to the surface roughness, these atmospheric effects may be observed in SAR imagery. Such effects are also more significant at low wind speeds. At low wind speeds, the effects of stability on convective eddies are more pronounced than at higher wind speeds where wind shear effects would dominate. Therefore, it is important to consider the contribution of atmospheric water vapor to static stability when interpreting SAR imagery over the ocean, particularly at low wind speeds.

References

[1] T. W. Gerling, "Structure of the surface wind field from SEASAT SAR," *J. Geophys. Res.*, vol. 91, pp. 2308–2320, 1986.

[2] W. Alpers and B. Brummer, "Atmospheric boundary layer rolls observed by the synthetic aperture radar aboard the ERS-1 satellite," *J. Geophys. Res.*, vol. 99, pp. 12 613–12 621, 1994.

[3] T. D. Sikora, G. S. Young, R. C. Beal, and J. B. Edson, "Use of spaceborne synthetic aperture radar imagery of the sea surface in detecting the presence and structure of the convective marine atmospheric boundary layer," *Month. Weath. Rev.*, vol. 123, pp. 3623–3632, 1995.

[4] C. C. Wackerman, C. L. Rufenach, R. A. Shuchman, J. A. Johannessen, and K. L. Davidson, "Wind vector retrieval using ERS-1 synthetic aperture radar imagery," *IEEE Trans. Geosci. Remote Sensing*, vol. 34, pp. 1343–1352, 1996.

[5] A. J. Faller, "The angle of windrows in the ocean," *Tellus B, Chem. Phys. Meteorol.*, vol. 16, pp. 363–370, 1964.

[6] ——, "Large eddies in the atmospheric boundary layer and their possible role in the formation of cloud rows," *J. Atmos. Sci.*, vol. 22, pp. 176–184, 1965.

[7] D. Etling and R. A. Brown, "Roll vortices in the planetary boundary layer: A review," *Bound. Layer Meteorol.*, vol. 65, pp. 215–248, 1993.

[8] M. A. LeMone, "The structure and dynamics of horizontal roll vortices in the planetary boundary layer," *J. Atmos. Sci.*, vol. 30, pp. 1077–1091, 1973.

[9] R. A. Brown, "Longitudinal instabilities and secondary flows in the planetary boundary layer: A review," *Rev. Geophys. Space Phys.*, vol. 18, pp. 683–697, 1980.

[10] P. J. Mason, "On the influence of variations in Monin–Obukhov length on horizontal roll vortices in an inversion-capped planetary boundary layer," *Bound. Layer Meteorol.*, vol. 27, pp. 43–68, 1983.

[11] A. S. Monin and A. M. Obukhov, "Basic laws of turbulent mixing in the atmosphere near the ground," *Tr. Akad. Nauk. SSSR Geofiz. Inst.*, vol. 24, pp. 163–187, 1954.

[12] U. Högström, "Non-dimensional wind and temperature profiles in the atmospheric surface layer: A re-evaluation," *Bound. Layer Meteorol.*, vol. 42, pp. 55–78, 1988.

[13] S. M. Babin, G. S. Young, and J. A. Carton, "A new model of the oceanic evaporation duct," *J. Appl. Meteor.*, vol. 36, pp. 160–201, 1997.

[14] S. M. Babin, "A new model of the oceanic evaporation duct and its comparison with current models," Ph.D. dissertation, Univ. Maryland, College Park, 1996.

[15] W. T. Liu, K. B. Katsaros, and J. A. Businger, "Bulk parameterization of air–sea exchanges of heat and water vapor including the molecular constraints at the interface," *J. Atmos. Sci.*, vol. 36, pp. 1722–1735, 1979.

[16] C. W. Fairall, E. F. Bradley, D. P. Rogers, J. B. Edson, and G. S. Young, "Bulk parameterization of air–sea fluxes for Tropical Ocean—Global atmosphere coupled—Ocean atmosphere response experiment," *J. Geophys. Res.*, vol. 101, pp. C2:3747–3764, 1996.

[17] J. S. Godfrey and A. C. M. Beljaars, "On the turbulent fluxes of buoyancy, heat and moisture at the air–sea interface at low wind speeds," *J. Geophys. Res.*, pp. C12:22 043–22 048, 1991.

[18] P. J. Webster and R. Lukas, "TOGA COARE: The coupled ocean atmosphere response experiment," *Bull. Amer. Meteorol. Soc.*, vol. 73, pp. 1377–1416, 1992.

[19] A. L. Buck, "New equations for computing water vapor pressure and enhancement factor," *J. Appl. Meteor.*, vol. 20, pp. 1527–1532, 1981.

[20] H. U. Sverdrup, M. W. Johnson, and R. H. Fleming, *The Oceans: Their Physics, Chemistry and General Biology*. Englewood Cliffs, NJ: Prentice-Hall, 1942, p. 1087.

[21] W. Brutsaert, *Evaporation into the Atmosphere: Theory, History, and Applications*. Norwood, MA: Kluwer, 1991, p. 299.

[22] R. B. Stull, *An Introduction to Boundary Layer Meteorology*. Norwood, MA: Kluwer, 1991, p. 666.