A Markovian Approach for InSAR Phase Reconstruction with Mixed Discrete and Continuous Optimization

A. Shabou, J.Darbon, and F. Tupin

***Abstract*—In this paper, we propose a Markovian approach for InSAR phase reconstruction. Recently, Markovian models based on multichannel InSAR likelihood statistics and Total Variation prior have been proposed to reconstruct the noisy and wrapped phase. Efficient discrete optimization algorithms based on the graph-cut technique are used to efficiently minimize the energy. Our contribution consists in extending these works to cope with continuous label sets providing more precise and accurate reconstructed profiles. The proposed approach provides also a good way to estimate local hyperparameters to adjust the prior model and well preserve discontinuities in profiles. This task is useful when working with real InSAR data where the quantization of the continuous label set leads to a loss of some physical information. The proposed method is compared to other Markovian approaches with discrete multilabel optimization algorithms. Experiments show better quality results both on simulated and real InSAR data.**

***Index Terms*—Multichannel phase unwrapping, MRF, Graph-cut, continuous optimization.**

I. Introduction

Three Dimensional (3D) reconstruction of Earth surface is becoming a task of increasing importance for the last years, thanks to the high resolution of new SAR sensors (TerraSAR-X, ALOS, CSK, RadarSat-2). The 3D reconstruction is mainly performed by estimating the absolute phase from the interferograms generated by interferometric synthetic aperture radar (InSAR) systems. First, the observed phase is known in the principal interval $[-\pi, \pi[$ and has to be unwrapped and denoised. Then, knowing the parameter acquisition of the InSAR system, the Digital Elevation Model (DEM) of the observed surface can be computed.

Unwrapping and denoising the InSAR phase is a difficult inverse problem due to both the multiplicative noise corrupting the observed data and the wrapping operation. More precisely, this problem is known to be ill-posed if the so-called *Itoh* condition is not satisfied [1], i.e., there are jumps of more than π in the estimated phase profile.

Although many methods have been proposed to address this problem, robust phase unwrapping and denoising remains a challenge, especially when dealing with highly noisy interferograms with high discontinuity rate profiles. One of the possible methods to solve this problem is the use of the multichannel Maximum A Posteriori (MAP) estimation method introduced in [2]. In this approach, the multichannel likelihood statistics are exploited to overcome the problem of phase unwrapping, and a Markovian a priori is used to regularize the phase. Then, the problem is resolved by optimizing the MAP criterion using optimization algorithms like the simulated annealing one.

Recently, similar multichannel phase unwrapping (MCPU) approaches have been proposed [3], [4], where new a priori models are introduced with adapted discrete optimization algorithms, like the efficient *graph-cut* based algorithms [5].

These MAP-MCPU approaches work by transforming the original problem of phase reconstruction into a labeling one. To fix notations, let us assume that the estimated phase is defined on a lattice denoted by $\mathcal{V}$. The value of the unwrapped phase x at the site p will be referred to as x_p and is supposed to take its value in a linearly ordered finite set of phase labels $\mathcal{L} = \{l_1, l_2, ..., l_k\}$. This quantization of the phase values is necessary for most graph-cut optimization algorithms. This point will be discussed in this paper. Using the Markovian formalism, the lattice is endowed with a neighborhood system and pairwise interactions are considered, where two sites p and q that are in interaction with each other are denoted by (p, q). The set of all considered pairwise interactions is referred to as $\mathcal{E}$. Then, the labeling problem is solved by minimizing an objective function defined as the following first order Markovian energy

$$E(x) = \sum_{p \in \mathcal{V}} E_p(x_p) + \sum_{(p,q) \in \mathcal{E}} E_{p,q}(x_p - x_q) \ , \qquad (1)$$

where terms E_p encode the multichannel likelihood statistics, while terms $E_{p,q}$ correspond to the prior model.

A popular model to regularize the phase while preserving the discontinuities in the estimated profile is the total variation (TV). Some results of this kind of regularization model on height estimation in case of urban area are presented in [6]. Furthermore, this *a priori* model is well adapted to the graph-cut based optimization algorithms.

The graph-cut optimization approach is a well known technique in computer vision for MRF minimization. A specific graph is constructed, where topology and edge capacities are defined from the energy terms, in such a manner that a s,t-minimum-cut on this graph gives the optimal configuration x (i.e., with the minimal energy value). The s,t-minimum-cut is obtained by computing the maximum-flow on the graph [7].

A. Shabou and F. Tupin are with Institut TELECOM, TELECOM ParisTech, CNRS LTCI, France. E-mail:{aymen.shabou, florence.tupin}@telecom-paristech.fr

J. Darbon is with CMLA, ENS Cachan, CNRS, PRES UniverSud, France. E-mail: darbon@cmla.ens-cachan.fr. Research of J.Darbon has been supported by the grant ONR 000140710810.

A necessary and sufficient condition for minimizing a first order Markovian energy with the graph-cut approach is the submodularity of the energy [8], [9]. For energies of the form of (1), the latter relies on the convexity of $E_{p,q}$ [9], [10].

Our contributions in this work are twofold. First, we improve the discrete optimization step for this class of Markovian models by exploiting the new optimization algorithms introduced in [11] that requires low memory compared to the original ones [9], [10]. Secondly, we increase the accuracy of the reconstructed profile by developping a new approach that improves the efficiency of the optimization step while allowing generating height construction with high precision. This is done thanks to a combination between discrete and continuous based optimization algorithms of the MAP criterion, that also provides a robust method for local prior hyperparameter estimation. The latter leads to a better discontinuity preservation and a reduced loss of contrast due to the TV *a priori*.

The remainder of this paper is as follows. In section II, we briefly review the MAP-MCPU approach. Section III describes the discrete optimization based algorithm used to compute a discrete 3D reconstruction of the surface height. Some drawbacks of this approach are discussed. Then in section IV, a second step is added to the algorithm allowing DEM generation with high precision. Finally some experiments on synthetic and real InSAR data are presented in section V.

II. MAP-MCPU WITH TV PRIOR

With InSAR multichannel systems, independent multifrequencies or multibaselines interferograms related to the same scene are provided [2]. The phase reconstruction problem from these noisy and wrapped phase data can be well solved using the MAP-MCPU approach.

The Markovian energy function proposed in [3] is defined, using the multichannel log-likelihood function for data fidelity term and TV for the prior term, as the following

$$E(x|y,\gamma) = \sum_{p\in\mathcal{V}}\sum_{c=1}^{M} E_p(y_{p,c},\gamma_{p,c}|x_p) + \sum_{(p,q)\in\mathcal{E}} w_{p,q}|x_p - x_q| \tag{2}$$

where x refers to the regularized phase (unwrapped and denoised), $y_{.,c}$ and $\gamma_{.,c}$ respectively encode the observed phase and coherence map of the c^{th} channel, one of the M available channels. Quantities w are some estimated discontinuity map which can be constant or varying spatially, this point will be described in IV-B, and

$$E_p(y_{p,c},\gamma_{p,c}|x_p) = -log\Big(\frac{1-\gamma_{p,c}^2}{2\pi(1-\gamma_{p,c}^2\cos{(y_{p,c}-x_p)^2})}\big(1 + \frac{\gamma_{p,c}\cos{(y_{p,c}-x_p)}Arccos(\gamma_{p,c}\cos{(y_{p,c}-x_p)})}{\sqrt{1-\gamma_{p,c}^2\cos{(y_{p,c}-x_p)^2}}}\big)\Big). \tag{3}$$

In [3], the authors exploit two different graph-cut based discrete minimization algorithms: the binary move based approximate optimization algorithm α-expansion [5] and the multilabel exact optimization algorithm [10]. Nevertheless, these two algorithms present some limits when coping with low coherent interferograms with high discontinuities. To overcom this problem, a more refined quantization of the continuous phase labels is needed to define $\mathcal{L}$. This fact leads to a huge graph that is prohibitive for the exact minimization, while binary partition moves minimization algorithms, like the α-expansion one, reach non good enough local minima [11].

One possible solution to get results with good precision is proposed in [4]. The idea is to perform binary jump partition moves based optimization algorithm with a multiresolution approach, i.e., by suggesting smaller quantified intervals from one iteration to the other. The algorithm is fast thanks to the use of binary moves. However, in case of high discontinuities, non good enough local minima may be reached.

Another solution is given in [11], where new multilabel and large partition moves based optimization algorithms are proposed. In fact, the continuous label set could be quantified with high precision and to overcome the need of large memory for the exact minimization, multilabel moves (i.e., implying a subset of labels from $\mathcal{L}$) are performed. These moves have shown to be able to reach good local minima (in practice, global minima are reached) while keeping memory use low. However, using this optimization method is usually not efficient for time computation. A large time is needed to reach a good solution in term of precision and optimum quality.

We propose in the next section a brief review of this optimization approach as well as its limits.

III. MULTILABEL OPTIMIZATION ALGORITHM

In [11], the proposed optimization algorithm provides good approximate solutions while maintaining a low memory requirement. This is done by performing large and multilabel partition moves (LMPMs) of the estimated solution.

The concept of partition move consists of changing the estimated configuration x iteratively until convergence to a local minimum $\widetilde{x}$. Each change is obtained by minimizing the energy function E_m defined on a set of labels $\mathcal{L}_m$, where m is the size of $\mathcal{L}_m$. A move is called large and multilabel if a large set of sites can change their labels in a large set of labels, i.e., $Card(x) \gg 1$ and $Card(\mathcal{L}_m) \gg 1$. An optimal move is obtained if an exact minimization of the energy function is performed on the considered set of labels. This is possible using graph-cut technique when the prior function $E_{p,q}(\cdot)$ is convex [10]. If the convexity property is satisfied, a specific graph $\mathcal{G}_m$ with specific capacities on arcs is constructed and a maximum-flow is computed on the graph to get the s,t-minimum-cut. This cut gives the configuration with the minimum energy with respect to the considered label set. We show in Fig. 1 an example of such a construction. For more details about the graph topography, we refer the reader to [11]. In Fig. 2, the LMPM algorithm is presented.

Such an optimization technique is very useful for SAR applications. Indeed, in case of high dimensional data with a high range of labels, good approximate minima could be reached with a low memory use. In practice, with a good choice of label set size, a global minimum of the energy is reached. However, there is still a problem of computational time if label sets are of large size. The algorithm has to iterate on all subsets $\mathcal{L}_m$ of the label set $\mathcal{L}$, where at each iteration a maximum-flow is computed on the constructed graph $\mathcal{G}_m$.

In the next section, a new algorithm is proposed to overcome these problems, while providing more accurate reconstruction.

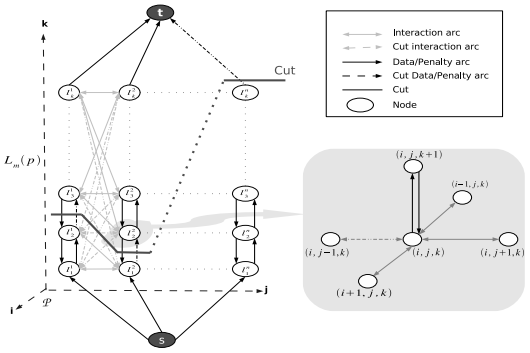


Fig. 1. Graph construction for an optimal multilabel move. On the left, a part of the graph $\mathcal{G}_m$ defined on three pixels. For each pixel, the column of nodes correspond to the possible label values chosen in $\mathcal{L}_m$. A cut is depicted and arcs in the cut are dotted whereas continuous ones are not. A part of the graph is highlighted on the right. Capacities on the edges are defined in [11].

Initialization: Initialize $\widetilde{x}$;
For each cycle c **do**
  success := false;
  For each iteration i **do**
    1. Create the label set of size m: $\mathcal{L}_m^{(i)}$;
    2. Find the optimal LMPM move: $x^{(i)} = \underset{x \in \mathcal{X}}{\operatorname{argmin}}\ E(x)$;
    If $(E(x^i) < E(\widetilde{x}))$ **then**
      $\widetilde{x} := x^{(i)}$;
      success := true ;
    end If
  end For
  If (success == false) **then**
    return $\widetilde{x}$;
  end If
end For

Fig. 2. Optimization algorithm based on the large and multilabel move.

IV. Proposed approach

The proposed approach is based on three steps, mixing discrete and continuous optimization algorithms, with an efficient local hyperparameter estimation method.

A. Discrete optimization

The first step consists in a discrete multilabel optimization of the multilabel energy function (2), where a reduced size of the original label set $\mathcal{L}$, through a quantization of the continuous set $[l_{min}, l_{max}]$, is exploited. The multilabel optimization is performed using the LMPM approach described previously. Since a reduced size of label set is used, an efficient optimization, both in terms of computational time and memory use, is performed. However, the obtained solution, that we note by $\widetilde{x}_1$, could not be good enough in term of precision due to the label set quantization. We note that, for the regularization terms, a global hyperparameter $\beta = w_{p,q}$, $\forall (p,q)$ is used. This scalar could be adjusted manually or automatically through the use of the $L-curve$ method [3].

B. Local hyperparameter estimation

The second step consists in estimating discontinuities map from $\widetilde{x}_1$. Indeed, the latter, as it represents an unwrapped and regularized phase, could bring us a prior knowledge about the true profile discontinuities. Two kinds of discontinuities could be identified. The first one is the set of weak discontinuities related to the under-precision of the reconstructed profile. The second one is the set of sharp discontinuities that are related to the true profile. So, by thresholding the existing discontinuities, we are able to define the local hyperparameters $\{w_{p,q}\}_{(p,q)\in\mathcal{E}}$ to adjust the energy function of (2), i.e.,

$$\begin{cases} w_{p,q} = 0 & \text{if } |\widetilde{x}_{1_p} - \widetilde{x}_{1_q}| > \mu \\ w_{p,q} = 1 & \text{otherwise}\,, \end{cases}$$

where μ is a prescribed threshold. Note that in practice, fixing μ could be performed easily for our case, since $\widetilde{x}_1$ is a regularized image. A low phase threshold on the latter is good enough to preserve true profile discontinuities while smoothing the false ones. Local hyperparameter estimation is a necessary task to overcome the known loss of contrast problem of TV based regularization approaches. Then, by adjusting the prior energy model of (2) using these hyperparameters and starting from $\widetilde{x}_1$, a continuous minimization of the new energy function is performed allowing higher precise reconstruction.

C. Continuous optimization

The third step, which is a continuous based minimization algorithm, is based on the gradient-descent technique. As well known, different continuous optimization algorithms of such objective function are possible, and for a convex minimization, the gradient-descent one is efficient and sufficient. In our case, the first-step optimization algorithm gives usually in practice an exact discrete minimum if the label set size m is well chosen. Thus, only a continuous minimization on a convex set is needed to reach the nearest continuous minimum. This explains the use of the gradient-descent as a second step minimization process. Note also that the continuous optimization may converge quickly since a good initialization is given. Other continuous minimization algorithms could be applied and present more efficient convergence properties, however this is not the aim of this work.

Let us remind the gradient descent based optimization algorithm. We first denote by $x^{(0)}$ the initial guess for the estimated solution, which is in our case the discrete minimum obtained at the end of the first step optimization algorithm. $x^{(i)}$ is the intermediate solution at the i^{th} iteration of the gradient descent algorithm. At each iteration, $x^{(i)}$ is given by

$$x^{(i)} = x^{(i-1)} - \lambda^{(i)} \nabla E(x^{(i-1)})\,, \tag{4}$$

where $\lambda^{(i)}$ designs the step size of the gradient descent. In practice a line search [12] could be applied to select the step size that gives sufficient decrease in the objective functional. This leads to the method of the steepest descent. ∇ is the gradient operator, i.e., derivative respect to x. Because of the differentiation problem of the TV function in (2), an approximate model need to be used instead. A direct approximation of this function could be

$$TV(x) = \sum_{(p,q)} w_{p,q} \sqrt{|x_p - x_q|^2 + \varepsilon}\,, \tag{5}$$

where ε is a precision constant close to 0. In this work, we have rather chosen the l_2 form of TV, given by

$$TV(x) = \sum_{(p,q)(p,r)} \sqrt{w_{p,q}(x_p - x_q)^2 + w_{p,r}(x_p - x_r)^2 + \varepsilon}, \quad (6)$$

where pixels q, r are neighbors to p such that if $p = (k, l)$ then $q = (k + 1, l)$ and $r = (k, l + 1)$. (k, l) encode the coordinate of p in the image grid. This form is more isotropic and appropriate to the gradient descent technique than (5). However, it could not be used in the first step discrete optimization algorithm (LMPM) since it does not satisfy the submodularity condition, needed by the graph-cut technique.

In the next section, some experimental results are performed to highlight the efficiency and accuracy of this algorithm.

V. Experiments and discussion

In this section, several experiments are proposed to show the effectiveness and robustness of the proposed approach, that we call later (MCPU-DC). First, our method has been tested on simulated InSAR data of different profile scenarios such as urban structures and mountain elevations. Quantitative and qualitative evaluations are presented with comparisons to other approaches. Then the method is tested on real InSAR data.

A. Simulated InSAR data

In the first experiment, we consider a synthetic height profile Fig. 3(a), of size (200×400), with a maximum height of 14 rad exhibiting both smooth and discontinuous areas. We used four frequencies ($\{5, 6.33, 7.66, 9\}$ GHz) to generate four noisy independent interferograms with a constant coherence of $\{\gamma_{p,c} = 0.6; \forall p, c\}$. In Fig. 3(b), we show one noisy interferogram. Note that the profile is ambiguous for these frequencies. In fact, there are phase jumps greater than π violating the *Itoh* condition.

In the second experiment, we consider a more representative height profile of natural image Fig. 3(c) with the same sizes as the previous one, and a maximum height of 19.1 rad exhibiting area characterizing highland scenes. The same InSAR parameters as the ones used in the previous experiment are applied on this profile to generate independent interferograms.

In figures 4 and 5, we show the reconstruction results obtained respectively on the two scenarios using the proposed approach. Results given by the multichannel approach (MCPU-GC) in [3] are also presented in figures 4(b) and 5(b), where the continuous label set is quantified into 200 labels and an exact minimization algorithm is applied based on the Ishikawa graph construction. We note that a huge graph has to be build to provide an accurate solution.

Reconstructions obtained at the end of the first-step optimization approach are presented in figures 4(c) and 5(c), where the continuous label set is quantified into 64 labels for both images and the LMPM optimization algorithm is applied using label sets of sizes respectively $m = 8$ and $m = 16$ to converge to a good local minimum of the energy.

Starting from these results, local hyperparameters are estimated by thresholding the existing discontinuities using the described method in IV-B. Note that the threshold parameter μ is fixed to $\frac{\pi}{8}$ in these experiments. Then the continuous optimization algorithm is performed. Results on both data sets are presented in 4(d) 5(d). We show that smooth areas are well reconstructed while discontinuities are also preserved. This is not the case in figures 4(b) and 5(b), where discontinuities are well preserved thanks to the TV prior but smooth areas are noisy, even when a large label set is considered.

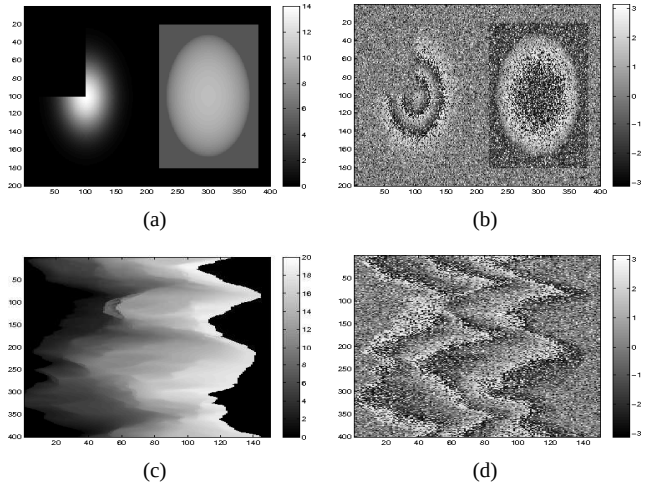


Fig. 3. Simulated data sets. The 2D view of the first (a) original profile, (b) and one generated interferogram, and the second (c) original profile, (b) and one generated interferogram.

TABLE I
Quantitative results of the two experiments.

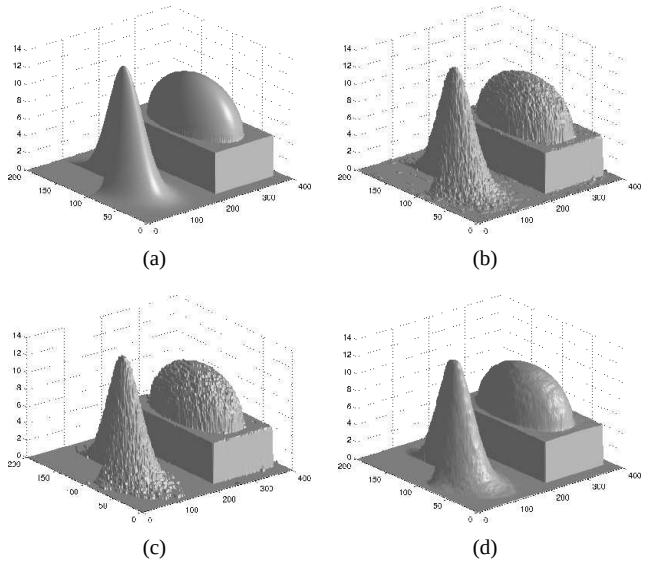
Data	Approach	MSE	RMSE	Memory	Time(mn:sec)
3 (a)	**MCPU-GC**	0.0474	0.2176	16×10^6	12:09
	MCPU-DC	0.0071	0.0840	64×10^4	03:25
3 (c)	**MCPU-GC**	0.0343	0.1852	12×10^6	18:14
	MCPU-DC	0.0325	0.1802	96×10^4	04:27

Quantitative measures, in term of reconstruction quality using the mean and root mean square deviation errors (MSE, RMSE) and computational time and memory needed for graph allocation, are presented in table I. These results show that both the accuracy and the efficiency of the two-step optimization algorithm and its effectiveness compared to other approaches. We also note that improvements in term of MSE and RMSE are less important for the second profile compared to the first one, since less smooth regions are characterizing this profile. However, we can clearly see that in term of computational time and qualitative visual results, our approach is always more interesting compared to the MCPU-GC one.

B. Real InSAR data

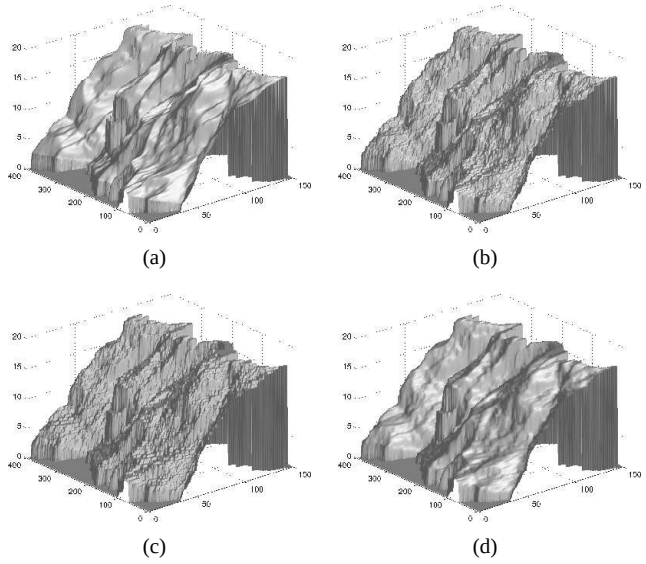
We have tested the new algorithm on a real data set of an urban scenario. A set of 8 L-Band E-SAR interferograms (2 interferograms for each of the four polarizations) are acquired on the city of Dresden. The smallest orthogonal baseline is of about $8.42m$ and the biggest is of about $28.34m$. Due to the presence of noise, the Itoh condition is violated in some areas.

We can see in figures 6(c) and 6(d) the good quality of the reconstructed phase, compared to the result obtained using the MCPU-GC 6(b). In fact, it is clearly seen that discontinuities



(a) (b) (c) (d)

Fig. 4. Results on the simulated data 3-(a). 3D view of (a) the original profile, and the reconstructed phases obtained respectively (b) using the MCPU-GC approach, (c) at the end of the first step optimization algorithm ($\widetilde{x}_1$), and (d) at the end of the continuous minimization ($\widetilde{x}_2$).



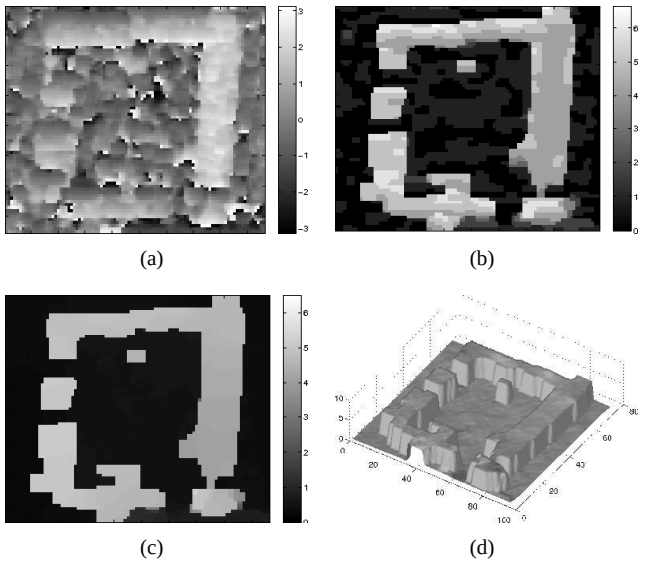
(a) (b) (c) (d)

Fig. 5. Results on the simulated area 3-(c). 3D views of (a) the original profile, and the reconstructed phases obtained respectively (b) using the MCPU-GC approach, (c) at the end of the first step optimization algorithm (LMPM) ($\widetilde{x}_1$), (d) at the end of the continuous minimization ($\widetilde{x}_2$).

and smooth or homogeneous areas are well preserved using the mixed discrete and continuous optimization approach, while using only a discrete optimization algorithm provides results with low precision. We have to note that only topographic fringes are present in the processed data. Thus, to deal with differential interferograms, other problems need to be adressed, such as the atmospheric phase contributions. Improvements in this direction will be proposed in a future work.

VI. Conclusion

In this work, we have developed a new 3D height reconstruction methodology based on a multichannel phase unwrapping approach and efficient and accurate optimization



(a) (b) (c) (d)

Fig. 6. 3D reconstruction of real InSAR data. (a) The first noisy interferogram, (b) the 2D view of the reconstructed profile with the MCPU-GC approach, (c) the 2D view of the reconstructed profile with the proposed approach, (d) the 3D view of the reconstruction.

algorithm. The proposed algorithm overcomes the limits that characterize other MAP-MCPU approaches both in term of computational complexity and solution accuracy. We have tested this approach on simulated and real InSAR data. Good quantitative and qualitative results are obtained. Future work will focus on the exploitation of these approaches for the differential InSAR applications (D-InSAR).

References

[1] K. Itoh, "Analysis of the phase unwrapping problem," *Journal of Applied Optics*, vol. 21, no. 14, 1982.

[2] G. Ferraiuolo, V. Pascazio, and G. Schirinzi, "Maximum a posteriori estimation of height profiles in InSAR imaging," *IEEE Geoscience and Remote Sensing Letters*, vol. 1, pp. 66–70, 2004.

[3] G. Ferraioli, A. Shabou, F. Tupin, and V. Pascazio, "Multichannel phase unwrapping with graph-cuts," *IEEE Geoscience and Remote Sensing Letters*, vol. 6, no. 3, pp. 562–566, 2009.

[4] G. Valado and J. Bioucas-Dias, "Phase imaging: Unwrapping and denoising with diversity and multi-resolution," in *Proceedings of 2008 International Workshop on Local and Non-Local Approximation in Image Processing (LNLA'08), Lausanne, Switzerland*, 2008.

[5] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1222–1239, 2001.

[6] G. Ferraioli, A. Shabou, F. Tupin, and V. Pascazio, "Fast InSAR multichannel phase unwrapping for DEM generation," in *International Urban Remote Sensing Conference (URBAN'09), China*, 2009.

[7] R. Ahuja, T. Magnanti, and J. Orlin, *Network Flows: theory, algorithms and applications*. Prentice Hall, 1993.

[8] V. Kolmogorov and R. Zabih, "What energy functions can be minimized via graph cuts?" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 2, pp. 147–159, 2004.

[9] J. Darbon, "Global optimization for first order Markov random fields with submodular priors," *Journal of Discrete Applied Mathematics*, vol. 157, no. 16, pp. 3412–3423, 2009.

[10] H. Ishikawa, "Exact optimization for Markov Random Fields with convex priors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 10, pp. 1333–1336, 2003.

[11] A. Shabou, F. Tupin, and J. Darbon, "A graph-cut based algorithm for approximate MRF optimization," in *International Conference on Image Processing (ICIP'09), Cairo, Egypt*, 2009, pp. 2413–2416.

[12] R. Fletcher, *Practical Methods of Optimization*. Wiley, 2000.