A Novel Phase Unwrapping Method Based on Network Programming

Mario Costantini

***Abstract*— Phase unwrapping is the reconstruction of a function on a grid given its values mod 2π. Phase unwrapping is a key problem in all quantitative applications of synthetic aperture radar (SAR) interferometry, but also in other fields. A new phase unwrapping method, which is a different approach from existing techniques, is described and tested. The method starts from the fact that the phase differences of neighboring pixels can be estimated with a potential error that is an integer multiple of 2π. This suggests the formulation of the phase unwrapping problem as a global minimization problem with integer variables. Recognizing the network structure underlying the problem makes for an efficient solution. In fact, it is possible to equate the phase unwrapping problem to the problem of finding the minimum cost flow on a network, for the solution of which there exist very efficient techniques. The tests performed on real and simulated interferometric SAR data confirm the validity of our approach.**

***Index Terms*—Minimum cost network flow, phase unwrapping, network programming, SAR interferometry.**

I. Introduction

PHASE unwrapping is the reconstruction of a function on a grid given the value mod 2π of the function on the grid. We will refer to these two functions as the unwrapped and wrapped phases, respectively, while we will often refer to grid points as pixels or simply points. In the last few years, an increasing interest has been devoted to phase unwrapping, mainly due to the development of synthetic aperture radar (SAR) interferometry [1], [2], although applications of phase unwrapping can be found in several other fields [3]. We will concentrate on two-dimensional (2-D) phase unwrapping, which is the most interesting for applications.

Basically, all existing phase unwrapping techniques start from the fact that it is possible to determine the discrete "derivatives" of the unwrapped phase, that is, the neighboring pixel differences, when these differences are less than π in absolute value. From the discrete derivatives, the unwrapped phase can be reconstructed up to an additive constant. The methods differ in the way they overcome the difficulty posed by the fact that the hypothesis above may be false at some points. This causes the estimated unwrapped phase discrete derivatives to be inconsistent, that is, they do not form an "irrotational" vector field.

Branch-cut methods [4] unwrap by "integrating" the estimated discrete partial derivatives of the unwrapped phase along paths where the integration give self-consistent results. The allowed integration paths are delimited by cuts, which cannot be crossed, made between the local inconsistencies of the estimated discrete partial derivatives. The problem of building these cuts is not solved by a global approach, which can prevent these methods from being robust.

In least-squares methods [5]–[7], unwrapping is achieved by minimizing the mean square deviation between the estimated and unknown discrete derivatives of the unwrapped phase. Least-squares methods are very efficient computationally when they make use of fast Fourier transform (FFT) techniques [8], [9]. The resulting unwrapping is not very accurate, however, because least-squares procedures tend to spread the errors rather than containing them within a limited set of points. To overcome this problem, a weighting of the wrapped phase can be useful [9], [10]. However, the weighted least-squares algorithms proposed are iterative and not as efficient as the unweighted ones. Moreover, the accuracy of the results depends critically on the weighting mask used.

Recently it has been recognized that choosing the L_p norm with $p < 2$ instead of the mean square (which is the L_2 norm) as the error criterion can reduce the spread of errors [11]. However, the iterative procedure proposed is not computationally efficient and does not allow external weighting. Other investigations have pointed out how more sophisticated estimations of the discrete derivatives of the unwrapped phase can reduce unwrapping errors [12]. Finally, it is worth mentioning that a different unwrapping approach has been proposed for the case in which several related SAR interferometric data sets are available [13].

Here we propose a new method for phase unwrapping that we have recently developed [14], [15]. The method exploits the fact that the discrete derivatives of the unwrapped phase are estimated with possibly an error that is an integer multiple of 2π. This leads to formulating the phase unwrapping problem as a global minimization problem with integer variables: the weighted deviation between the estimated and the unknown discrete derivatives of the unwrapped phase is minimized, subject to the constraint that the two functions must differ by integer multiples of 2π. With this constraint, the spread of errors is prevented and the resultant unwrapped phase is identical to the original wrapped phase when rewrapped; in addition, the unwrapping results are less sensitive to slight changes of the weighting mask used.

Minimization problems with integer variables are usually very complex computationally. However, recognizing the network structure underlying our problem makes it possible to employ very efficient strategies for its solution. In fact, by

Manuscript received May 8, 1997; revised October 13, 1997.

The author is with ESRIN, European Space Agency, 00044 Frascati, Italy (e-mail: marioc@mail.esrin.esa.it).

Publisher Item Identifier S 0196-2892(98)02852-6.

choosing the weighted L_1 norm as the error criterion, the phase unwrapping problem can be equated to the problem of finding the minimum cost flow on a network, for the solution of which there exist very efficient algorithms [16].

Our phase unwrapping method has been tested on simulated and real SAR interferometric data. The results are very encouraging and demonstrate the robustness, accuracy, and efficiency of the method.

The phase unwrapping algorithm proposed is described in Section II. In Section III, we report the results of some validation tests performed. In Section IV, we briefly draw some conclusions.

II. Phase Unwrapping Method

Let us establish some notation and a preliminary mathematical framework. Let S be a rectangular grid of points (i,j) such that $i = 0,1,\dots N-1$ and $j = 0,1,\dots M-1$. Moreover, let S_0 be the subset of the $(i,j) \in S$ such that $i = 0,1,\dots N-2$ and $j = 0,1,\dots M-2$, whereas S_1 is the subset corresponding to $i = 0,1,\dots N-2$ and $j = 0,1,\dots M-1$, and S_2 is the subset corresponding to $i = 0,1,\dots N-1$ and $j = 0,1,\dots M-2$. We define for any real "function" $F(i,j)$, $(i,j) \in S$, its discrete partial "derivatives" to be

$$\Delta_1 F(i,j) = F(i+1,j) - F(i,j), \quad (i,j) \in S_1 \tag{1}$$
$$\Delta_2 F(i,j) = F(i,j+1) - F(i,j), \quad (i,j) \in S_2. \tag{2}$$

It is easy to see that given two real valued functions $F_1(i,j)$, $(i,j) \in S_1$, and $F_2(i,j)$, $(i,j) \in S_2$, there exists a real valued function $F(i,j)$, $(i,j) \in S$, such that

$$\Delta_1 F(i,j) = F_1(i,j), \quad (i,j) \in S_1 \tag{3}$$
$$\Delta_2 F(i,j) = F_2(i,j), \quad (i,j) \in S_2 \tag{4}$$

if and only if

$$F_1(i,j+1) - F_1(i,j) = F_2(i+1,j) - F_2(i,j), \quad (i,j) \in S_0. \tag{5}$$

The condition expressed in (5) is the equivalent on a discrete space of the property that the gradient of a function is an irrotational vector field. When (5) holds, the function $F(i,j)$, $(i,j) \in S$, is determined unambiguously up to an additive real constant A by "integrating" the functions $F_1(i,j)$, $(i,j) \in S_1$, and $F_2(i,j)$, $(i,j) \in S_2$, according to an integration formula such as the following (a different integration path would give the same result):

$$F(i,j) = A + \sum_{i'=0}^{i-1} F_1(i',0) + \sum_{j'=0}^{j-1} F_2(i,j'), \quad (i,j) \in S. \tag{6}$$

Consider now a real-valued function $\Phi(i,j)$, $(i,j) \in S$, and let

$$\Psi(i,j) = \Phi(i,j) + 2\pi n(i,j), \quad (i,j) \in S \tag{7}$$

where $n(i,j)$ are the integers such that $\Psi(i,j) \in [-\pi,\pi)$. We refer to $\Phi(i,j)$ and $\Psi(i,j)$, $(i,j) \in S$, as the unwrapped and the wrapped phase functions, respectively. The inversion of (7), that is, the reconstruction of $\Phi(i,j)$ from $\Psi(i,j)$, $(i,j) \in S$, is the 2-D phase unwrapping process.

We define preliminary estimates of the discrete partial derivatives of the unwrapped phase according to

$$\Psi_1(i,j) = \Delta_1 \Psi(i,j) + 2\pi n_1(i,j), \quad (i,j) \in S_1 \tag{8}$$
$$\Psi_2(i,j) = \Delta_2 \Psi(i,j) + 2\pi n_2(i,j), \quad (i,j) \in S_2 \tag{9}$$

where $n_1(i,j)$ and $n_2(i,j)$ are integers selected based on *a priori* knowledge so that $\Psi_1(i,j) = \Delta_1 \Phi(i,j)$ for most of $(i,j) \in S_1$ and $\Psi_2(i,j) = \Delta_2 \Phi(i,j)$ for most of $(i,j) \in S_2$. For example, when it is assumed that there are not many big value jumps between neighboring points of the unwrapped phase function $\Phi(i,j)$, $(i,j) \in S$, the simplest option is to choose $n_1(i,j)$, $(i,j) \in S_1$, and $n_2(i,j)$, $(i,j) \in S_2$, to have $\Psi_1(i,j) \in [-\pi,\pi)$ and $\Psi_2(i,j) \in [-\pi,\pi)$, respectively: in fact, in this case it follows from (7)–(9) that $\Psi_1(i,j) = \Delta_1 \Phi(i,j)$ when $\Delta_1 \Phi(i,j) \in [-\pi,\pi)$, $(i,j) \in S_1$, and $\Psi_2(i,j) = \Delta_2 \Phi(i,j)$ when $\Delta_2 \Phi(i,j) \in [-\pi,\pi)$, $(i,j) \in S_2$.

In general, the functions $\Psi_1(i,j)$, $(i,j) \in S_1$, and $\Psi_2(i,j)$, $(i,j) \in S_2$, cannot be consistently interpreted as discrete partial derivatives of the unwrapped phase because they do not fulfill the irrotational property (5). It is convenient to restate the phase unwrapping problem of inverting (7) as the problem of finding the following discrete derivative residuals:

$$K_1(i,j) = \frac{1}{2\pi}[\Delta_1 \Phi(i,j) - \Psi_1(i,j)], \quad (i,j) \in S_1 \tag{10}$$
$$K_2(i,j) = \frac{1}{2\pi}[\Delta_2 \Phi(i,j) - \Psi_2(i,j)], \quad (i,j) \in S_2. \tag{11}$$

From the residuals $K_1(i,j)$, $(i,j) \in S_1$, and $K_2(i,j)$, $(i,j) \in S_2$, the discrete partial derivatives $\Delta_1 \Phi(i,j)$ and $\Delta_2 \Phi(i,j)$ are determined; the phase $\Phi(i,j)$, $(i,j) \in S$, can then be reconstructed, for example, according to (6), up to an additive constant that is an integer multiple of 2π. The reader familiar with branch-cut unwrapping method terminology [4] should note that the points where the functions $\Psi_1(i,j)$, $(i,j) \in S_1$, and $\Psi_2(i,j)$, $(i,j) \in S_2$, do not fulfill the irrotational property (5) are usually called residues, while the nonzero discrete derivative residuals $K_1(i,j)$, $(i,j) \in S_1$, and $K_2(i,j)$, $(i,j) \in S_2$, identify the branch cuts.

Let $c_1(i,j)$, $(i,j) \in S_1$, and $c_2(i,j)$, $(i,j) \in S_2$, be nonnegative real numbers weighting the *a priori* confidence that the residuals $K_1(i,j)$ and $K_2(i,j)$, respectively, must be small; that is, weighting the reliability of $\Psi_1(i,j)$ and $\Psi_2(i,j)$ as estimates of $\Delta_1 \Phi(i,j)$ and $\Delta_2 \Phi(i,j)$, respectively. Even when no independent knowledge is available, it can be useful to define weights based on information extracted from the data to be unwrapped themselves. For example, $c_1(i,j)$, $(i,j) \in S_1$, and $c_2(i,j)$, $(i,j) \in S_2$, could reflect the consistency of $\Psi_1(i,j)$ and $\Psi_2(i,j)$, respectively, as discrete partial derivatives of a function, that is, if $\Psi_1(i,j)$ and $\Psi_2(i,j)$ satisfy the irrotational property (5).

We can estimate the residuals $K_1(i,j)$, $(i,j) \in S_1$, and $K_2(i,j)$, $(i,j) \in S_2$, as the solution of the following mini-

mization problem:

$$\min_{\{k_1,k_2\}} \left\{ \sum_{i=0}^{N-2} \sum_{j=0}^{M-1} c_1(i,j)|k_1(i,j)| + \sum_{i=0}^{N-1} \sum_{j=0}^{M-2} c_2(i,j)|k_2(i,j)| \right\} \tag{12}$$

subject to the constraints

$$k_1(i,j+1) - k_1(i,j) - k_2(i+1,j) + k_2(i,j) = \frac{-1}{2\pi}[\Psi_1(i,j+1) - \Psi_1(i,j)\Psi_2(i+1,j) + \Psi_2(i,j)], \quad (i,j) \in S_0 \tag{13}$$

$$k_1(i,j) \text{ integer}, \quad (i,j) \in S_1 \tag{14}$$

$$k_2(i,j) \text{ integer}, \quad (i,j) \in S_2. \tag{15}$$

The objective function to be minimized in (12) comes from the assumption made in (8) and (9), which implies that the residuals to be estimated $K_1(i,j)$, $(i,j) \in S_1$, and $K_2(i,j)$, $(i,j) \in S_2$, are almost always zero. The weighted sum of the absolute values of the variables, which is the weighted L_1 norm, is chosen as the error criterion because it allows an efficient solution of the minimization problem through the transformation shown below. The constraints in (13) express the property that $\Psi_1(i,j) + 2\pi K_1(i,j)$, $(i,j) \in S_1$, and $\Psi_2(i,j) + 2\pi K_2(i,j)$, $(i,j) \in S_2$, are discrete partial derivatives of a function (the unknown function $\Phi(i,j)$, $(i,j) \in S$), which results from (10) and (11). These constraints ensure that the discrete partial derivatives of the unwrapped phase estimated from the solution of the minimization problem satisfy the irrotational property (5). Note that the right-hand side of (13) is different from zero when (5) is not satisfied by the discrete partial derivatives preliminary estimates $\Psi_1(i,j)$, $(i,j) \in S_1$, and $\Psi_2(i,j)$, $(i,j) \in S_2$. The constraints in (14) and (15) originate from the fact that $K_1(i,j)$, $(i,j) \in S_1$, and $K_2(i,j)$, $(i,j) \in S_2$, take integer values, as can be seen from (7)–(11). As a consequence, the spread of errors is prevented and the unwrapped phase achievable from the solution of the problem above is identical to the original wrapped phase when rewrapped. Moreover, by allowing only integer values to the variables, the solution of the minimization problem is less sensitive to a small change of the weights $c_1(i,j)$, $(i,j) \in S_1$, and $c_2(i,j)$, $(i,j) \in S_2$, used in (12).

The problem defined in (12)–(15) is a nonlinear minimization problem with integer variables. We will see that its solution can be found efficiently by solving the following problem:

$$\min_{\{x_1^+,x_1^-,x_2^+,x_2^-\}} \left\{ \sum_{i=0}^{N-2} \sum_{j=0}^{M-1} c_1(i,j)\left[x_1^+(i,j) + x_1^-(i,j)\right] + \sum_{i=0}^{N-1} \sum_{j=0}^{M-2} c_2(i,j)\left[x_2^+(i,j) + x_2^-(i,j)\right] \right\} \tag{16}$$

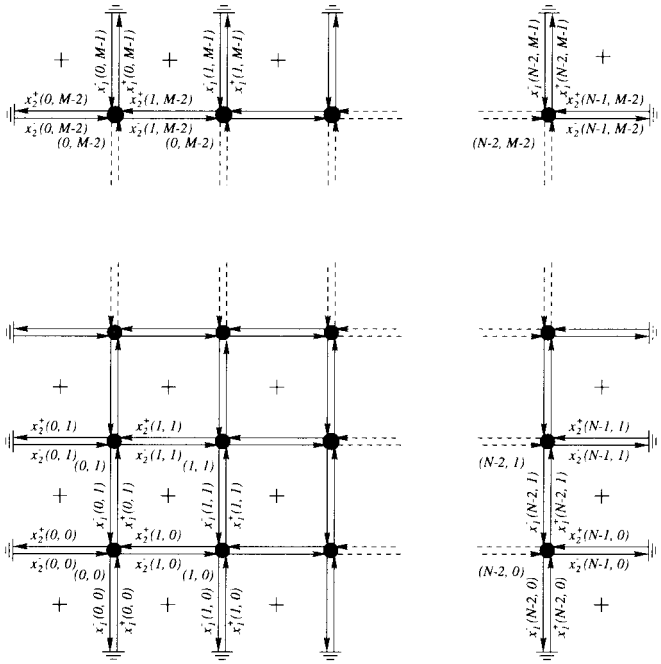


Fig. 1. Network associated with the phase unwrapping problem (the circles and the arrows represent the nodes and the arcs of the network, respectively; the boundary arcs are connected to the "earth"—by analogy with electrical networks—node; the crosses denote the phase grid points).

subject to the constraints

$$x_1^+(i,j+1) - x_1^-(i,j+1) - x_1^+(i,j) + x_1^-(i,j) - x_2^+(i+1,j) + x_2^-(i+1,j) + x_2^+(i,j) - x_2^-(i,j) = \frac{-1}{2\pi}[\Psi_1(i,j+1) - \Psi_1(i,j) - \Psi_2(i+1,j) + \Psi_2(i,j)], \quad (i,j) \in S_0, \tag{17}$$

$$x_1^+(i,j) \geq 0,\ x_1^-(i,j) \geq 0, \quad (i,j) \in S_1 \tag{18}$$

$$x_2^+(i,j) \geq 0,\ x_2^-(i,j) \geq 0, \quad (i,j) \in S_2. \tag{19}$$

The problem stated in (16)–(19) is a linear minimization problem with real variables. Moreover, it expresses a minimum cost flow problem on a network. In fact, consider the network consisting of an earth node plus the nodes labeled by (i,j), $(i,j) \in S_0$, with two arcs connecting each pair of nodes in the two directions (Fig. 1). Let the variables $x_1^+(i,j)$ and $x_1^-(i,j)$, $(i,j) \in S_1$, describe the flow along the arcs from node $(i,j-1)$ to node (i,j) and from node (i,j) to node $(i,j-1)$, respectively, while the variables $x_2^+(i,j)$ and $x_2^-(i,j)$, $(i,j) \in S_2$, describe the flow along the arcs from node (i,j) to node $(i-1,j)$ and from node $(i-1,j)$ to node (i,j), respectively, where the nodes with indexes $(i,j) \notin S_0$ correspond to the earth node. Moreover, let $c_1(i,j)$, $(i,j) \in S_1$, be the unit cost of the flows $x_1^+(i,j)$ and $x_1^-(i,j)$, whereas $c_2(i,j)$, $(i,j) \in S_2$, is the unit cost of the flows $x_2^+(i,j)$ and $x_2^-(i,j)$. Then, the objective function in (16) is the total cost of the flow. The constraints in (17) express the conservation of flow at nodes $(i,j) \in S_0$, with the right-hand side indicating (depending on whether it is positive or negative) the flow supply or demand at the nodes. Note that the elements of the right-hand side of (17) are integer, as can

be deduced from (8) and (9). Finally, the constraints in (18) and (19) define the capacities of the arcs.

The solution to the original minimization problem defined in (12)–(15) corresponds to a solution of the minimum cost network flow problem stated in (16)–(19) through the following change of variables:

$$x_1^+(i,j) = \max(0, k_1(i,j)),\ x_1^-(i,j) = \min(0, k_1(i,j)), \quad (i,j) \in S_1 \quad (20)$$

$$x_2^+(i,j) = \max(0, k_2(i,j)),\ x_2^-(i,j) = \min(0, k_2(i,j)), \quad (i,j) \in S_2. \quad (21)$$

In fact, it can be seen that the minimum cost network flow problem given in (16)–(19) has a solution at which at least one variable in each of the pairs $x_1^+(i,j), x_1^-(i,j), (i,j) \in S_1$, and $x_2^+(i,j), x_2^-(i,j), (i,j) \in S_2$, is zero, while all the variables $x_1^+(i,j), x_1^-(i,j), (i,j) \in S_1$, and $x_2^+(i,j), x_2^-(i,j), (i,j) \in S_2$, are integer. The first property can be easily demonstrated, and it becomes evident when you interpret $x_1^+(i,j)$ and $x_1^-(i,j), (i,j) \in S_1$, like $x_2^+(i,j)$ and $x_2^-(i,j), (i,j) \in S_2$, as opposite flows between two nodes. The second property can be demonstrated in the framework of the network programming theory and is related to the unimodularity of the matrix that can be associated to the constraints defined in (17) (see [16, Sections 9.6 and 11.12]). Note that the variables $x_1^+(i,j)$ and $x_1^-(i,j), (i,j) \in S_1$, as the variables $x_2^+(i,j)$ and $x_2^-(i,j), (i,j) \in S_2$, could be assigned different costs in (16). In view of (20) and (21), this would correspond to a different weighting of the possibilities that the residuals to be estimated $K_1(i,j), (i,j) \in S_1$, and $K_2(i,j), (i,j) \in S_2$, are positive or negative.

The minimum cost network flow problem stated in (16)–(19) can be solved very efficiently, both regarding the memory and the computation required. Many different efficient strategies can be used to solve minimum cost flow problems on a network, and their description would go beyond the scope of this paper (see [16, chs. 9–11]). Further investigations are needed to choose the most efficient minimum cost network flow algorithm for the phase unwrapping problem. In addition, the computational efficiency of our phase unwrapping method can be optimized by developing an algorithm that exploits the properties of the specific minimum cost network flow problem arising in phase unwrapping.

Let $b_1(i,j), (i,j) \in S_1$, and $b_2(i,j), (i,j) \in S_2$, be nonnegative integer numbers. It should be noted that, when useful, it is possible to add to the minimization problem defined in (12)–(15) the following constraints:

$$|k_1(i,j)| \le b_1(i,j), \quad (i,j) \in S_1 \quad (22)$$

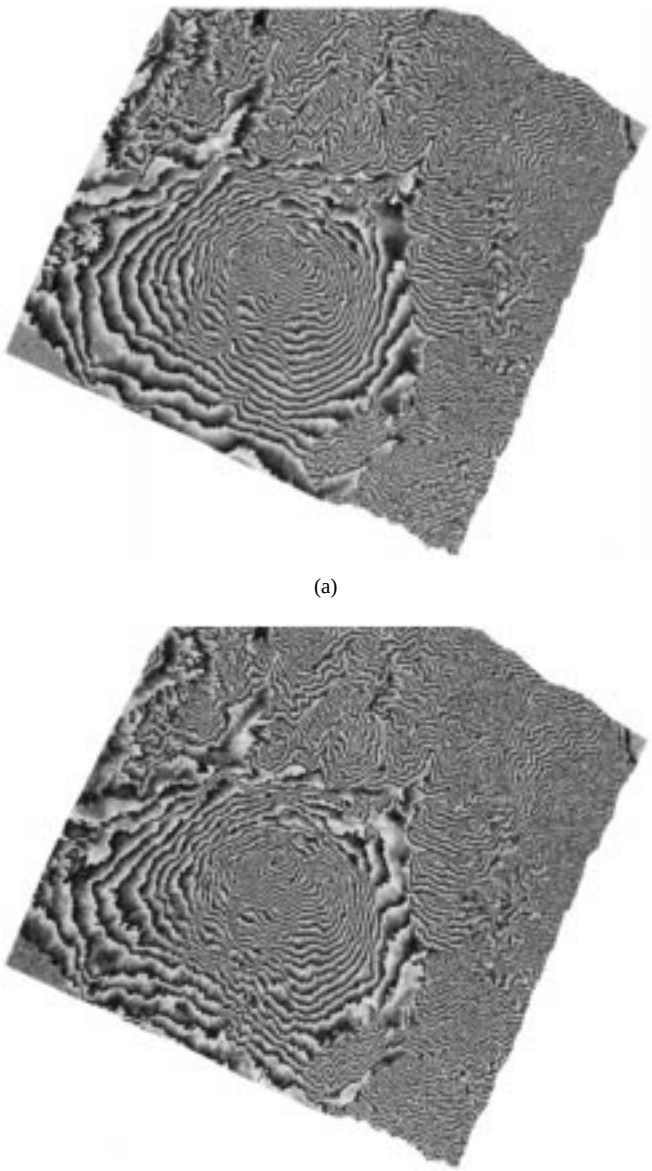
$$|k_2(i,j)| \le b_2(i,j), \quad (i,j) \in S_2 \quad (23)$$

and the corresponding arc capacity constraints

$$x_1^+(i,j) \le b_1(i,j),\ x_1^-(i,j) \le b_1(i,j), \quad (i,j) \in S_1 \quad (24)$$

$$x_2^+(i,j) \le b_2(i,j),\ x_2^-(i,j) \le b_2(i,j), \quad (i,j) \in S_2 \quad (25)$$

to the minimum cost network flow problem given in (16)–(19). These or other supplementary constraints could reflect *a priori* considerations, or could be useful in speeding the computation.



(a)

(b)

Fig. 2. (a) Wrapped phase simulating the interferometric phase obtained from the September 5 and 6, 1995, ERS-1 and ERS-2 SAR images of the Etna volcano, Sicily, Italy (the image size is 1644 × 1938 pixels; the grayscale represents the interval $[-\pi, \pi]$). (b) Real data corresponding to those simulated in (a).

Finally, it is important to note that, to process large data sets, it can be useful to subdivide the data into more manageable blocks that partially overlap. The minimum cost network flow problem can be solved sequentially in each of the blocks, subject to the boundary conditions imposed by the solution found in the blocks already processed to ensure that (17) holds globally. For example, suppose that the nth block, $n = 1, 2, \ldots N_b$, consists of the subset of the grid points $(i,j) \in S$ and such that $i = K_n, K_n + 1, \ldots K_n + N_n - 1$ and $j = L_n, L_n + 1, \ldots L_n + M_n - 1$. Suppose that the blocks are processed sequentially from the block with the smallest K_n and L_n, $n = 1, 2, \ldots N_b$, to that with the largest. Let A_n and B_n, $n = 1, 2, \ldots N_b$, be the subsets of the block grid points (i,j) such that $i = K_n, K_n + 1, \ldots K_n + N_n - 2$,

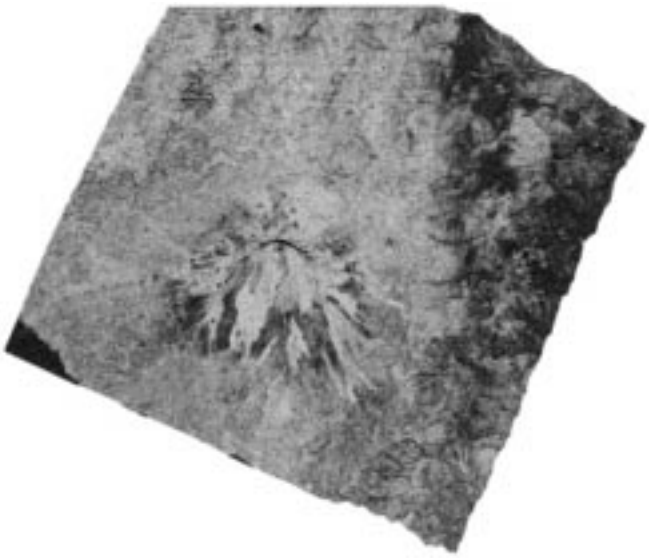


Fig. 3. Amplitude of the coherence of the interferometric phase shown in Fig. 2 (the grayscale represents the interval [0,1]).

$j = L_n, L_n + M_n - 1$ and $i = K_n, K_n + N_n - 1$, $j = L_n, L_n + 1, \ldots L_n + M_n - 2$, respectively, and such that they also belong to blocks already processed. Therefore, the solutions $X_1^+(i,j)$, $X_1^-(i,j)$, $(i,j) \in A_n$, and $X_2^+(i,j)$, $X_2^-(i,j)$, $(i,j) \in B_n$, have already been found for the variables $x_1^+(i,j)$, $x_1^-(i,j)$ and $x_2^+(i,j)$, $x_2^-(i,j)$, respectively. Then, for the nth block, a minimum cost network flow problem analogous to the one stated in (16)–(19) can be defined, but with the addition of constraints

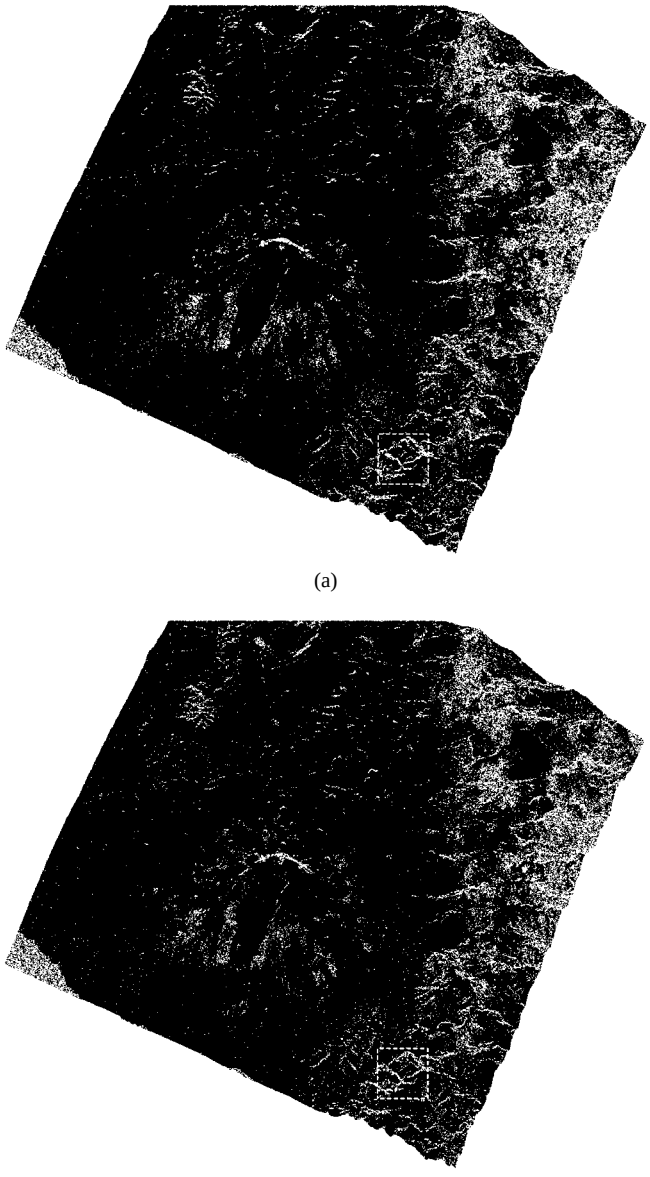
$$x_1^+(i,j) = X_1^+(i,j), \quad x_1^-(i,j) = X_1^-(i,j), \quad (i,j) \in A_n \tag{26}$$

$$x_2^+(i,j) = X_2^+(i,j), \quad x_2^-(i,j) = X_2^-(i,j), \quad (i,j) \in B_n. \tag{27}$$

The sequential solution of these minimum cost network flow problems for all the blocks n, $n = 1, 2, \ldots N_b$, can give a suboptimal solution of the global problem for virtually sized unlimited data with a fixed memory requirement, and with computational time linearly increasing with the size. In order for this decomposition to work, it is useful that the size of the blocks and their overlap are not too small. The block size should be sufficiently large to prevent incorrect preliminary estimates (8), (9) of the unwrapped phase discrete partial derivatives from cutting a block in disconnected regions. The overlap should allow the results to be discarded that are close to the boundaries of a block; these results are less reliable because less data are available to be unwrapped consistently.

III. Validation Tests

SAR interferometry provides one of the most difficult and interesting applications of phase unwrapping. We have tested our algorithm on simulated and real SAR interferometric phases. The simulated data allow a precise quantitative validation of the unwrapping method, while the real data are useful to verify the robustness of the algorithm in a less controlled

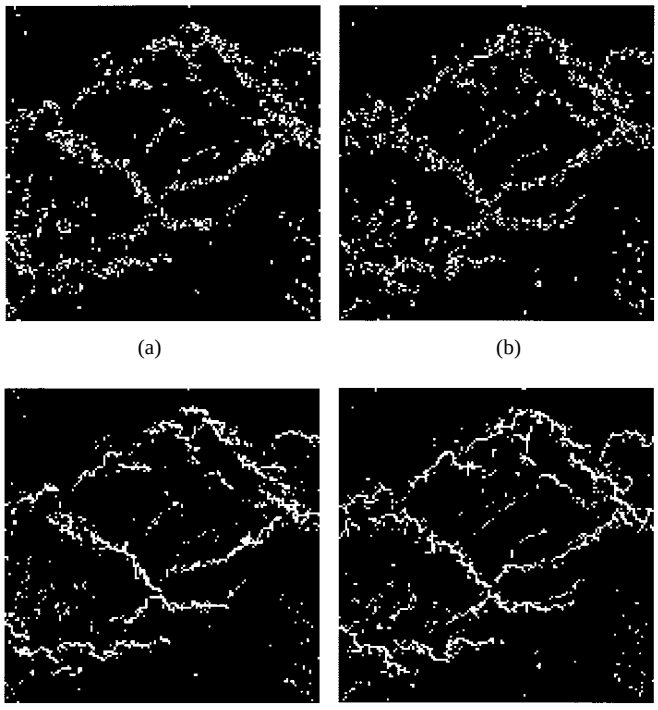


(a)

(b)

Fig. 4. (a) Inconsistencies ("residues") in the preliminary estimates (8), (9) of the discrete partial derivatives of the unwrapped phase for the data of Fig. 2(a) (in white are the points where the irrotational property (5) is not satisfied by these preliminary estimates; the dashed rectangle identifies the area enlarged in Fig. 5). (b) Same as (a), but referring to Fig. 2(b).

situation. Both the simulated and real interferometric phases to be unwrapped (Fig. 2) have been generated by means of the interferometric processor DIAPASON [17], starting from the SAR images taken on September 5 and 6, 1995, by the ERS-1 and ERS-2 satellites over the Etna volcano, Sicily, Italy, and using a digital elevation model (DEM) of the same region produced from SPOT satellite optical images. The interferometric phase data consist of 1644 $\times$ 1938 pixels of size 15.81 $\times$ 39.86 m^2 (2 $\times$ 10 looks) in the across and the along-track directions, respectively, which corresponds to an observed scene of about 66 $\times$ 77 km^2. The simulated phase has been made more realistic by introducing noise based on the amplitude of the coherence of the corresponding real interferometric phase (Fig. 3), that is, the amplitude of the



(a) (b)

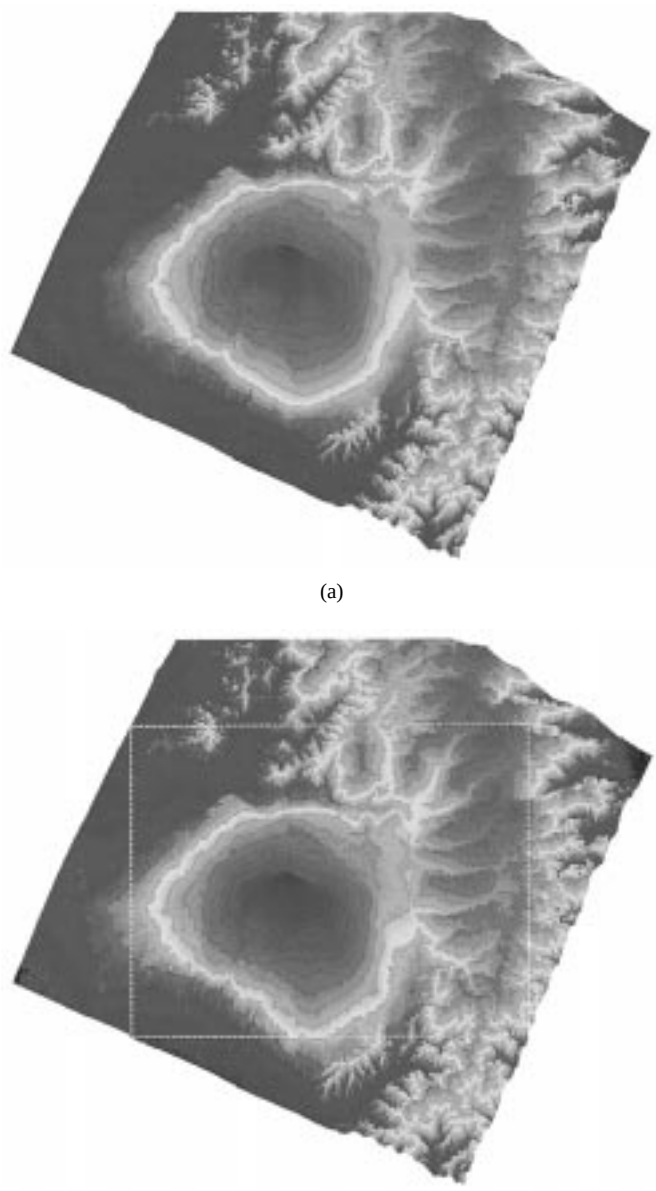
(c) (d)

Fig. 5. (a) Dashed area of Fig. 4(a) enlarged. (b) Dashed area of Fig. 4(b) enlarged. (c) Discrete derivative residuals (10), (11) ("cuts") for the data of Fig. 2(a) estimated by solving the minimization problem given in (12)–(15) (in white are the points corresponding to nonzero residuals). (d) Same as (c), but referring to Fig. 2(b).

correlation coefficient of the complex images from which the interferometric phase is derived. This amplitude is related to the error standard deviation of the phase [18].

For both the simulated and real data, preliminary estimates of the discrete partial derivatives of the unwrapped phase have been calculated according to (8) and (9), with $n_1(i,j)$, $(i,j) \in S_1$, and $n_2(i,j)$, $(i,j) \in S_2$, chosen to have $\Psi_1(i,j) \in [-\pi, \pi)$ and $\Psi_2(i,j) \in [-\pi, \pi)$, respectively. Many inconsistencies can be noted in these preliminary estimates of the unwrapped phase discrete partial derivatives (Fig. 4); that is, they do not satisfy the irrotational property (5). The errors in these preliminary estimates can be related to well-known phenomena in SAR interferometry. In particular, there is low coherence at the left corner of the image (corresponding to the sea) and at the right side (corresponding to mountains), while layover structures are recognizable at the right side and at the center, corresponding to peaks of mountains and the volcano.

The inconsistencies in the preliminary estimates (8), (9) of the unwrapped phase discrete partial derivatives can be overcome by solving the minimization problem defined in (12)–(15) to determine where these estimates must be wrong; that is, where the derivative residuals (10), (11) are found to be different from zero because the discrete partial derivatives of the unwrapped phase must be greater than π in absolute value (Fig. 5). We have chosen the weights $c_1(i,j)$, $(i,j) \in S_1$, and $c_2(i,j)$, $(i,j) \in S_2$, in (12) to reflect the amplitude of the coherence of the interferometric phase (Fig. 3); moreover, the weights have been set to zero when the corresponding preliminary estimates (8), (9) of the unwrapped phase discrete



(a)

(b)

Fig. 6. (a) Phase of Fig. 2(a) unwrapped by integrating, according to (6), its discrete partial derivatives determined by substituting in (10) and (11) the residuals that solve the minimization problem given in (12)–(15) (the color scale represents the interval [0, 234.55]). (b) Same as (a), but referring to Fig. 2(b) (the dashed rectangle identifies the area shown in Fig. 11).

partial derivatives do not satisfy the irrotational property (5) and then gradually increased in the neighborhood of these points. Note that the latter weighting is defined based on the phase data themselves; it has been found that this weighting is able to guarantee good unwrapping results on its own when no external information like phase coherence is available for weighting. However, further tests and investigations can be useful to find the best weighting. A discussion on the problem of selecting the weights for the analogous least-squares minimization problem can be found in [10].

The solution to the problem given in (12)–(15) is found, through (20) and (21), by solving the minimum cost network flow problem stated in (16)–(19). We have not enforced the constraints stated in (22) and (23) or (24) and (25): further tests

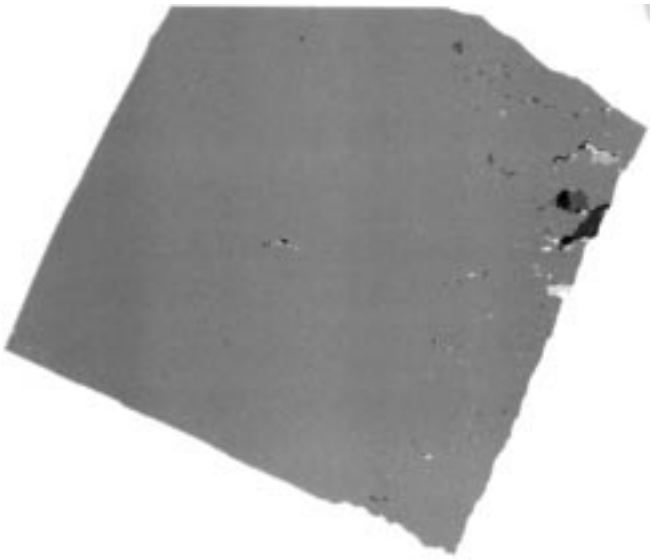


Fig. 7. Difference between the reconstructed unwrapped phase of Fig. 6(a) and the corresponding simulated unwrapped phase (the grayscale represents the interval $[-6\pi, 6\pi]$).

TABLE I
STATISTICS OF THE PHASE UNWRAPPING ERROR SHOWN IN FIG. 7 (PIXEL PERCENTAGE CORRESPONDING TO DIFFERENT ABSOLUTE VALUES OF THE DATA)

Absolute Error	0	2π	4π	$\geq 6\pi$
Pixel Percentage	98.46	1.04	0.30	0.20

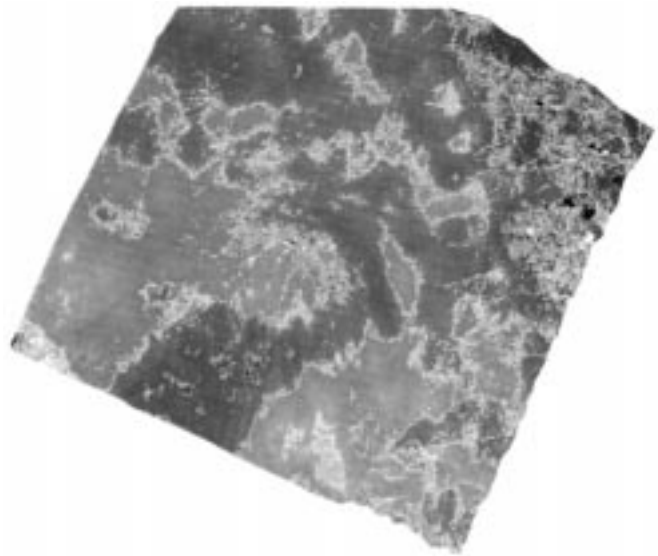


Fig. 8. Difference between the reconstructed unwrapped phase of Fig. 6(b) and the noise-free simulated unwrapped phase (the grayscale represents the interval $[-6\pi, 6\pi]$; the contour lines are spaced at 2π intervals).

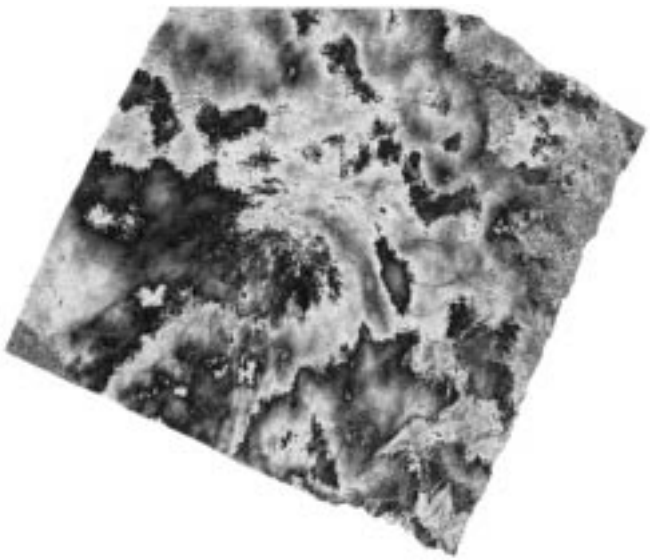


Fig. 9. Wrapped difference between the real wrapped phase of Fig. 2(b) and the noise-free version of the simulated wrapped phase shown in Fig. 2(a) (the grayscale represents the interval $[-\pi, \pi]$).

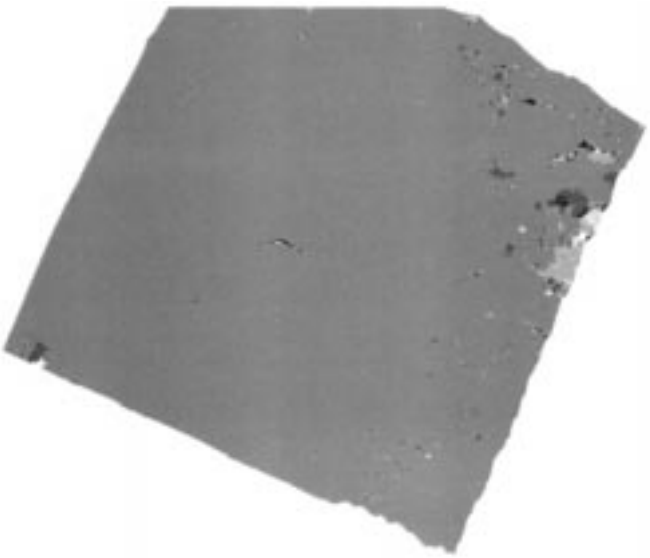


Fig. 10. Difference between the phase of Fig. 9 unwrapped with our algorithm and the data of Fig. 8 (the grayscale represents the internal $[-6\pi, 6\pi]$).

are needed to establish when they can be useful. To check the reliability of the block decomposition described in (26) and (27), we have split the data in four equal blocks, with the blocks overlapping by 25% of their size.

Then (Fig. 6), the unwrapped phase is reconstructed up to an additive constant integer multiple of 2π by integrating, according to (6), the discrete partial derivatives calculated by substituting in (10) and (11) the derivative residuals that solve the minimization problem stated in (12)–(15). We remember that the reconstructed unwrapped phase is identical to the original wrapped phase when rewrapped.

We have implemented two different minimum cost network flow algorithms in our phase unwrapping software: the CPLEX [19] network simplex algorithm (see [16, ch. 11] for a description of the network simplex method) and the RELAX [20] relaxation algorithm (see also [16, Section 9.10]). With the experimental setup chosen, unwrapping the 1644×1938 pixel data on a Silicon Graphics Power Onyx RE2 (using one R8000 CPU rated at about 300 MFlops, 75 MHz) took approximately 2 h when using the CPLEX software and 2 min with the RELAX algorithm. Note, however, that the time needed with

Fig. 11. Dashed area of Fig. 6(b) (perspective view; the grayscale represents the amplitude of the corresponding complex interferogram).

TABLE II
STATISTICS OF THE PHASE UNWRAPPING ERROR SHOWN IN FIG. 10 (PIXEL PERCENTAGE CORRESPONDING TO DIFFERENT ABSOLUTE VALUES OF THE DATA)

Absolute Error	0	2π	4π	$\geq 6\pi$
Pixel Percentage	97.91	1.67	0.26	0.16

the relaxation algorithm can vary significantly depending on the data. Of course, the computational time for unwrapping increases linearly with the data size when subdividing the data in blocks of fixed size; tests performed without using a block subdivision have indicated that the time is approximately $O((NM)^{1.6})$ and $O(NM)$, where NM is the number of pixels, with the CPLEX and the RELAX software, respectively. It is possible that even better performance can be obtained with different minimum cost network flow algorithms, in particular, if specifically optimized for the phase unwrapping minimum cost network flow problem.

To quantify the accuracy of the reconstructed unwrapped phase, we have calculated the difference between this and the known simulated one. Of course, the noisy version of the simulated phase has been used for comparison with the unwrapped phase reconstructed from the simulated data, while the noise-free version has been compared with the unwrapped phase reconstructed from the real data. When unwrapping the simulated phase, the error committed (Fig. 7 and Table I) is concentrated in layover and very low coherence areas, where interferometry does not furnish accurate phase information and the density of inconsistencies in the preliminary estimates (8), (9) of the unwrapped phase discrete partial derivatives is very high; in addition, these areas are located at the boundaries of the image, where less data have to be unwrapped consistently and the results are therefore less reliable. For the large majority of pixels, the error is zero.

In the case of the real data, the difference between the unwrapped phase and the noise-free version of the simulated unwrapped phase is greater than half a cycle for a large portion of the data (Fig. 8). However, this difference is not due to phase unwrapping errors, but it already exists between the real wrapped phase and the noise-free version of the simulated wrapped phase (Fig. 9). The discrepancy arises from a combination of the inaccuracy of the DEM and satellite orbits used for the simulation as well as atmospheric artifacts affecting the real interferometric phase. In fact, the difference between the unwrapped real interferometric phase and the noise-free simulated unwrapped phase matches very well with the wrapped difference between the real and the noise-free simulated wrapped phases (compare Figs. 8 and 9). A more quantitative estimation of the unwrapping error for the real interferometric phase can be obtained by unwrapping this wrapped difference and comparing the result with the difference between the unwrapped phase and the noise-free simulated unwrapped phase (Fig. 10 and Table II). This again gives strong evidence that the unwrapping error is concentrated in the same layover and very low coherence areas at the boundaries of the image as for the simulated data, while it is zero for the large majority of pixels.

Finally, we show a perspective view of the unwrapped real phase (Fig. 11) to give just an indication of one of the main applications of phase unwrapping in SAR interferometry; that

is, the reconstruction of the elevation of the observed scene, which is related to the value of the unwrapped phase [1], [2].

IV. Conclusion

We propose a new method for automated phase unwrapping that has been shown to be accurate and efficient. The key points are to formulate the phase unwrapping problem exploiting globally its integer qualities, which ensures accurate results, and to recognize the network structure underlying our formulation of the phase unwrapping problem, which makes for an efficient solution. The tests performed demonstrate the validity of this approach.

Acknowledgment

The author wishes to acknowledge D. Massonet of the CNES, Toulouse, France, for providing the Etna DEM used for the phase simulation, and F. Rossi, of the Pure and Applied Mathematics Department, University of L'Aquila, Italy, for useful discussions on network programming algorithms. He would also like to acknowledge A. Farina of the Systems Area, Alenia-Finmeccanica Spa, Rome, Italy, and F. Zirilli of the Mathematics Department, University "La Sapienza," Rome, for their advice. Finally, he would like to thank S. Coulson, R. Cravey, B. Rosich Tell, N. Walker, and many other people at ESRIN, European Space Agency, Frascati, Italy, for their generous support.

References

[1] L. C. Graham, "Synthetic interferometer radar for topographic mapping," *Proc. IEEE*, vol. 62, pp. 763–768, June 1974.
[2] H. A. Zebker and R. M. Goldstein, "Topographic mapping from interferometric synthetic aperture radar observations," *J. Geophys. Res.*, vol. 91, pp. 4993–4999, Apr. 1986.
[3] A. V. Oppenheim and J. S. Lim, "The importance of phase in signals," *Proc. IEEE*, vol. 69, pp. 529–541, May 1981.
[4] R. M. Goldstein, H. A. Zebker, and C. L. Werner, "Satellite radar interferometry: Two-dimensional phase unwrapping," *Radio Sci.*, vol. 23, pp. 713–720, July–Aug. 1988.
[5] D. L. Fried, "Least-squares fitting a wave-front distortion estimate to an array of phase-difference measurements," *J. Opt. Soc. Amer.*, vol. 67, pp. 370–375, Mar. 1977.
[6] R. H. Hudgin, "Wave-front reconstruction for compensated imaging," *J. Opt. Soc. Amer.*, vol. 67, pp. 375–378, Mar. 1977.
[7] B. R. Hunt, "Matrix formulation of the reconstruction of phase values from phase differences," *J. Opt. Soc. Amer.*, vol. 69, pp. 393–399, Mar. 1979.
[8] H. Takajo and T. Takahashi, "Noniterative method for obtaining the exact solution for the normal equation in least-squares phase estimation from the phase difference," *J. Opt. Soc. Amer. A*, vol. 5, pp. 1818–1827, Nov. 1988.
[9] D. C. Ghiglia and L. A. Romero, "Robust two-dimensional weighted and unweighted phase unwrapping that uses fast transforms and iterative methods," *J. Opt. Soc. Amer. A*, vol. 11, pp. 107–117, Jan. 1994.
[10] M. D. Pritt, "Phase unwrapping by means of multigrid techniques for interferometric SAR," *IEEE Trans. Geosci. Remote Sensing*, vol. 34, pp. 728–738, May 1996.
[11] D. C. Ghiglia and L. A. Romero, "Minimum L^p-norm two-dimensional phase unwrapping," *J. Opt. Soc. Amer. A*, vol. 13, pp. 1999–2013, Oct. 1996.
[12] G. W. Davidson and R. Bamler, "Multiresolution phase unwrapping for SAR interferometry," *IEEE Trans. Geosci. Remote Sensing*, to be published.
[13] A. Ferretti, C. Prati, F. Rocca, and A. Monti Guarnieri, "Multibaseline SAR interferometry for automatic DEM reconstruction," in *Proc. 3rd ERS Symp.*, Florence, Italy, ESA SP–414, 1997; (http://florence97.ers-symposium.org/data/ferretti/index.html).
[14] M. Costantini, "A phase unwrapping method based on network programming," in *Proc. "Fringe '96" Workshop*, Zurich, Switzerland, ESA SP-406, pp. 261–272; (http://www.geo.unizh.ch/rsl/fringe96/papers/costantini/).
[15] ———, "Validation of a novel phase unwrap algorithm using true and simulated ERS tandem SAR interferometric data," in *Proc. 3rd ERS Symp.*, Florence, Italy, ESA SP–414, 1997, vol. 3, pp. 1701–1708; (http://florence97.ers-symposium.org/data/costantini/index.html).
[16] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network Flows: Theory, Algorithms, and Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
[17] CNES, *Logiciel de traitement interferometrique automatise DIAPASON*, Toulouse, France, CNES/833/2/95/0145, 1996.
[18] J. S. Lee, K. W. Hoppel, S. A. Mango, and A. R. Miller, "Intensity and phase statistics of multilook polarimetric and interferometric SAR imagery," *IEEE Trans. Geosci. Remote Sensing*, vol. 32, pp. 1017–1028, Sept. 1994.
[19] CPLEX, *Using the CPLEX Callable Library—Version 4.0*, CPLEX Optimization, Inc., Incline Village, NV, 1995.
[20] D. P. Bertsekas and P. Tseng, "Relaxtion methods for minimum cost ordinary and generalized network flow problems," *Oper. Res.*, vol. 36, pp. 93–114, Jan.–Feb. 1988.

Mario Costantini received the dottore degree in physics in 1991 from the University "La Sapienza," Rome, Italy.

He was with the University "La Sapienza," Alenia-Finmeccanica Spa, and Vitrociset Spa, Rome. Since 1996, he has been a Research Fellow of the ESRIN, European Space Agency, Frascati, Italy. He has conducted research in the fields of statistical mechanics and fractal geometry. In the last few years, his scientific activity has focused on applied mathematics and, in particular, image processing.