Urban mapping using coarse SAR and optical data: outcome of the 2007 GRSS data fusion contest

Fabio Pacifici, Fabio Del Frate, William Emery, Paolo Gamba, Jocelyn Chanussot

To cite this version:


HAL Id: hal-00348851
https://hal.archives-ouvertes.fr/hal-00348851
Submitted on 22 Dec 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Urban Mapping Using Coarse SAR and Optical Data: Outcome of the 2007 GRSS Data Fusion Contest

Fabio Pacifici, Student Member, IEEE, Fabio Del Frate, Member, IEEE, William J. Emery, Fellow, IEEE, Paolo Gamba, Senior Member, IEEE, and Jocelyn Chanussot, Senior Member, IEEE

Abstract—The 2007 Data Fusion Contest that was organized by the IEEE Geoscience and Remote Sensing Data Fusion Technical Committee was dealing with the extraction of a land use/land cover maps in and around an urban area, exploiting multitemporal and multisource coarse-resolution data sets. In particular, synthetic aperture radar and optical data from satellite sensors were considered. Excellent indicators for mapping accuracy were obtained by the top teams. The best algorithm is based on a neural classification enhanced by preprocessing and postprocessing steps.

Index Terms—Data fusion, European Remote Sensing satellite (ERS), image classification, Landsat, neural networks (NNs), urban remote sensing.

I. INTRODUCTION

In the past two decades, monitoring urban centers and their peripheries at a regional scale has become an increasingly relevant topic for public institutions to keep track of the loss of agricultural land and natural vegetation due to urban development. In recent years, significant attention has focused on multisensor data fusion for remote sensing applications and, more specifically, for land cover mapping. Data fusion techniques combine information from multiple sources, providing potential advantages over a single sensor in terms of classification accuracy [1].

The characteristics of data acquired by optical and synthetic aperture radar (SAR) sensors greatly differ. Multispectral satellites such as Landsat provide information on the energy scattered and radiated by the Earth’s surface in different wavelengths, from the visible to the thermal infrared, providing the ability to discriminate between different land cover classes such as vegetated areas, water surfaces, and urban centers. SAR sensors such as the European Remote Sensing satellites (ERS) 1/2 provide measurements in amplitude and phase related to the interaction of the Earth’s surface with microwaves. These acquisitions (C-band) are characterized by high returns from buildings and wide streets and the presence of trees. This means that SAR sensors provide information that may not be obtained from optical sensors, and therefore, data fusion potentially provides improved results in the classification process compared to the conventional single-source classification results [2].

Data fusion may be accomplished at different information levels such as signal, pixel, feature, or decision. A signal-based fusion combines data from different sensors, creating a new input signal with improved characteristics over the original (e.g., a better signal-to-noise ratio). Information from different images can be merged in the pixel-based fusion to improve the performance of the processing tasks. A feature-based fusion combines features extracted from different signals or images, whereas a decision-level fusion consists of merging very dissimilar data at a higher level of abstraction [3].

In the data fusion literature, many alternative methods have been proposed for combining multisensor decisions by weighting the influence of each sensor. A common approach to multisource classification is to concatenate the data in a stacked vector and treat it as a unique set of measurements [4], but statistical classifiers can be difficult to deal with because it is not always possible to formulate reasonable assumptions about the distribution of features. Contextual information from neighboring pixels improves the accuracy of a pixel-based classification. For instance, the reliability of each information source can be estimated for each pixel using spatial features and can be integrated in a fuzzy-logic-based fusion scheme [5]. Markov random fields (MRFs) also provide a powerful methodological framework for modeling spatial and temporal context, allowing the images from different sensors and map data to be merged in a consistent way [6], [7]. Nonparametric approaches such as neural networks (NNs) [8], [9] or support vector machines [10] can be exploited because they do not require any specific probabilistic assumptions for class distribution. Hybrid approaches combining parametric methods and NNs have been proposed by Benediktsson et al. [11] by first

Manuscript received August 23, 2007; revised November 19, 2007.
F. Pacifici and F. Del Frate are with the Department of Computer Science, Systems, and Production, Tor Vergata University, 00133 Rome, Italy (e-mail: f.pacifici@disp.uniroma2.it).
W. J. Emery is with the Department of Aerospace Engineering Science, University of Colorado at Boulder, Boulder, CO 80309 USA.
P. Gamba is with the Dipartimento di Elettronica, Università di Pavia, 27100 Pavia, Italy.
J. Chanussot is with the GIPSA-Lab, Grenoble Institute of Technology, 38000 Grenoble, France.
Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.
Digital Object Identifier 10.1109/LGRS.2008.915939
separately treating each data source using statistical methods and then using an NN to obtain the consensus decision.

This letter is organized as follows. The 2007 Data Fusion Contest is presented in Section II, whereas Section III introduces the methodology of the winning team, discussing how it was designed. Finally, in Section IV, the classification results are presented and discussed.

II. 2007 DATA FUSION CONTEST

The Data Fusion Contest has been organized by the Data Fusion Technical Committee (DFTC) of the Geoscience and Remote Sensing Society (GRSS) of the IEEE and has been annually proposed since 2006. It is a contest open not only to DFTC members, but to everyone. The aim of the Data Fusion Contest is to evaluate existing methodologies at the research or operational level to solve remote sensing problems using data from different sensors. The main aim of this contest is to provide a benchmark to the researchers interested in a class of data fusion problems, starting with a contest and then allowing the data and results to be used as reference for the wider community, inside and outside the DFTC. The first issue of the contest was devoted to pansharpening [12]. In 2007, the contest was related to urban mapping using radar and optical data. A set of satellite radar and optical images (ERS amplitude data and Landsat multispectral data) was made available, and the task was to obtain a classified map as accurate as possible, relative to ground reference data. The reference data depicted land cover and land use classes for the urban area of interest. The final results led to an award for the methodology proposed in this letter, and the award was announced at the 2007 International Geoscience and Remote Sensing Symposium (IGARSS07) held in July 2007 in Barcelona, Spain.

The contest attracted considerable attention in the remote sensing research community. More than 70 individuals registered to download the data sets and try their own approach. In the end, nine different teams uploaded more than 100 classification maps, most of them continuously refining their algorithm performance. The best result, awarded with a Certificate of Recognition during the Technical Committees’ Dinner at IGARSS07, is presented in Section III.

III. WINNING ALGORITHM

The winning algorithm is based on an NN approach. The classification procedure can be divided into three steps: 1) preprocessing; 2) NN classification; and 3) postprocessing.

A. Preprocessing

Different classification schemes may be developed to achieve the desired classification accuracy. The first step to be considered regards the choice of inputs, which will successively feed the NN. The simplest scheme, shown in Fig. 1(a), feeds an NN with 21 inputs (9 SAR and 6 + 6 optical), which requires a long training time. Therefore, the reduction of the input dimensionality is generally desirable. The principal component analysis (PCA) was applied to decrease the number of inputs used to train the NN. The PCA maps image data into a new uncorrelated coordinate system in which the data have greatest variance.

The data sets and the resulting evaluations were all provided using a web portal specifically designed for the 2007 Data Fusion Contest. The portal allowed anyone to register, read the contest rules, understand the task for the contest, download the data, and upload the results. It also provided immediate feedback on the quality of the uploaded map by computing a few quantitative indices and correspondingly upgrading the top ten ranking list. In particular, because a classification map was the final output of the algorithms proposed by the contest’s participants, this map was evaluated against ground reference data (unknown to the contestants), and the kappa index was used as a ranking value. The top three classification maps of the final list provided a kappa index larger than 0.93, an extremely good result, showing the excellence of the teams and the methodologies prompted by the contest task.

The contest attracted considerable attention in the remote sensing research community. More than 70 individuals registered to download the data sets and try their own approach. In the end, nine different teams uploaded more than 100 classification maps, most of them continuously refining their algorithm performance. The best result, awarded with a Certificate of Recognition during the Technical Committees’ Dinner at IGARSS07, is presented in Section III.
function used is a weighted linear combination of inputs such as neuron consists of three main parts: 1) the synaptic weight vector generally characterized by several inputs and one output. Each ... and used only to have a rough estimation on the performances of the networks during the designing process.

Once the input and the output of the network are established, the critical step is to find the optimal number of units to be considered in the hidden layers. In fact, if the number of these neurons is too small, the input–output associative capabilities of the network are too weak. At the same time, this number should not be too large, otherwise, these capabilities might show a lack of generality being too narrowly tailored to the training set, and the computational complexity of the algorithm would be increased in vain [21]. Therefore, a compromise has to be found to select the most suitable number of hidden neurons for the optimal NN topology.

Two different approaches can be used to find the best architecture:

1) the growing method in which the starting network is small and the neurons are subsequently added until the optimization criteria is reached;
2) the pruning method in which the starting network is relatively large and the neurons are subsequently removed until the optimization criterion is satisfied.

The latter approach is used here. Therefore, after a reasonable evaluation in terms of classification accuracy, the chosen topology was 14-200-200-5. This estimation involved the analysis of the output variance of different topologies characterized by an increasing number of hidden neurons (by a factor of 40) starting from 14-40-40-5. In general, we expect that increasing the number of hidden neurons is effective up to a given number, after that, the overall error value does not significantly change.

3) Learning Algorithm: During the training phase, the network learns how to approximate an unknown input-output relation by adjusting the weight connections. This is done by minimizing an error function that quantifies the difference between the actual outputs and the desired outputs. There are several learning algorithms designed to minimize this error function. The learning algorithm used here is the scaling conjugate gradient method [22]. This is a member of the class of conjugate gradient methods, general-purpose second-order techniques that help minimize the goal functions of several variables. Second order indicates that such methods use the second derivatives of the error function, whereas a first-order technique, like standard backpropagation, only uses the first derivatives. To decide on when the training has to be interrupted, the early-stopping criterion based on the validation set was used [16].

A central problem in pattern recognition consists of minimizing the system complexity. In NNs, this issue often consists of minimizing the number of connection weights [23]. Pruning algorithms try to make NNs smaller by pruning unnecessary links or units. The advantages are the following:

1) The cost of an NN can be reduced (runtime, memory, and cost for hardware implementation).
2) The generalization of the NN can be improved.
3) Unnecessary input units can be pruned to give evidence of the relevance of input values.

To decide which links or units are less important and, thus, are candidates for removal, it is necessary to assess the relative importance of weight, namely its saliency. There are different approaches to determine the saliency of an element. The magnitude-based pruning is the simplest weight pruning algorithm [24], which is based on deleting links with small saliency. After the standard training, the link having the smallest magnitude value is removed (thus, the saliency of a link is just the absolute size of its weight). The network is then retrained, and the process is iteratively repeated until the training error reaches a specified limit. Although this method is very simple, it rarely yields worse results than the more sophisticated algorithms such as optimal brain damage, optimal brain surgeon, or skeletonization [25], [26].

Based on the validation set, the pruning process tried to remove neurons and connections following two distinct goals:

1) to reach the minimum of the classification error (on the validation set). The network is then retrained and the pruning is iteratively repeated until the classification error does not increase;
2) to reach the minimum number of connections taking into account a maximum decrease of the classification accuracy of about 10% with respect to the fully connected network. Note that this task is only academic and does not address the contest.

After 4227 epochs of training, the fully connected NN (43 800 connections) correctly classified 86.4% of the validation patterns, as shown in Table III, which lists the accuracy details of the different network topologies.

The number of retraining epochs of the pruning phase has been set to 50. This means that the network was retrained for 50 cycles after the pruning of each connection. The fully connected NN has been pruned, reaching the minimum classification error with 43 657 connections (overall accuracy of 89.7%). Therefore, less than 0.5% of the initial connections were removed, for an improvement, in terms of classification accuracy, of 3.3%. However, no neuron was removed by the procedure. Successively, the pruning continued to further reduce the number of connections, expecting a decrease in the classification accuracy. As shown in Table III, 57 connections were sufficient to decrease the classification accuracy from 89.7% to 84.3%. The decrease in the classification accuracy of about 10% with respect to the fully connected network was reached with 42 470 connections (3.0% of the initial connections). These accuracies were based on the validation set obtained by the visual inspection of the scene and were only used to have a rough estimation on the performances of the different topologies during the designing process.

C. Postprocessing

Classified images often suffer from a lack of spatial coherence, which results in speckle or holes in homogeneous areas. This noise phenomenon appears as many isolated pixels or small groups of pixels whose classifications are different from those of their neighbors [27]. Postclassification processing techniques can be applied to further increase the classification accuracy. This is often achieved by analyzing the neighborhood for each pixel and removing the isolated pixels (sieve process) and then merging the small groups of pixels together to make more continuous and coherent units (clump process) [28]. These sieve and clump procedures were used here to reduce the effect of isolated pixels removing all regions smaller than the designated cluster dimension.

The trial-and-error strategy was used to define the optimal size with respect to the validation set, starting with a small cluster dimension (10 pixels). The highest classification accuracy was reached using a dimension of 142 pixels. Therefore, even though the smaller clusters may have been correctly classified, they were considered unreliable.

The classification map obtained after the postclassification processing reached an accuracy of 96.98% in terms of k-coeff, relative to the validation set obtained by visual inspection.

---

**TABLE III**

<table>
<thead>
<tr>
<th>Net ID</th>
<th>Number of Connections</th>
<th>Overall Accuracy (%)</th>
<th>k-coeff</th>
<th>Epochs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full</td>
<td>43800</td>
<td>86.4</td>
<td>0.816</td>
<td>4227</td>
</tr>
<tr>
<td>net1</td>
<td>43684</td>
<td>86.9</td>
<td>0.824</td>
<td>10027</td>
</tr>
<tr>
<td>net2</td>
<td>43657</td>
<td>89.7</td>
<td>0.861</td>
<td>11377</td>
</tr>
<tr>
<td>net3</td>
<td>43600</td>
<td>84.3</td>
<td>0.787</td>
<td>14227</td>
</tr>
<tr>
<td>net4</td>
<td>42470</td>
<td>75.4</td>
<td>0.667</td>
<td>70727</td>
</tr>
</tbody>
</table>
and the ground reference data are shown. The color code is in Table II.

(b) optical (bands RGB-431) sensors. In (c) and (d), the final classification map
Fig. 3. City of Pavia imaged by (a) SAR (backscattering amplitude) and
PACIFICI
in Fig. 3, reached an accuracy of 0.9393 in terms of
suitable network topology. The final classification map, shown
highlighting the careful analysis that is required to find the most
ERS 1/2 and Landsat data set has been presented in Section III,
be assumed [21].
require that a particular form of a probability density function
input–output discriminant relations from the data and do not

Vegetated areas, sparse buildings, and water surfaces have shown higher classification accuracies (stemming from a higher class separability) with respect to the other two classes, which is an expected result because the responses of the areas characterized by high or moderate density of buildings (such as city center and residential, respectively) are quite similar in both optical and SAR sensors.

ACKNOWLEDGMENT

The authors would like to thank G. Pulina for designing and managing the 2007 contest website and implementing the automatic evaluation software.

REFERENCES