Phase unwrapping algorithms for radar interferometry: residue-cut, least-squares, and synthesis algorithms

Howard A. Zebker and Yanping Lu

Department of Geophysics, Stanford University, Stanford, California 94305-2215

Received May 5, 1997; accepted September 18, 1997; revised manuscript received October 9, 1997

The advent of interferometric synthetic aperature radar for geophysical studies has resulted in the need for accurate, efficient methods of two-dimensional phase unwrapping. Inference of the lost integral number of cycles in phase measurements is critical for three-pass surface deformation studies as well as topographic mapping and can result in an order of magnitude increase in sensitivity for two-pass deformation analysis. While phase unwrapping algorithms have proliferated over the past ten years, two main approaches are currently in use. Each is most useful only for certain restricted applications. All these algorithms begin with the measured gradient of the phase field, which is subsequently integrated to recover the unwrapped phases. The earliest approaches in interferometric applications incorporated residue identification and cuts to limit the possible integration paths, while a second class using least-squares techniques was developed in the early 1990's. We compare the approaches and find that the residue-cut algorithms are quite accurate but do not produce estimates in regions of moderate phase noise. The least-squares methods yield complete coverage but at the cost of distortion in the recovered phase field. A new synthesis approach, combining the cuts from the first class with a least-squares solution, offers greater spatial coverage with less distortion in many instances.
 [S0740-3232(98)01403-3]
OCIS codes: 280.0180, 280.6730, 120.3180.

1. INTRODUCTION

Algorithms that relate individual phase measurements on a two-dimensional field, motivated largely by interest in interferometric synthetic aperture radar (SAR) techniques, have proliferated over the past ten years.[1–4] These algorithms seek to infer the integral number of cycles lost when a phase measurement is made from a two-dimensional complex signal amplitude observation, which uniquely identifies only the phase value modulo 2π. We refer to such algorithms here as phase unwrapping, as distinguished from the use of that term for reconstruction of signal amplitudes from frequency-domain phase data, a common problem associated with one-dimensional signal processing.

We report here a comparison of several of the phase unwrapping algorithms in more common use today and identify and contrast the strengths and weaknesses of each. We also present a synthesis approach that combines some of the more effective features of the existing algorithms and can extend the range of phase unwrapping situations amenable to automated solution. Rather than revolutionize present capabilities, these new algorithms represent another approach for phase unwrapping procedures that aid in some situations where the existing algorithms fare poorly. Because we (1) do not review each existing algorithm comprehensively, (2) do propose new variations of the algorithms suited to particular styles of input data, and (3) do not believe that the existing set of approaches is the final word on phase unwrapping procedures, this article serves more as a progress report rather than as a review of phase unwrapping procedures.

The advent of interferometric SAR for geophysical studies, in particular, has resulted in the need for accurate, efficient methods of two-dimensional phase unwrapping. Radar methods for fast and precise measurement of topographic data,[5,6] determination of centimeter-level surface deformation fields,[7–9] and surface velocity fields[10,11] all require that the relative phases over large areas be known. Strictly speaking, each of these techniques requires, in addition, knowledge of the absolute phase values;[12] however, in practice contextual clues or known fiducial values (tie points) often permit geophysical inference with only the relative phases, given that the phase field is unwrapped.

Although many insights into underlying surface processes may be obtained by visual inspection of the initial, wrapped radar interferogram,[13,14] this unwrapping by eye may be applied only to simple phase fields without significant and complicated structure. Moreover, automated topographic mapping approaches and the application of three-pass deformation algorithms are precluded by the necessity of human interaction on a pixel-by-pixel level with the interferometric data. For instance, even a small-area topographic map may contain millions of meter-spaced posts. Accuracy also drives the requirement. In regions of finely spaced fringes, it may be difficult to estimate the phase manually to better than a large fraction of a cycle accuracy, whereas performance of the radar system itself allows accuracies corresponding to

perhaps 3 deg of phase.[12,15] Requirements for future radar system performance typically imply that the interferogram phase can be determined to similar if not better levels.

All of the commonly used phase unwrapping algorithms relate the phase values by first differentiating the phase field and subsequently reintegrating, adding back the missing integral cycles to obtain a more continuous result. The many algorithms proposed for phase unwrapping over the past few years fall into two basic classes: (1) algorithms based on identification of residues in the wrapped phase field and cuts, or "trees," connecting a group of residues to limit the integration paths, and (2) algorithms that derive a smooth field by integrating the gradient of the observations subject to smoothness constraints as determined by least-squared-difference criteria. Least-squares algorithms may be weighted or unweighted; that is, they may consider the closeness of fit as dependent on ancillary data for each measurement point such as the radar brightness or correlation, giving greater weight to those points deemed to be more accurate. Note that weighted least-squares solutions need not be smooth or even continuous, if zero weights are assigned to some points in the phase data. This property will form the basis of our proposed improvement on the traditional weighted least-squares solution, consisting essentially of a method to assign the zero weights that is dependent on phase observables.

Another solution, derived from a mathematical theory utilizing a Green's-function reconstruction,[4] has recently been shown to be equivalent to the least-squares solution, differing only in terms of computational efficiency.[16] Other approaches in today's literature use nested combinations of least-squares techniques[17] and methods employing measures of data integrity to guide unwrapping paths,[18] as well as several employing neural network or "genetic" algorithms.[19] However, these latter approaches have not gained widespread acceptance and we will not analyze them here. They may be useful, or even outperform the more common algorithms, for specialized situations.

This paper is constructed as follows. First, we briefly illustrate and summarize the phase unwrapping problem and describe the solution approaches of the two principal algorithm classes. We also introduce a synthesis algorithm that combines features of both classes in a single algorithm. Next, we illustrate the performance of each algorithm in a variety of situations that differ in the character of the phase field to be unwrapped. We conclude by stating that phase unwrapping remains a significant issue in interferometric SAR analysis and that the synthesis approach presented here contributes one more technique that is to be added to the collection of algorithms available for any given application.

2. PHASE UNWRAPPING: THE PROBLEM AND TWO SOLUTION APPROACHES

In interferometric SAR analysis, two radar images of the same surface area are acquired and differenced in phase, forming a radar interferogram. This is usually implemented by cross multiplying the complex reflectivity at each point of one image by its corresponding conjugate value in the second image, so that the interferogram also preserves useful information about the signal amplitudes.

The complex signal amplitude reflected from each radar resolution element in a single image consists of the vector sum of contributions from many individual scattering elements within that resolution element. Because the size of the resolution element is typically many wavelengths, randomness in the spatial distribution of the scatterers yields an approximately uniform phase distribution in the resultant coherent sum of the scattered waves. The phases, however random, are nonetheless deterministic in that they are calculable from knowledge of the precise locations of each scattering center. If the pair of images forming the radar interferogram are duplicates precisely regarding viewing geometry and antenna polarization, if receiver noise effects are minimal, and if the surface itself is unchanged between observations, the two images will be identical and the interferogram will exhibit zero phase everywhere.

On the other hand, if the viewing geometry is altered within certain limits, the phase differences comprising the interferogram will vary across the terrain in a manner related to the surface topography.[5,20] In addition, if the viewing geometry is unchanged but one region of the surface itself is displaced spatially with respect to the rest of the image, the pixels corresponding to that section will exhibit phase differences proportional to the line-of-sight component of the displacement.[7] Finally, if displacements are not coherent across each resolution element but instead randomize the positions of each scatterer with respect to others inside the element, the pair of echoes will be less well related and the coherence, or correlation, of the interferogram will decrease,[21] resulting in a noise that may mask the underlying phase signature.

The interferometric phase signatures vary relatively smoothly from point to point in the interferogram and may be inverted to recover surface topography, velocity, or displacement fields. However, the phase observables are measured modulo 2π; that is, the integral number of phase cycles on each measurement is lost. Consequently, if the surface displacement in a scene is greater than one-half radar wavelength and the resulting interferogram phase excursion is greater than one cycle, or if the combination of interferometric baseline and surface topography yields more than one fringe of topographic signature, the interferogram cannot be uniquely inverted without a procedure to recover the missing cycles. We refer to such procedures as phase unwrapping.

Phase unwrapping algorithms share a common initial approach: the phase change, or gradient, from one point to the next in an interferogram is computed and then integrated to form a single, smoother phase function incorporating the previously missing cycles. To date all algorithms applied to interferometric SAR have used variations of measured phase gradient integration procedures.

The first constraint on a phase unwrapping algorithm is that it produce consistent results; that is, the same phase field should be recovered independently of the direction and order chosen for the phase difference integra-

tion. The problem can be illustrated by the following set of measurements, corresponding to a four-by-four pixel interferogram array, each expressed in cycles:

0.2	0.0	0.8	0.0
0.4	0.2	0.2	0.4
0.6	0.8	0.8	0.6
0.8	0.8	0.8	0.8

We assume that the data are adequately sampled; to satisfy Nyquist we cannot have a jump of more than one-half cycle from point to point. To unwrap these data, we can integrate the top line across from left to right, adding integral cycles to ensure that the jump from one point to the next in all directions is less one-half cycle, and then integrate down each column. The result of this procedure is

0.2	0.0	−0.2	0.0
0.4	0.2	0.2	0.4
0.6	−0.2	−0.2	0.6
0.8	−0.2	−0.2	0.8

Now, transpose the directions of integration. First unwrap down the leftmost column and then across each row:

0.2	0.0	−0.2	0.0
0.4	0.2	0.2	0.4
0.6	0.8	0.8	0.6
0.8	0.8	0.8	0.8

These results are different, and only one of these solutions at most can be correct. They differ in that four of the measurements are one cycle greater in the second instance than in the first. This inconsistency must be eliminated in any practical algorithm.

A little reflection leads to the conclusion that for a consistent result, any closed line integral in the phase field must equal zero; that is, the phase along the path must return to its starting value. Using this approach, the inconsistency in the above example can be traced to the existence of residues in the measured phase field, regions that are self-inconsistent. Consider a short circular phase path extracted from the above example:

0.4	0.2
0.6	0.8

Integrating clockwise around the loop yields a net −1 cycle instead of the zero sum required by consistency. We thus term this set of four phase measurements a negative phase residue, drawing on a mathematical analogy with Cauchy integration in the complex plane. In fact any closed path on the phase field containing this residue will yield the same net −1 cycle unless the path contains additional residues. If the net charge within the integration paths is zero, the result will be consistent. Therefore it is the goal of the phase unwrapping procedure to eliminate potential integration paths enclosing unequal numbers of positive and negative residues. The residue-cut algorithms do this explicitly, whereas the least-squares procedures described to date exhibit artifacts generated by the existence of the residues. The synthesis approach proposed here incorporates residue avoidance with the least-squares solution.

Residues derive from two sources in the radar measurements.[1] The first is actual discontinuities in the data. The fringe spacing may be so fine on certain topographic slopes or from large interobservation displacement as to exceed the Nyquist criterion of half-cycle spacing. The second is noise in the data set, whether from thermal and other noise sources or from decorrelation due to baseline length and temporal change in the scene.[20–22] Residues from whatever source require compensation in the phase unwrapping procedure.

A. Residue-Cut Tree Algorithms

The initial residue-cut phase unwrapping procedure proposed by Goldstein *et al.*[1] is implemented by first identifying the locations of all residues in an interferogram and then connecting the residues with branch cuts to prevent the existence of integration paths that can encircle unbalanced numbers of positive and negative residues. Because of the terminology of branch cuts and the dendritic appearance of the complete set of cuts, the interconnection of the residues was referred to as growing “trees” by the authors.

The residue-cut algorithm of Goldstein *et al.*, in common use today, is a relatively conservative algorithm in that it tends to grow rather dense networks of trees in residue-rich regions. The algorithm initially connects closely spaced, oppositely charged, pairs of residues with cuts that prevent integration paths between them, helping to ensure consistency. If all permitted integration paths enclose equal numbers of positive and negative residues—that is, each tree connects the same number of plus and minus charges and is in that sense uncharged—consistency is assured. Progressively longer trees are permitted until all residues are connected to at least one other residue and until the net charge on each tree is zero. Networks of small trees are used to prevent any single branch from becoming too long and isolating large subareas from the rest of the image.

Specifically, the tree-growth algorithm of Goldstein *et al.*[1] consists of the following steps:

1. Residues are identified and marked as either positive or negative.
2. The interferogram phase field is scanned systematically until a residue is located.
3. From that residue, the surrounding area is searched until another residue is found. A cut is formed between the residues and the total charge is computed.
4. If the total charge along the tree is zero, the tree is considered complete and the systematic scan of step (2) is continued.
5. If the total charge is nonzero, the search continues for nearby residues. As each residue is encountered, it is connected to the tree by means of a branch cut and the total charge is computed. Only when the total charge is zero is the tree considered complete and the step (2) scan allowed to proceed.

In the course of the search for residues, all residues, regardless of whether each has been previously assigned to

a tree, are added to the new tree. This yields the dendritic appearance of the cuts and the nomenclature "trees."

Finally, when a boundary of the scene is encountered, a cut is drawn to the boundary and the tree is deemed complete and uncharged. This superconducting property of the edges prevents overly long trees from isolating large areas of the image.

A consequence of the indiscriminate branch growth until charge neutrality is achieved for all trees, and all residues are accounted for, is that in residue-rich regions, such as heavily laid over or very noisy regions, the tree growth is so dense that the region is isolated from the remainder of the image and no unwrapped phase estimate can be obtained. This conservative approach nearly eliminates mistakes at the expense of providing an incomplete unwrapping result.

An unpublished variant of this algorithm was developed by Atsushi Hiramatsu at Caltech in the early 1990's; nevertheless, the algorithm and code survive among the phase unwrapping underground. In the Atsushi approach, residues are determined as before, but the trees are prevented from closing on themselves and thus a solution is forced at every point in the image. This solution, although consistent and complete, is not always correct—the algorithm typically generates regional errors in the noisy and laid over portions of the interferogram image. In many situations the gain from complete unwrapping solutions outweighs the cost of the errors, but this must be determined on a case-by-case basis.

Another version of the residue-cut algorithm again begins with the method of Goldstein *et al.*[1] but allows operator interaction to edit the trees manually before phase integration, thus permitting more integration paths than the conservative initial algorithm allows. In this manner integration paths may be opened up into areas that were too densely packed with residues to unwrap. Here again the trade-off is for increased areal coverage at the expense of possible errors, but the additional coverage and errors may be precisely controlled by the skill of the operator in identifying appropriate trees to prune. However, it transforms the procedure from a fully automated one to one requiring intense operator interaction on each interferogram. For limited applications this may be acceptable. (This editing approach was used by Zebker *et al.*[9] to maximize the unwrapped area in the analysis of surface deformation resulting from the Landers 1992 earthquake.)

B. Least-Squares Algorithms

The second major approach to phase unwrapping in common use today was presented by Ghiglia and Romero,[2] who applied a mathematical formalism first developed by Hunt[23] to the radar interferometry phase unwrapping problem. Hunt developed a matrix formulation suitable for general phase reconstruction problems; Ghiglia and Romero found that a discrete cosine transform technique permits accurate and efficient least-squares inversion even for the very large matrices encountered in the radar interferometry special case. They examined both unweighted and weighted least-squares solution procedures. In the least-squares methods, the vector gradient of the phase field is determined and then integrated, subject to the constraint of a smooth solution.

One major difference between the residue-cut and least-squares solutions is that in the residue-cut approach, only integral numbers of cycles are added to the measurements to produce the result. In the least-squares approach, any value may be added to ensure smoothness and continuity in the solution. Thus the spatial error distribution may differ between the approaches, and the relative merits of each method must be determined, depending on the application.

The unweighted algorithm is implemented as follows. Consider a sampled wrapped phase function $\varphi_{i,j}$, evaluated at discrete points i,j corresponding to the row and column locations, respectively, of a two-dimensional data matrix. We want to determine a smooth, unwrapped phase function $\phi_{i,j}$ that minimizes the difference between the gradients calculated from the wrapped phase and the presumed smooth, unwrapped phase. Hunt[23] shows that these may be related by a matrix-vector equation

$$\mathbf{s} = \mathbf{P}\phi + \mathbf{n} \tag{1}$$

in which $\mathbf{s}$ is derived from the measured row and column phase differences of φ, $\mathbf{P}$ is a matrix containing 1's, -1's, and zeros describing row and column differencing operations, ϕ is the unwrapped phase field, and $\mathbf{n}$ is a vector representing measurement noise. The least-squares solution is the well-known

$$\phi = (\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T\mathbf{s}. \tag{2}$$

Specifically, in the radar interferometry case we minimize the function

$$\sum_{i=0}^{M-2}\sum_{j=0}^{N-1}(\phi_{i+1,j} - \phi_{i,j} - \Delta^x_{i,j})^2 + \sum_{i=0}^{M-1}\sum_{j=0}^{N-2}(\phi_{i,j+1} - \phi_{i,j} - \Delta^y_{i,j})^2, \tag{3}$$

where $\Delta^x_{i,j}$ and $\Delta^y_{i,j}$ are the row differences and column differences of the wrapped phases, respectively. The row and column differences are calculated from the wrapped phases as

$$\Delta^x_{i,j} = \varphi_{i+1,j} - \varphi_{i,j},$$
$$\Delta^y_{i,j} = \varphi_{i,j+1} - \varphi_{i,j}, \tag{4}$$

with appropriate cycles added to ensure that $-\pi \leqslant \Delta^x_{i,j}, \Delta^y_{i,j} \leqslant \pi$.

The least-squared-error solution is obtained by differentiating Eq. (3) with respect to $\phi_{i,j}$ and setting the result equal to zero, so that

$$(\phi_{i+1,j} - 2\phi_{i,j} + \phi_{i-1,j}) + (\phi_{i,j+1} - 2\phi_{i,j} + \phi_{i,j-1}) = (\Delta^x_{i,j} - \Delta^x_{i-1,j}) + (\Delta^y_{i,j} - \Delta^y_{i,j-1}). \tag{5}$$

Equation (5), the unweighted case, is a discrete form of Poisson's equation and may be solved efficiently by using a discrete cosine transform approach.[2]

Now, if some points in an interferogram are deemed more reliable than others, for example, possessing higher

correlation, a weighted algorithm may be used. Including in Eq. (1) a matrix $\mathbf{W}$ of weights for each point yields

$$\mathbf{Ws} = \mathbf{WP}\phi + \mathbf{n}. \tag{6}$$

The resulting normal equations for the weighted least-squares problem are then expressed as

$$\phi = (\mathbf{P}^T\mathbf{W}^T\mathbf{WP})^{-1}\mathbf{P}^T\mathbf{W}^T\mathbf{Ws}. \tag{7}$$

Unfortunately, the weighted least-squares equation does not reduce to the same simple Poisson's equation form, and thus the efficient discrete cosine algorithm cannot be directly applied. Ghiglia and Romero[2] show, however, that several iterative approaches using repeated discrete cosine transforms are yet relatively efficient at achieving an accurate solution. Following Ghiglia and Romero, we used conjugate gradient methods to solve Eq. (7) iteratively. For the convergence criterion we chose

$$\frac{\sum_{\text{all } i,j} |\phi_{i,j}^{k+1} - \phi_{i,j}^{k}|^2}{\sum_{\text{all } i,j} |\phi_{i,j}^{k+1}|^2} \leqslant \epsilon,$$

where ϵ is set equal to a conservative value of 10^{-12}, which for a single-precision computation ensures that we are limited by round-off errors.

The accuracy of the solution depends on the choice of weights to apply to each point of the measured phase. Choosing the same value for all weights reduces the problem to the unweighted solution. More typical choices are functions of the signal-to-noise ratio or the observed interferometric correlation, both of which give greater emphasis to the phase estimates deemed more reliable. Various methods for selecting weights are described by Pritt.[17]

All of the least-squares algorithms produce a continuous solution unless zero weights are assumed at some data points. In fact this constraint is fundamental to the way the problem is framed. Thus we would expect the algorithm to perform poorly in the presence of actual discontinuities in the underlying phase field, as can be present if there is any layover or extreme foreshortening in the radar image, unless the locations of the zero weights are properly and carefully chosen. Weighted solutions derived from correlation or signal-to-noise measures can lessen the errors by tying down the solution less in the discontinuous areas; but some, albeit smaller, smoothness is assumed everywhere the weights are nonzero. If the continuity constraint is removed completely in laid over regions but not in the remainder of the image, the weighted least-squares solution can minimally distort the result. This will be the synthesis approach that we present in Section 3.

One advantage of the least-squares algorithms over residue-cut algorithms is that results may be obtained more readily in the residue-rich regions, permitting use on noisy data that would have been difficult or impossible to unwrap because of the dense tree network in the Goldstein *et al.*[1] algorithm. Although the Atsushi algorithm would force unwrapping in the noisy regions, enough phase unwrapping errors are often present to make these data nearly unusable. The least-squares method provides estimates in these regions that in many cases are more accurate than those provided with the Atsushi residue-cut algorithm.

Neither of the above least-squares algorithms explicitly utilizes residue information. As we discussed above, the residues are inherent in any measured data set and must be accounted for in the phase unwrapping procedure to ensure consistency and minimize error in the final result. We will demonstrate in the next section that the existence of residues in a phase field induces distortion in the least-squares solution unless they are compensated for. This compensation forms the basis of a synthesis algorithm that we will introduce below, which under some conditions permits more complete unwrapping than would be possible with residue-cut algorithms yet less distortion than is present under least-squares techniques.

C. Other Existing Algorithms

Additional algorithms proposed for interferometric SAR phase unwrapping applications include Green's-function approaches,[4] multigrid algorithms,[17] methods using estimates of data integrity to guide phase unwrapping paths,[18] and neural-network or genetic algorithms.[19] The Green's-function methods have recently been shown to be mathematically equivalent to least-squares solutions,[16] but they differ in claimed computational efficiency. Although it is beyond the scope of the present paper to consider these efficiency issues in detail, a comprehensive study of them would be helpful for anyone considering a large-scale application. The multigrid methods achieve an increase in efficiency with a nested procedure for evaluating phase fields on different size scales, but the fundamental unwrapping considerations are the same as described in the least-squares algorithms discussed above. Data integrity algorithms unwrap preferentially towards regions of, say, high correlation, but without explicit reference to the residue locations. The genetic algorithms are random-path integration procedures with probabilistic quality metrics that seem to work well on data with sparse residue distributions, but they have not received the widespread testing and application of the residue-cut or least-squares approaches. Thus they have not yet been challenged by the variety of vexing phase unwrapping situations, occurring in actual data analysis, that the more common algorithms have confronted.

3. SYNTHESIS ALGORITHM

In this section we propose a new algorithm for phase unwrapping designed to overcome limitations in both of the above major approaches. As we shall illustrate below, the principal limitation of the residue-cut method is that the tree density may become so pronounced that large areas of the scene cannot be unwrapped. In more extreme cases the image is subdivided into small, unwrappable islands that we cannot relate to one another. Forcing the residue-cut algorithm to unwrap the dense regions, as is done in the Atsushi algorithm, leads to unpredictable and incorrect phase estimates in the dense areas that may propagate long distances in the image.

The strength of the least-squares approach is that phase values are obtained everywhere in the image. But we will observe below that errors in the least-squares solutions follow from the assumption that the unwrapped phase field is everywhere continuous. This will be seen to induce distortions that tend to underestimate recovered phase slopes. The Goldstein *et al.*[1] algorithm does not suffer from this problem, as the cuts in the surface enable the integration to proceed without enforcing continuity at these sites. Adding this capability to the least-squares result greatly improves performance of the weighted least-squares approach, and it results in the synthesis algorithm. In this sense the synthesis approach is an application of least-squares-solution weights defined by the residue-cut algorithm. The methods of solution, once the weights are determined, are identical to those given in the least-squares discussion above.

We add the capability by first calculating the trees, using either the Goldstein *et al.* or the Atsushi residue-cut algorithm to calculate the surface cuts that define the integration paths and then adopting zero as the weight for the points that lie along the trees. If these points are given zero weight in the unwrapping algorithm, there is no constraint implying continuity at these points. The remaining points are weighted according to their interferometric correlation. Therefore solutions with true discontinuities (layover) are permitted, as are solutions in very noisy areas where the traditional Goldstein *et al.* al-

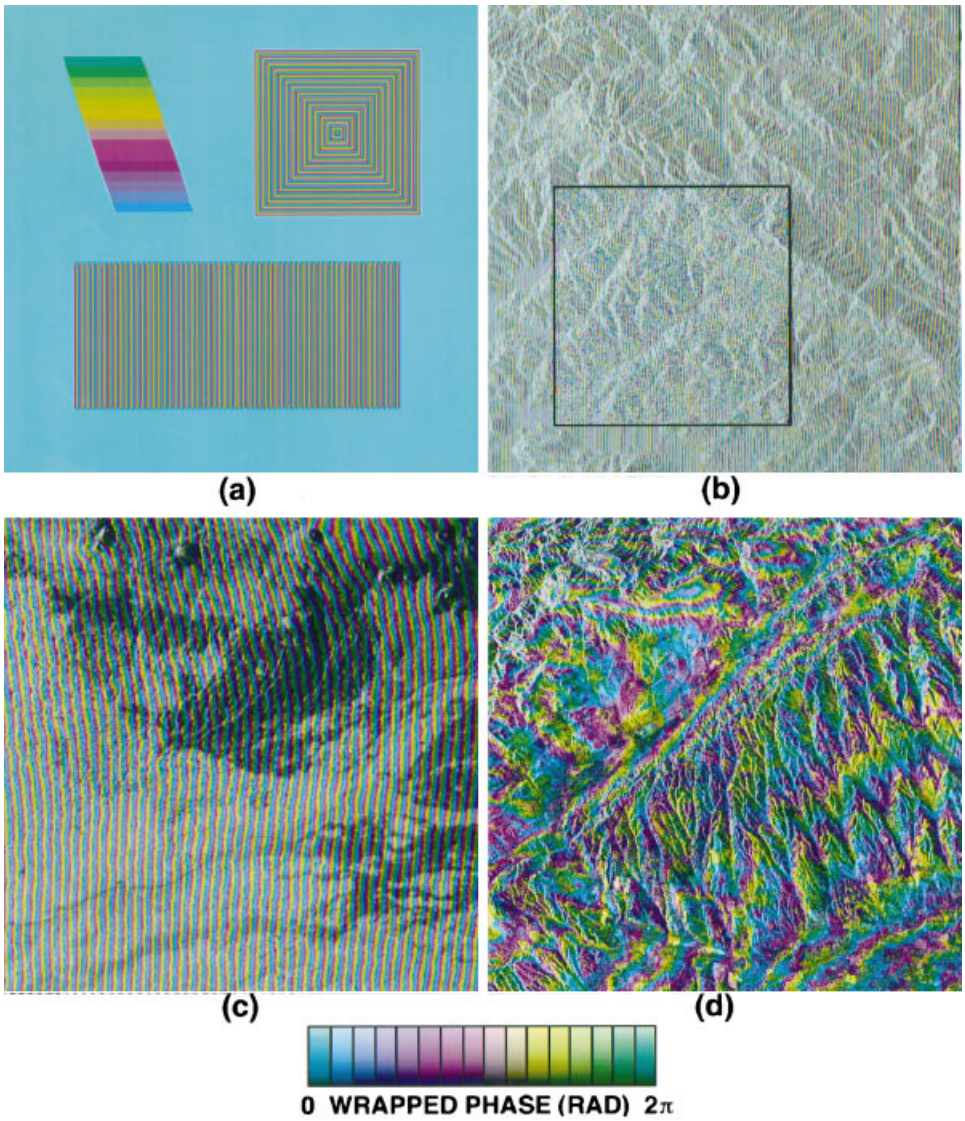

Fig. 1. Test data sets for algorithm intercomparison, in wrapped phase (raw interferogram) form. Two sets, (a) and (b), are simulated data. One contains simple geometric shapes designed to illustrate the importance of residue distributions and proper tree locations, and one, derived from topographic data, includes layover and thermal noise. The box outlined in the lower left of the scene depicts added noise. Actual data comprise two more scenes: (c) a very high signal-to-noise interferogram of a fairly flat area in Hawaii and (d) rugged terrain in central California. Radar brightness appears as the intensity at each point and the phase as color, with one color cycle corresponding to one fringe. (a) The simple geometrical target scene includes a pyramid structure that exhibits no phase discontinuities at its edges, a two-sided ramp structure that is continuous with the background on the left and right edges but discontinuous along the top and bottom, and a slanted wedge structure with exactly 2π phase change along its length. (b) The simulated topographic interferogram is most rugged in the upper left-hand corner, leading to significant layover in this area and less elsewhere. Scene (b) possesses a very high fringe rate, scene (c) a moderate rate, and in scene (d) the background flat-Earth fringes have been removed.

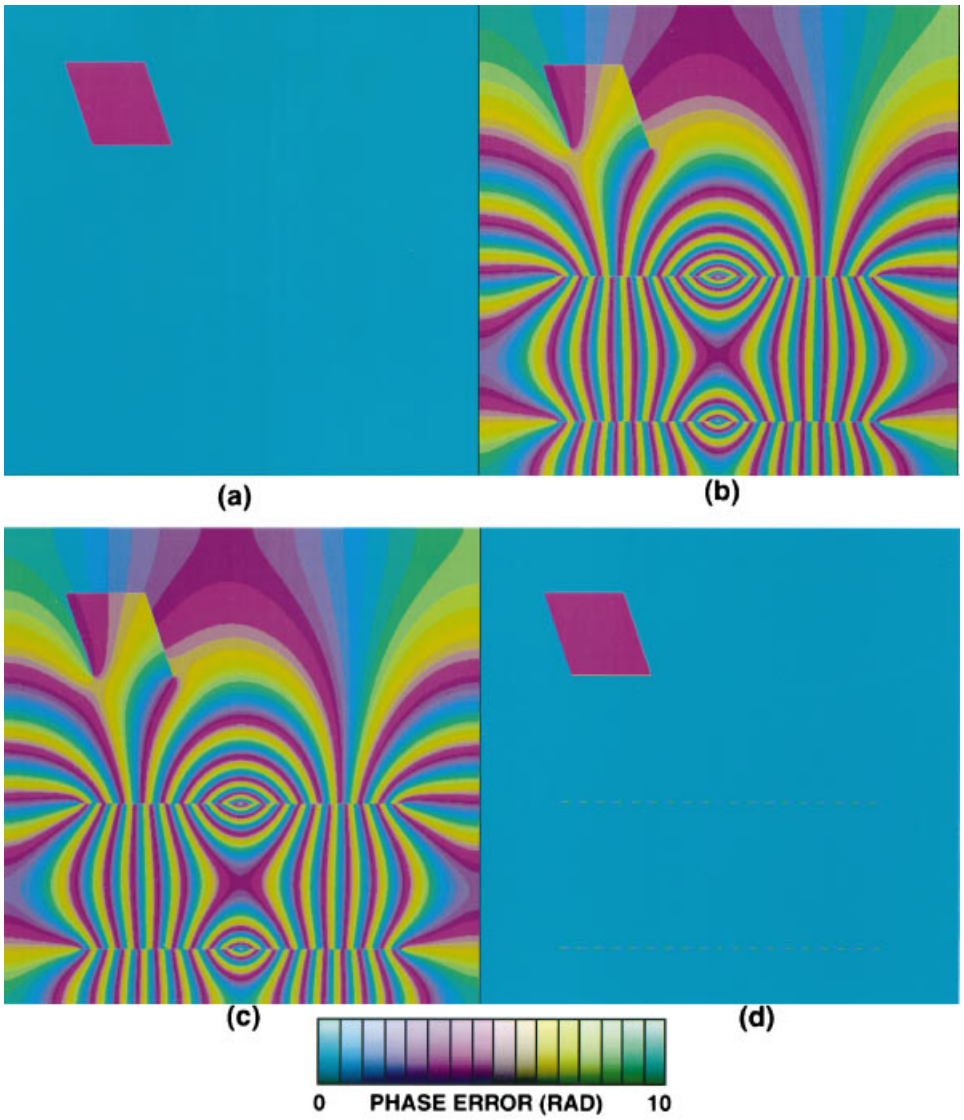

Fig. 2. Phase errors after application of four unwrapping algorithms to the geometric-shape data set. (a) Results from the Goldstein *et al.*[1] residue-cut tree algorithm; only a one-cycle error on half of the wedge is generated. The rest of the image unwraps perfectly. Both the unweighted [(b)] and the weighted [(c)] least-squares algorithms exhibit large distortion fields associated with discontinuities at the top and bottom of the ramp structure and from the half-wedge that showed the error in the tree case [(a)]. These results are identical, as the weights assigned were unity everywhere (see text). The synthesis algorithm [(d)] produces the same result as the residue-cut algorithm here.

gorithm fails to return an answer. The mathematical description of the algorithm is that of the individual parts described in the sections above.

The accuracy of the synthesis algorithm, as in the least-squares solution, depends on the appropriate choice for the weights accorded each phase estimate. Zero weight is assigned along the branch cuts of the residue-cut algorithm, and the remainder of the weights may be assigned according to correlation or signal strength as previously mentioned. We have found, though, that the greatest gain in accuracy follows from the assignment of the zero weights on the cuts, and the other weights affect the solution minimally.

It is also true that the synthesis algorithm will generate errors if the branch cuts defining the zero weights are placed in error. Unfortunately, no completely reliable algorithm for placing the weights has yet been demonstrated. Simply balancing the number of positive and negative residues along the cuts is an ambiguous method of cut identification. The search for improved methods of cut definition will have to continue if algorithms such as those presented here are to be assuredly effective.

4. PERFORMANCE OF THE ALGORITHMS

In this section we illustrate the performance of the above algorithms in different phase unwrapping situations and compose a table that summarizes the relative merits of each approach. We examine the accuracies achieved and errors generated by several simple geometrical surface shapes, as well as interferograms generated by using topographic data and radar imaging geometries leading to

significant layover. We also inject noise into the simulated data to illustrate performance as a function of signal-to-noise level.

The performance of both phase unwrapping approaches has been shown in the existing literature to be excellent in the case of high signal-to-noise ratio and continuous, adequately sampled underlying phase fields. In other words, these cases correspond to very low numbers of residues. We therefore skip the simple illustrations and proceed to cases in which the algorithms begin to fail.

We illustrate algorithms on several data sets, comprising simulated data, where we know the true unwrapped phase field, and also two sets of actual data where the true phases are unknown. In the former cases we calculate the errors explicitly, whereas for the latter we may still examine the results visually and comment on algorithm performance. The simulated data consist of two scenes, one containing simple geometric shapes designed to illustrate the importance of residue distributions and proper tree locations and one derived from topographic data. The topographic data scene has been used to construct a synthetic interferogram exhibiting layover, and thermal noise has been added in one region to demonstrate its deleterious affect. The real data also comprise two scenes, one a very high signal-to-noise interferogram of a fairly flat area in Hawaii acquired by NASA's Spaceborne Imaging Radar–C (SIR–C), which is straightforward to unwrap by using all algorithms, and one of rugged terrain in Central California acquired by the European Space Agency's ERS–1 satellite. This second data set poses a challenge to all algorithms.

All four input interferograms are shown in wrapped

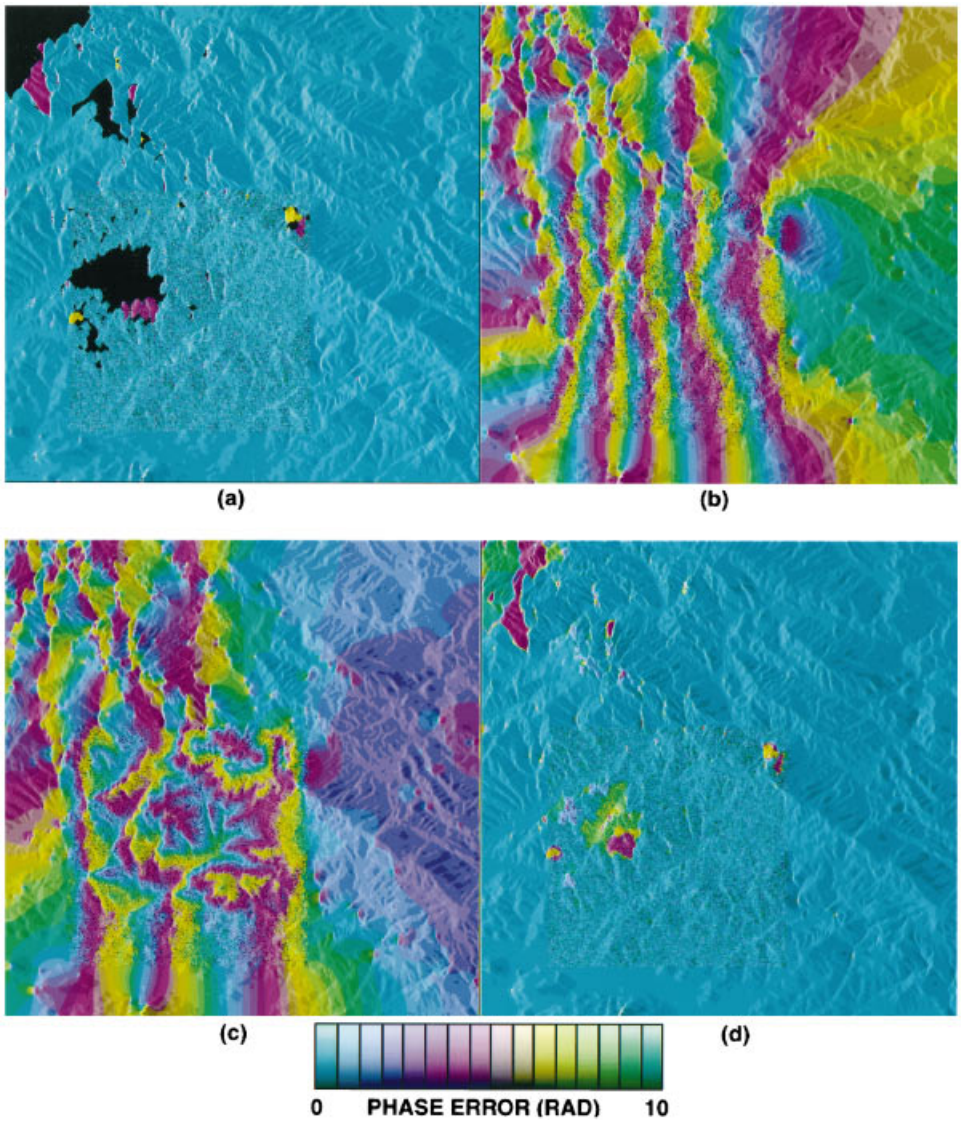

Fig. 3. Phase errors after application of four unwrapping algorithms to the synthetic topography data set. (a) Results from the Goldstein *et al.*[1] residue-cut tree algorithm. The algorithm fails to unwrap the heavily laid-over region in the upper-left and part of the noisy inset and also generates a few local one-cycle (2π) error regions. (b) The unweighted least-squares algorithm unwraps the image completely but exhibits distortion associated with residues from layover and thermal noise. It also underestimates the overall slope of the scene from left to right. (c) The results from the weighted least-squares algorithm differs in detail but is similar. (d) The synthesis algorithm produces a complete result much closer to the actual answer than the traditional least-squares approach, although errors associated with layover in the upper left and the noise in the box remain.

form in Fig. 1, where in each case we plot the radar brightness as the intensity at each point and the phase as color. One color cycle corresponds to one fringe, so that the underlying fringe density may be compared. The simple geometrical target scene [Fig. 1(a)] includes a pyramid structure that exhibits no phase discontinuities at its edges, a two-sided ramp structure that is continuous with the background on the left and right edges but discontinuous along the top and bottom, and a slanted wedge structure with exactly 2π phase change along its length. We will see that this last target embodies an error type that no existing algorithm can remove properly.

The simulated topographic interferogram [Fig. 1(b)] is most rugged in the upper left-hand corner. We have used an imaging geometry that generates significant layover in this area and less elsewhere. We have also added noise to a box approximately one quarter the size of the scene so that we may compare performance in noisy and noise-free regions—the box is visible and outlined in the lower left part of the image. For this image the signal-to-noise ratio in the box is 3 dB, and we examine the effects of varying noise levels in a subsequent section below. The two actual data sets [Figs. 1(c) and 1(d)] are the Hawaii and the California data, respectively, described above.

A. Error Magnitudes

In Fig. 2 we plot the phase errors after applying the four unwrapping algorithms to the geometric-shape data set. Panel (a) shows the results from the Goldstein *et al.*[1] residue-cut tree algorithm. The algorithm generates only a one-cycle error on half of the wedge, and the rest of the image unwraps perfectly. Since exactly 2π of phase change occurs on the wedge, it is impossible to identify the end that is continuous with the background in a wrapped image. Consequently, no algorithm will resolve this target properly. Similar structures are all too common in real interferograms, so that there will always be impossible unwrapping situations.

Both the unweighted [Fig. 2(b)] and the weighted [Fig. 2(c)] least-squares algorithms generate large distortion fields emanating from discontinuities at the top and bottom of the ramp structure and from the half-wedge that showed the error in the tree case. These two algorithms produce identical results because the weights assigned were unity everywhere. If appropriate weights had been used, the results would have been superior in the weighted solution. These synthetic data are assumed to have equal validity everywhere, leading to the choice of unit weights in the weighted case. If data were known to be on the edge of a discontinuity, reducing the weights here would produce a less distorted result. This is what the synthesis algorithm is designed to do—identify possible discontinuities and weight the least-squares solution appropriately. The synthesis algorithm [Fig. 2(d)] produces the same result as the residue-cut algorithm for this scene.

Figure 3 illustrates the difference between the unwrapped estimates and the true phase values at each point for the synthetic-topography data set. In this case also we can calculate the actual error distribution, because we know the original unwrapped solution. Again

Table 1. Rms Errors from Four Algorithms

Region	RMS Phase Error (rad)			
	Goldstein *et al.*[1]	Unweighted Least-Squares	Weighted Least-Squares	Synthesis
Scene (a)				
Pyramid	0.000002[a]	0.000002[a]	0.000002[a]	0.000002[a]
Ramp	0.000002[a]	24.7	24.7	0.00002[b]
Wedge	1.07[c]	0.92[c]	0.92[c]	1.06[c]
Scene (b)				
Noise-free area				
SNR: 1	0.827[d]	54.42	29.6	2.15
SNR: 3	0.756[d]	22.05	17.3	2.15
SNR: 10	0.744[d]	13.03	14.6	2.15
SNR: 30	0.744[d]	12.74	14.5	2.15
Noisy area				
SNR: 1	2.533[d]	35.01	200	2.57
SNR: 3	1.282[d]	11.84	8.34	1.03
SNR: 10	0.591[d]	4.48	6.14	0.32
SNR: 30	0.543[d]	4.23	6.17	0.27

[a] Quantization noise only.
[b] Errors along cuts only.
[c] Half of wedge always one cycle (2π) off.
[d] Includes only unwrapped area.

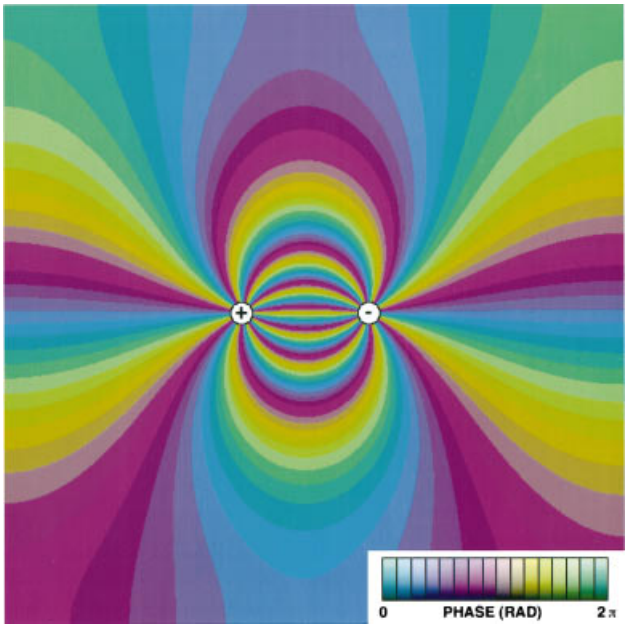

Fig. 4. Error signature of a dipole formed by a positive and negative residue pair.

panel (a) shows the results from the Goldstein *et al.*[1] residue-cut tree algorithm. The algorithm fails to unwrap the heavily laid-over region in the upper left and part of the noisy inset and also generates a few local one-cycle (2π) error regions. The results are generally accurate, but incomplete, with the exceptions being a few small parts of the laid-over and the noisy areas. The unweighted least-squares algorithm [Fig. 3(b)] produces a complete result but exhibits distortion associated with residues generated by layover and thermal noise. It also underestimates the overall slope of the scene from left to

right. The results from the weighted least-squares [Fig. 3(c)] differs in detail but is similar. Here we used the value of the interferogram correlation as the weight for each point, and the correlation is assumed to be unity everywhere except in the noise inset, where we calculate it from the signal-to-noise ratio following Zebker and Villasenor.[21] The synthesis algorithm [Fig. 3(d)] produces a complete result that is much closer to the actual answer than with the traditional least-squares approach, although errors associated with layover in the upper left and the noise in the box remain.

We calculate the root-mean-squared phase error for portions of these two synthetic data scenes and present the results for each of the four algorithms in Table 1. For the simple geometrical targets, we calculate the error over the object itself plus a border around the object 30 pixels in size so that the effect of discontinuities may be counted. For the synthetic topography scene, we calculate the errors in both the noise-free areas and the noise regions independently. The noise-free areas are still subject to residues caused by layover. For the Goldstein *et al.*[1] case, we did not count the areas that did not unwrap, so these numbers may be considered lower bounds on the error.

We note that the Goldstein *et al.*[1] algorithm performs quite well wherever it unwraps, with the single exception here being the wedge feature. The wedge as chosen here is ambiguous when shown in wrapped form, as it was devised to exhibit exactly one cycle of phase along its length. Thus when viewed as a wrapped data set, it appears continuous at both ends against the background, whereas in reality only one end can be continuous. This situation cannot be unwrapped without prior knowledge as to which end the discontinuity applies to, independent of the algorithm chosen. In this particular case two residues are found on the edges halfway along the target, leading to a cut across the middle of the wedge. Thus half the wedge unwraps properly with this algorithm, and the other half is off by one cycle.

The least-squares approaches offer completeness of solution but at the cost of underestimating large-scale slopes significantly. If we examine the error distribution

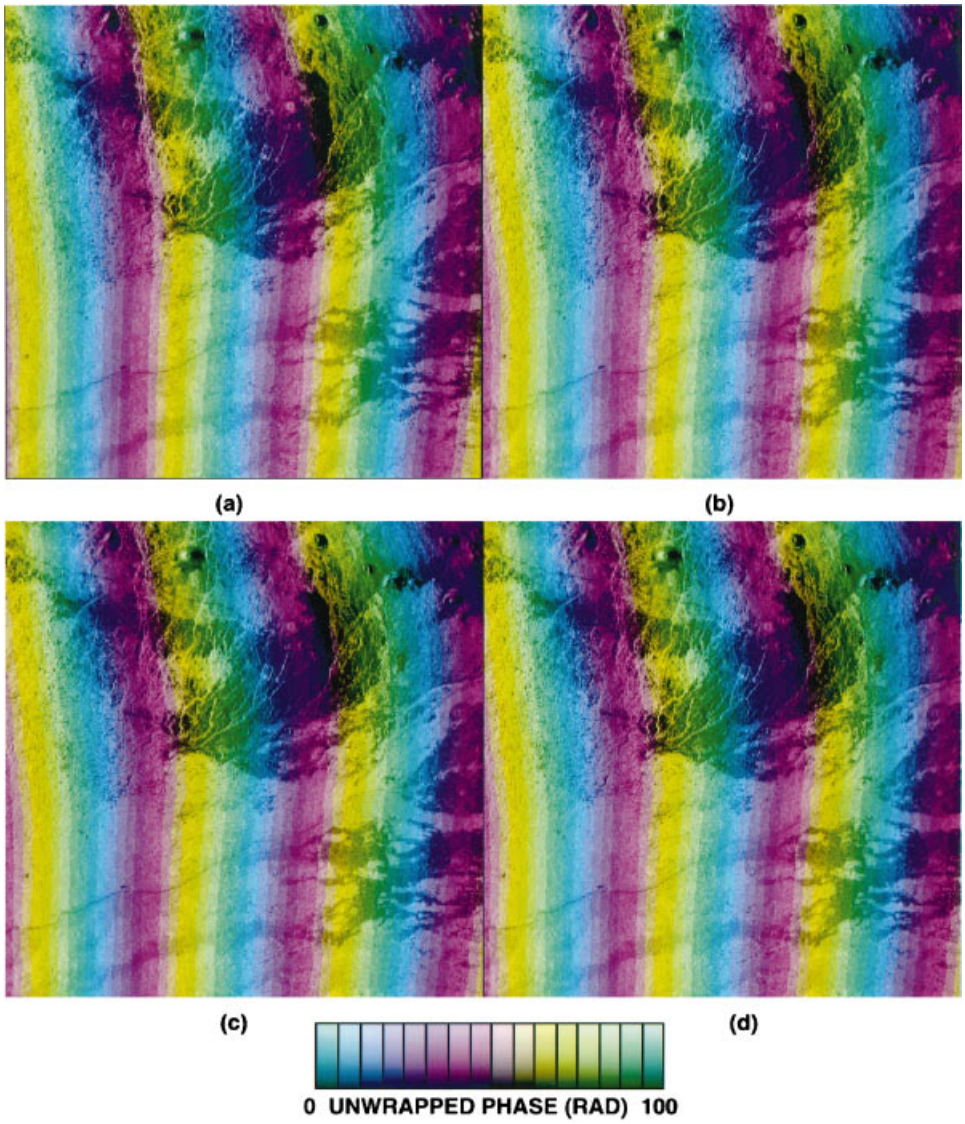

Fig. 5. Unwrapping of an easy scene, the Hawaii SIR-C data, by each of the four algorithms: (a) residue-cut trees, (b) unweighted least-squares, (c) weighted least-squares, and (d) synthesis. All unwrap completely and get nearly the same result.

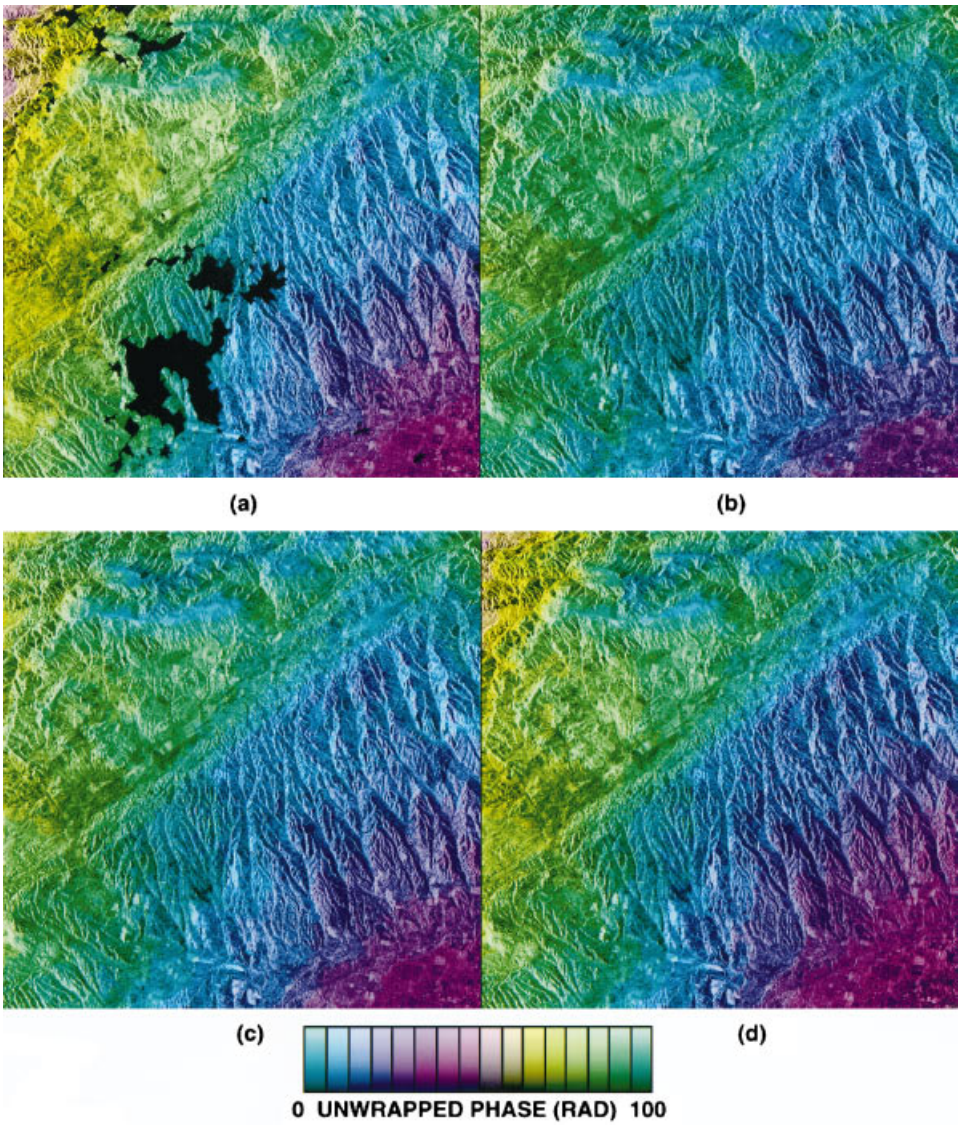

Fig. 6. Results of applying the four algorithms [(a) residue-cut trees, (b) unweighted least-squares, (c) weighted least-squares, and (d) synthesis] to a more difficult unwrapping problem, ERS-1 data over Parkfield, Calif. The residue-cut algorithm leaves significant gaps. Both of the least-squares algorithms produce subtly different results that significantly underestimate the tilt from upper left to lower right in the scene. The synthesis algorithm gives a complete answer and recovers the global tilt accurately. There are undoubtedly errors of $\sim 2\pi$ in magnitude in this solution, but they are not visible at this scale.

from the least-squares algorithm in greater detail, we note that an interesting distortion pattern is produced from each pair of residues in the input image. Illustrated in Fig. 4, the error signal radiates outward from the residue pair and is reminiscent of an electromagnetic dipole pattern. The more-complex error fields found in the more-complicated scenes are a combination of many of these individual dipole patterns. Therefore local discontinuities in the phase field generate residues and propagate errors throughout the image. These errors decay with increasing distance from the residue dipole.

The weighted method outperformed the unweighted method for the noisy region in cases of high noise and is useful if we know the location of the noisy area and set the weighting value in this area to be relatively low. For those areas with discontinuities, the results are poor because of the least-squares algorithm property to tend to smooth out the solution to make it everywhere continuous. Therefore this method works poorly for the layover region. But if we know the location of jumps and we set the weighting values equal to zero at those discontinuities, we can get a reasonably good solution. Once again, the shortcoming of the existing least-squares algorithm is in the choice of proper weights, not in any mathematical limitations of the approach.

Other techniques for generating zero weights along cuts have been examined. One approach is to threshold the correlation coefficients so that values below some cutoff are set equal to zero. This method works in laid-over regions, which tend to have low correlation. It will, however, miss placing cuts along cliffs oriented perpendicular to the flight path, as well as areas where a phase discontinuity is a result of certain ground motions between interferometric observations. These cases are identified by

using the synthesis algorithm, which generates weights without any additional knowledge of the scene characteristics.

Figures 5 and 6 present the results of applying all four algorithms to the Hawaii (Fig. 5) and the California (Fig. 6) data. The Hawaii data set unwraps well in all four cases. But in the case of the California data, the Goldstein *et al.*[1] algorithm leaves significant gaps in coverage. The unweighted and weighted least-squares algorithms give complete coverage but do not match the overall slope as seen in the Goldstein *et al.* result. Only the synthesis algorithm yields full coverage plus the correct image tilt.

B. Computational Efficiency

In addition to accuracy, computation time is important for all applications requiring moderate to high throughput. In Table 2 we give the time required on our Hewlett–Packard 712 workstation for each of the four scenes, for the Goldstein *et al.*[1] unweighted least-squares, weighted least-squares, and synthesis algorithms. The variation among algorithms is quite large, ranging from 2.7 to 1844.5 s. The synthesis algorithm is by far the least efficient on synthetic data but is comparable to weighted least-squares for real data. It is likely that optimizing the convergence criterion could lead to a drastic reduction in execution time for the weighted least-squares and synthesis approaches.

Table 2. Computational Efficiencies of the Algorithms

	Scene	Execution Time (s) Goldstein *et al.*[1]	Unweighted Least-Squares	Weighted Least-Squares	Synthesis
A.	Geometrical targets	12.9	19.7	19.7	1223.0
B.	Synthetic topography				
	SNR: 1	7.2	18.9	225.1	1337.1
	SNR: 3	5.2	19.0	224.5	1240.0
	SNR: 10	9.0	18.9	224.3	939.7
	SNR: 30	5.9	19.7	224.8	949.3
C.	Hawaii	2.7	18.8	1844.5[a]	306.1
D.	California	3.2	18.8	231.8	211.7

[a] Hawaii scene converged with 10^{-10} error after approximately 2 min. Continuing until 10^{-12} required 30 min.

Table 3. Summary of Algorithm Intercomparisons

Algorithm	Coverage	Accuracy	Efficiency
Goldstein *et al.*	Limited	**Excellent**	**Fast**
Unweighted least-squares	**Complete**	Much distortion if many residues exist	Moderate
Weighted least-squares	**Complete**	Can be better than unweighted	Slow
Synthesis	**Complete**	**Good/Excellent**	Slow

5. CONCLUSIONS

Table 3 summarizes the comparison of the four algorithms that we examined in this paper: the Goldstein *et al.*[1] algorithm, the unweighted least-squares algorithm, the weighted least-squares algorithm, and the synthesis algorithm. The conservative Goldstein *et al.* algorithm is fast and gives accurate results where it can unwrap. Errors inherent in certain singular cases such as the 2π wedge shown above are unwrapped in error by all algorithms. The algorithm is limited to areas of moderate residue density.

The unweighted least-squares algorithm also is reasonably efficient but performs poorly in all but the most benign situations. A significant improvement is afforded by using weighted algorithms, although a large penalty in computational time is incurred. Substituting weighted for unweighted solutions would be a useful trade-off for analysis of a very few interferograms.

A synthesis algorithm that combines the integration path isolation of the residue-cut methods with the better noise-region performance of the least-squares algorithms yields the greatest coverage with least error. However, it suffers from the same computational complexity that affects the weighted least-squares algorithm. Nonetheless it remains the most accurate and complete method available. Algorithm improvements, along with increases in the speed of inexpensive computers, may in time ameliorate this problem.

Phase unwrapping remains a significant issue in radar interferometry. Particularly for cases involving large amounts of data to be processed, the issue will require further study. However, for scientific analysis of a limited number of interferograms, the algorithms presented here may be usefully employed.

Correspondence should be sent to Howard A. Zebker, tel: 650-723-8067; fax: 650-725-7344; e-mail: zebker @jakey.stanford.edu.

REFERENCES

1. R. M. Goldstein, H. A. Zebker, and C. Werner, "Satellite radar interferometry: two-dimensional phase unwrapping," Radio Sci. **23**, 713–720 (1988).
2. 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. Am. A **11**, 107–117 (1994).
3. M. D. Pritt and J. S. Shipman, "Least-squares two-dimensional phase unwrapping using FFT's," IEEE Trans. Geosci. Remote Sens. **32**, 706–708 (1994).
4. G. Fornaro, G. Franceschetti, and R. Lanari, "Interferometric SAR phase unwrapping using Green's formulation," IEEE Trans. Geosci. Remote Sens. **34**, 720–727 (1996).
5. H. Zebker and R. Goldstein, "Topographic mapping from interferometric SAR observations," J. Geophys. Res. **91**, 4993–4999 (1986).
6. H. A. Zebker, T. G. Farr, R. P. Salazar, and T. H. Dixon, "Mapping the world's topography using radar interferometry: the TOPSAT mission," Proc. IEEE **82**, 1774–1786 (1994).
7. A. K. Gabriel, R. M. Goldstein, and H. A. Zebker, "Mapping small elevation changes over large areas: differential radar interferometry, " J. Geophys. Res. **94**, 9183–9191 (1989).
8. D. Massonnet, M. Rossi, C. Carmona, F. Adragna, G.

Peltzer, K. Feigl, and T. Rabaute, "The displacement field of the Landers earthquake mapped by radar interferometry," Nature (London) **364**, 138–142 (1993).
9. H. A. Zebker, P. A. Rosen, R. M. Goldstein, A. Gabriel, and C. Werner, "On the derivation of coseismic displacement fields using differential radar interferometry: the Landers earthquake," J. Geophys. Res. **99**, 19617–19634 (1994).
10. R. M. Goldstein and H. A. Zebker, "Interferometric radar measurement of ocean surface currents," Nature (London) **328**, 707–709 (1987).
11. R. M. Goldstein, H. Engelhardt, B. Kamb, and R. M. Frolich, "Satellite radar interferometry for monitoring ice sheet motion: application to an Antarctic ice stream," Science **262**, 1525–1530 (1993).
12. S. N. Madsen, J. Martin, and H. A. Zebker, "Analysis and evaluation of the NASA/JPL TOPSAR interferometric SAR system," IEEE Trans. Geosci. Remote Sens. **33**, 383–391 (1995).
13. D. Massonnet, K. Feigl, M. Rossi, and F. Adragna, "Radar interferometric mapping of deformation in the year after the Landers earthquake," Nature (London) **369**, 227–230 (1994).
14. D. Massonnet and K. L. Feigl, "Discrimination of geophysical phenomena in satellite radar interferograms," Geophys. Res. Lett. **22**, 1537–1540 (1995).
15. H. A. Zebker, C. L. Werner, P. Rosen, and S. Hensley, "Accuracy of topographic maps derived from ERS-1 radar interferometry," IEEE Trans. Geosci. Remote Sens. **32**, 823–836 (1994).
16. G. Fornaro, G. Franceschetti, R. Lanari, and E. Sansoti, "Robust phase unwrapping techniques: a comparison," J. Opt. Soc. Am. A **13**, 2355–2366 (1996).
17. M. D. Pritt, "Phase unwrapping by means of multigrid techniques for interferometric SAR," IEEE Trans. Geosci. Remote Sens. **34**, 728–738 (1996).
18. D. J. Bone, "Fourier fringe analysis: the two-dimensional phase unwrapping problem," Appl. Opt. **30**, 3627–32 (1991).
19. A. Collaro, G. Franceschetti, F. Palmieri, and M. S. Ferreiro, "Phase unwrapping by means of genetic algorithms," J. Opt. Soc. Am. A **15**, 407–418 (1997).
20. E. Rodriguez and J. Martin, "Theory and design of interferometric SARs," Proc. IEEE **139**, 147–159 (1992).
21. H. A. Zebker and J. Villasenor, "Decorrelation in interferometric radar echoes," IEEE Trans. Geosci. Remote Sens. **30**, 950–959 (1992).
22. F. Li and R. M. Goldstein, "Studies of multi-baseline spaceborne interferometric synthetic aperture radars," IEEE Trans. Geosci. Remote Sens. **28**, 88–97 (1990).
23. B. R. Hunt, "Matrix formulation of the reconstruction of phase values from phase differences," J. Opt. Soc. Am. **69**, 393–399 (1979).