ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/279178350

Recent Methods for Reconstructing Missing Data in Multispectral Satellite Imagery

Chapter · September 2016

DOI: 10.1007/978-4-431-55342-7_19

CITATIONS
2

READS
249

4 authors:

Farid Melgani
Università degli Studi di Trento
212 PUBLICATIONS **8,419** CITATIONS
SEE PROFILE

G. Mercier
eXo maKina
228 PUBLICATIONS **3,489** CITATIONS
SEE PROFILE

Luca Lorenzi
sws
12 PUBLICATIONS **198** CITATIONS
SEE PROFILE

Edoardo Pasolli
University of Naples Federico II
71 PUBLICATIONS **2,845** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

MODENA View project

Landscape Archaeology in Northern Africa View project

All content following this page was uploaded by Edoardo Pasolli on 19 September 2016.

The user has requested enhancement of the downloaded file.

Recent Methods for Reconstructing Missing Data in Multispectral Satellite Imagery

Farid Melgani, Grégoire Mercier, Luca Lorenzi and Edoardo Pasolli

Abstract One of the major limitations of passive sensors is their high sensitivity to weather conditions during the image acquisition process. The resulting images are frequently subject to the presence of clouds, which makes the image partly useless for assessing landscape properties. The common approach to cope with this problem attempts to remove the clouds by substituting them with cloud-free estimations. The cloud removal problem can be viewed as an image reconstruction/restoration issue, in which it is aimed at recovering an original scene from degraded or missing observations. Two cloud removal approaches are detailed and discussed in this chapter. The first one is a single-channel method for the reconstruction in a sequence of temporal optical images. Given a contaminated image of the sequence, each area of missing measurements is recovered by means of a contextual prediction process that reproduces the local spectro-temporal relationships. The second approach exploits the Compressive Sensing (CS) theory, which offers the capability to recover an unknown sparse signal with a linear combination of a small number of elementary samples. The two reconstruction approaches are evaluated experimentally on a real multitemporal multispectral remote sensing image.

Keywords Cloud removal · Compressive sensing · Genetic algorithms · Image reconstruction · Optical imagery · Satellite image time series · Sparse representation

F. Melgani (✉) · L. Lorenzi
Department of Information Engineering and Computer Science,
University of Trento, 38123 Trento, Italy
e-mail: melgani@disi.unitn.it

G. Mercier
Télécom Bretagne, Technopôle Brest-Iroise, 29238 Brest, France
e-mail: gregoire.mercier@telecom-bretagne.eu

E. Pasolli
Centre for Integrative Biology, University of Trento, 38123 Trento, Italy
e-mail: edoardo.pasolli@gmail.com

© Springer Japan 2016

R.S. Anderssen et al. (eds.), *Applications + Practical Conceptualization + Mathematics = fruitful Innovation*, Mathematics for Industry 11,
DOI 10.1007/978-4-431-55342-7_19

1 Introduction

One of the major limitations of passive sensors is their high sensitivity to weather conditions during the image acquisition process. The resulting images are frequently subject to the presence of clouds, whose extent depends on the season and the geographic position of the study region. For instance, in Canada, from 50 to 80 % of the Earth's surface can be obscured by clouds in mid-morning [1]. Depending on the application and the end-user requirements, clouds can be viewed: as a source of information for measuring important parameters such as cloud liquid water useful in meteorological forecasting and hydrological studies [2]; or as a source of contamination that makes the image partly useless for assessing landscape properties. In the latter case, which represents the subject of this paper, clouds distort the spectral response of land covers, thereby resulting in missing data for high-frequency passive sensors including multispectral optical sensors and microwave radiometers. The common approach to cope with this problem first detects the contaminated regions and, in a second instance, attempts to remove the clouds by substituting them with cloud-free estimations.

The cloud removal problem, which is the focus of this contribution, can be viewed as an image reconstruction/restoration issue, in which it is aimed at recovering an original scene from degraded or missing observations. Image reconstruction/restoration has been intensively and extensively studied in various application fields, such as radio astronomy, biomedical engineering, and machine vision, because of its practical importance as well as theoretical interest [3]. In the remote sensing field, significant attention has been devoted to the reconstruction/restoration of images subject to various problems, such as acquisition blur and geometric distortions, phase distortions, resampling problems, or problems related to applications like buried object detection. By contrast, less attention has been paid to the specific problem of cloud removal. Among the relatively few works available in the literature, one can find for instance [4, 5]. The main drawbacks that can be identified from these cloud removal algorithms are: sensor-dependence, very high temporal resolution, ground cover type-dependence, cloud-type dependence, or/and high methodological complexity. Two alternative recent approaches, based on linear contextual prediction and compressive sensing [6, 7], respectively, that aim at circumventing most of these drawbacks, are detailed and discussed in this paper.

2 Problem Formulation

We consider a set of multitemporal multispectral images $I^{(i)}$ acquired over the same geographical area by an optical sensor at times t_i (with $i \in S = \{1, 2, \ldots, T\}$). Let us suppose that the images have been registered. We assume that (1) the images of the sequence may convey changes in the spectral appearance of objects on the ground and (2) they are characterized by an almost similar spatial structure. The last assumption

can be considered realistic if the acquisition dates are close to each other (i.e., high-temporal resolution) or if the spatial dynamics of the geographical area under analysis is slow compared to the total time interval of the sequence (e.g., forest, mountainous, and urban areas). Moreover, we assume that the images have first been processed to generate a sequence of cloud/noncloud classification maps $M^{(i)}$ $(i \in S)$ by using an automatic cloud-masking method or simply by photointerpretation. Given $M^{(i)}$, cloudy and non cloudy areas are represented by $\Omega^{(i)}$ and $\Phi^{(i)}$, respectively, subject to $I^{(i)} = \Omega^{(i)} \cup \Phi^{(i)}$. The specific problem of the detection of clouds (and their shadow) is not dealt with in this paper. The objective of the investigated methods is to reconstruct any area contaminated by clouds (or by cloud shadows) for each image of the sequence. Therefore each classification map $M^{(i)}$ will be used to guide the cloud removal process.

For the first investigated approach based on linear contextual prediction, image channels are processed separately. Let us denote by $X^{(i)}$ $(i \in S)$ one of the available T single-channel temporal images. We denote by C a cloudy area of the image $X^{(i)}$. The reconstruction problem of C in $X^{(i)}$ can be expressed as a problem of generating an image $Y^{(i)}$ such that:

$$Y^{(i)}(u, v) = \begin{cases} X^{(i)}(u, v) \text{ if } (u, v) \notin C \\ f\left[X^{(k)}(u, v),\ k \in S_c\right],\ \text{otherwise} \end{cases} \tag{1}$$

where $f[\cdot]$ represents a contextual prediction function and S_c stands for the subset of indices corresponding to images $X^{(k)}$ $(k \neq i)$ that are cloud-free in the spatial area $N(C)$ including C and its neighborhood. In other words, any image of the sequence obscured by the presence of a cloud in $N(C)$ will not contribute to the reconstruction process of C. In each image $X^{(k)}$ $(k \in S_c)$, the spatial area $N(C)$ can be subdivided into two cloud-free areas C and $\overline{C}$ such that $N(C) = C \cup \overline{C}$ and $C \cap \overline{C} = \emptyset$ where C represents the spatial area that corresponds to the cloudy area in $X^{(i)}$ and $\overline{C}$ stands for the neighboring spatial area.

For the second approach based on compressive sensing, we simplify the problem by supposing to have just two images ($T = 2$) and we relax the above channel-based processing constraint (all channels are processed together). At this level, we make the hypothesis that the image $I^{(2)}$ has clouds, while the image $I^{(1)}$ is cloud-free. We assume that any pixel $\mathbf{x}^{(1)} \in \Omega^{(1)}$ can be expressed as a linear combination of pixels in region $\Omega^{(1)}$ of $I^{(1)}$. In other words, in $I^{(1)}$ we have:

$$\mathbf{x}^{(1)} = \Phi^{(1)} \cdot \boldsymbol{\alpha},\ \forall \mathbf{x}^{(1)} \in \Omega^{(1)} \tag{2}$$

where $\boldsymbol{\alpha}$ is an unknown weight vector associated with the considered pixel $\mathbf{x}^{(1)}$ and having the same dimension as the number of pixels belonging to $\Phi^{(1)}$. The problem at this point is to infer $\boldsymbol{\alpha} = f\left(\Phi^{(1)}, \mathbf{x}^{(1)}\right)$. Once $\boldsymbol{\alpha}$ is computed, if we assume that $I^{(1)}$ and $I^{(2)}$ are temporally close, it will be possible to reuse the $\boldsymbol{\alpha}$ coefficients to reconstruct the spatially corresponding pixel in the missing area $\Omega^{(2)}$, adopting the previous formulation for $I^{(2)}$, i.e., $\hat{\mathbf{x}}^{(2)} = \Omega^{(2)} \cdot \boldsymbol{\alpha}$. In other words, for each pixel

$\mathbf{x}^{(1)} \in \Omega^{(1)}$, we evaluate $\boldsymbol{\alpha}$, and in a second moment we reuse this weight vector to return an estimation of $\mathbf{x}^{(2)} \in \Omega^{(2)}$:

$$\begin{aligned} \text{From } I^{(1)} : \boldsymbol{\alpha} &= f\left(\Phi^{(1)}, \mathbf{x}^{(1)}\right) \\ \text{To } I^{(2)} : \hat{\mathbf{x}}^{(2)} &= \Omega^{(2)} \cdot \boldsymbol{\alpha} \end{aligned} \tag{3}$$

where $f(\cdot)$ represents an estimation function. We recall that, differently from the first approach, all image channels are processed simultaneously.

3 Linear Contextual Prediction Approach

The first approach is a single-channel method for the reconstruction of cloud-contaminated areas in a sequence of temporal optical images. It is based on the assumptions that spectral non-stationarity is allowed while the spatial structure of the image should be identical over the image sequence. Given a contaminated image of the sequence, each area of missing measurements is recovered by means of a contextual prediction process that reproduces the local spectro-temporal relationships. These latter are deduced from the cloud-free areas in the spatial neighborhood of the contaminated area over the available series of temporal images. The contextual prediction process is implemented in two steps. First, a prediction system is trained to learn over $\overline{C}$ the temporal relationship between the set of available images $X^{(k)}$ ($k \in S_c$) that are cloud-free in $N(C)$ on the one hand and the image $X^{(i)}$ on the other. This is done by means of an ensemble of linear predictors, each trained in an unsupervised way over a local temporal region that is spectrally homogeneous in each temporal image of the sequence. In order to obtain such regions, each temporal image is locally classified in an unsupervised way by means of the Expectation-Maximization (EM) algorithm [8] assuming the data classes are Gaussian. The number of data classes is estimated automatically by minimizing the Minimum Descriptive Length (MDL) criterion [9]. Once the training is completed, the prediction system is used to provide an estimate of each contaminated pixel of image $X^{(i)}$ in C, based on the spatially corresponding pixel values in images $X^{(k)}$ ($k \in S_c$).

3.1 Contextual Prediction Process

The complexity of the relationship between images $X^{(k)}$ ($k \in S_c$) and image $X^{(i)}$ in $N(C)$ will depend mainly on the complexity of their statistical distribution, which is conditioned by the quantity and quality of ground-cover classes in $N(C)$ at each date t_k ($k \in S_c \cup \{i\}$). In multispectral imagery, the assumption that the distribution of images can be approximated as a mixture of normally distributed samples is generally well-accepted. Accordingly, the probability distribution function (pdf) of each image $X^{(k)}$ ($k \in S_c$) in $N(C)$ can be written as:

$$p_k(x) = \sum_{m=1}^{M_k} P(\omega_m^k) \cdot p(x \mid \omega_m^k) \tag{4}$$

where $P(\omega_m^k) = P_m^k$ and $p(x|\omega_m^k) = N(\mu_m^k, \sigma_m^k)$ are the prior probability and the conditional pdf associated with the mth gaussian mode in the $N(C)$ region of the kth image, respectively. Constant M_k stands for the number of modes characterizing the related pdf $p_k(x)$, while μ_m^k and σ_m^k are mean and standard deviation parameters, respectively.

Given a multitemporal pixel vector $\boldsymbol{x} = [x_1, x_2, \ldots, x_K]$ (K is the cardinality of S_c), such that x_j represents the pixel value in the jth image of the temporal sequence $X^{(k)}[k \in S_c; k = p(j)$, where $p(\cdot)$ is a mapping of the integers $\{1, 2, \ldots, K\}$ into $S_c]$ and $x_j \in \omega_{n_j}^j$ ($n_j \in \{1, 2, \ldots, M_j\}$), the contextual prediction function $f[\cdot]$ can be expressed as follows:

$$y = f[\boldsymbol{x}] = \widetilde{f}\left[x_1, x_2, \ldots, x_K \mid x_1 \in \omega_{n_1}^1, x_2 \in \omega_{n_2}^2, \ldots, x_K \in \omega_{n_K}^K\right] \tag{5}$$

where $\widetilde{f}[\cdot]$ is a multitemporal mapping associated with the combination of modes $\left(\omega_{n_1}^1, \omega_{n_2}^2, \ldots, \omega_{n_K}^K\right)$. Accordingly, for each possible multitemporal combination of modes, a prediction function $\widetilde{f}[\cdot]$ needs to be defined.

Under a linear prediction model, the function $\widetilde{f}[\cdot]$ associated with the multitemporal combination of modes $\left(\omega_{n_1}^1, \omega_{n_2}^2, \ldots, \omega_{n_K}^K\right)$ can be written as:

$$y = \widetilde{f}\left[x_1, x_2, \ldots, x_K \mid x_1 \in \omega_{n_1}^1, x_2 \in \omega_{n_2}^2, \ldots, x_K \in \omega_{n_K}^K\right] = \sum_{j=1}^{K} \beta_j \cdot x_j \tag{6}$$

where β_j stands for the weight assigned to the jth mode of the combination. The determination of the weight values can be addressed by the minimum square error pseudo-inverse technique. This consists of solving the following system of R linear equations with K unknown variables ($R > K$):

$$\begin{bmatrix} x_1^1 & x_2^1 & \ldots & x_K^1 \\ x_1^2 & x_2^2 & \ldots & x_K^2 \\ \ldots & \ldots & \ldots & \ldots \\ x_1^R & x_2^R & \ldots & x_K^R \end{bmatrix} \cdot \begin{bmatrix} \beta_1 \\ \beta_2 \\ \ldots \\ \beta_K \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \ldots \\ y_R \end{bmatrix} \quad \Leftrightarrow \quad \underline{P} \cdot \beta = \underline{Y} \tag{7}$$

where R represents the number of multitemporal vectors $\boldsymbol{x}^r = [x_1^r, x_2^r, \ldots, x_K^r]$ ($r \in \{1, 2, \ldots, R\}$) observed in $\overline{C}$ and collected in $\underline{P}$, such that $x_1^r \in \omega_{n_1}^1, x_2^r \in \omega_{n_2}^2, \ldots,$ and $x_K^r \in \omega_{n_K}^K$. The corresponding observations y_r in the cloudy image $X^{(i)}$ (to be reconstructed) are gathered in the target vector $\underline{Y}$. The estimate of the optimal weight vector β^* is given by the following equation based on the pseudoinverse $\underline{P}^{\#}$ of the matrix $\underline{P}$:

$$\beta^* = (\underline{P}^t \cdot \underline{P})^{-1} \cdot \underline{P}^t \cdot \underline{Y} = \underline{P}^{\#} \cdot \underline{Y} \tag{8}$$

The prediction system involved in the contextual multiple linear prediction (CMLP) method is thus made up of an ensemble of linear predictors, each trained to learn the relationship between images $X^{(k)}$ ($k \in S_c$) and image $X^{(i)}$ over a possible multitemporal combination of classes $\left(\omega_{n_1}^1, \omega_{n_2}^2, \ldots, \omega_{n_K}^K\right)$ in $N(C)$. In addition, we integrate the ensemble with an additional linear predictor, termed global predictor, which is trained over all samples of $\overline{C}$ independently of their class membership. The motivation behind such an integration is that the global predictor is useful to deal with one of the following two possible situations: (1) a combination of classes in C does not exist in the set of feasible combinations of classes identified in $\overline{C}$; or (2) the number of samples collected in $\underline{P}$ and $\underline{Y}$ for a given multitemporal combination of classes available in $\overline{C}$ is not enough to apply Eq. (8).

3.2 Unsupervised Classification with the EM Algorithm

The first step of the contextual prediction process is that of classifying the region $N(C)$ of each image $X^{(k)}$ ($k \in S_c$) into a set of M_k data classes. This can be done in an unsupervised way (i.e., without the need of training samples) by means of the EM algorithm [8]. The EM algorithm is an iterative procedure that converges to local but usually good Maximum Likelihood (ML) estimates of mixture parameters. It is based on the interpretation of $\widetilde{X}^{(k)} = \left\{x(u,v) : x(u,v) \in X^{(k)}, k \in S_c, (u,v) \in N(C)\right\}$ as incomplete data where the missing part is $Z^{(k)}$, i.e., its classification map. Assuming that L is the number of pixels in $\widetilde{X}^{(k)}$, the missing part can be evaluated as a set of L labels $Z^{(k)} = \left\{\boldsymbol{z}_k^{(1)}, \boldsymbol{z}_k^{(2)}, \ldots, \boldsymbol{z}_k^{(L)}\right\}$ associated with the L pixels, indicating which class is at the origin of each pixel realization. Each label is a binary vector $\boldsymbol{z}_k^{(i)} = [z_{k,1}^{(i)}, z_{k,2}^{(i)}, \ldots, z_{k,M_k}^{(i)}]$, such that $z_{k,r}^{(i)} = 1$ ($r \in \{1, 2, \ldots, M_k\}$) if the ith pixel x_k^i of $\widetilde{X}^{(k)}$ belongs to the rth data class ω_r^k, and $z_{k,r}^{(i)} = 0$ otherwise. The complete log-likelihood function, from which it would be possible to estimate the vector of parameters $\Theta^k = [P_1^k, P_2^k, \ldots, P_{M_k}^k, \mu_1^k, \mu_2^k, \ldots, \mu_{M_k}^k, \sigma_1^k, \sigma_2^k, \ldots, \sigma_{M_k}^k]$ if the complete data $\Psi^{(k)} = \{\widetilde{X}^{(k)}, Z^{(k)}\}$ were observed, is given by:

$$\log p(\Psi^{(k)} \mid \Theta^k) = \ell(\Psi^{(k)} \mid \Theta^k) = \sum_{i=1}^{L} \sum_{r=1}^{M_k} z_{k,r}^{(i)} \ln \left[P_r^k \, p(x_k^i \mid \theta_r^k)\right] \tag{9}$$

where $\theta_r^k = [\mu_r^k, \sigma_r^k]$.

The quantity $z_{k,r}^{(i)}$ can be estimated as the conditional expectation of $z_{k,r}^{(i)}$ given the observation x_k^i and the set of parameters Θ^k. The EM algorithm consists of expectation and maximization steps, which are iterated up to convergence. The expectation step is represented by the computations of $z_{k,r}^{(i)}$ ($i = 1, 2, \ldots, L$ and $r = 1, 2, \ldots, M_k$) using the current estimates of the set of parameters Θ^k. The maximization step allows updating such parameter estimates.

At convergence of the EM algorithm, the final parameter estimates will define completely the Gaussian data classes (modes) available in $\widetilde{X}^{(k)}$. The latter is then transformed into a classification map with minimum error by adopting the well-known maximum a posteriori probability (MAP) decision rule. Since the final estimates of $z_{k,r}^{(i)}$ represent the estimates of the posterior probabilities $P(\omega_r^k \mid x_k^i)$ ($i = 1, 2, \ldots, L$ and $r = 1, 2, \ldots, , M_k$), one can assign to each pixel x_k^i of $\widetilde{X}^{(k)}$ the optimal class label $\hat{\omega} \in \Omega = \left\{\omega_r^k : r = 1, 2, \ldots, M_k\right\}$, such that:

$$\hat{\omega} = \arg \max_{\omega_r^k \in \Omega} \{P(\omega_r^k \mid x_k^i)\} \tag{10}$$

3.3 Estimation of the Number of Classes

The number of data classes M_k, which is not known a priori, is estimated using the Minimum Description Length (MDL) criterion [9]:

$$MDL(M_k) = -\widetilde{\ell}(\Psi^{(k)} \mid \Theta^k) + \gamma \cdot \kappa \cdot \log(L) \tag{11}$$

where $\widetilde{\ell}(\Psi^{(k)} \mid \Theta^k)$ represents the log-likelihood function value found at convergence of the EM algorithm, κ is the number of parameters in Θ^k, and γ is a constant.

The optimal number of data classes $\hat{M}_k$ in $\widetilde{X}^{(k)}$ is estimated by minimizing the MDL criterion, i.e.,

$$\hat{M}_k = \arg \min_{M_k = 1, \ldots, M_{\max}} \{MDL(M_k)\} \tag{12}$$

where $M_{\max}$ is a predefined maximal number of data classes.

4 Compressive Sensing Reconstruction Approach

The second approach exploits the Compressive Sensing (CS) theory recently developed by Donoho [10], which offers the capability to recover an unknown sparse signal with a linear combination of a small number of elementary samples called atoms. The fundamental of the CS theory is the ability to recover with relatively few measurements $\mathbf{x} = D \cdot \boldsymbol{\alpha}$ by solving the following L_0-minimization problem:

$$\min \|\boldsymbol{\alpha}\|_0 \text{ subject to } \mathbf{x} = D \cdot \boldsymbol{\alpha} \tag{13}$$

where D is a dictionary with a certain number of atoms, $\mathbf{x}$ is the original signal which can be represented as a sparse linear combination of these atoms, and the minimization of $\| \cdot \|_0$, the L_0-norm, corresponds to the maximization of the number zeros in $\boldsymbol{\alpha}$, following this formulation: $\|\boldsymbol{\alpha}\|_0 = \#\{i : \alpha_i \neq 0\}$. Equation (13) represents

a NP-hard problem, but can be simplified in a relatively easily linear programming solution. Under some reasonable assumptions, minimizing L_1-norm is equivalent to minimizing L_0-norm, which is defined as $\|\boldsymbol{\alpha}\|_1 = \sum_i |\alpha_i|$. Accordingly, it is possible to rewrite (13) as:

$$\min \|\boldsymbol{\alpha}\|_1 \text{ subject to } \mathbf{x} = D \cdot \boldsymbol{\alpha} \tag{14}$$

The CS formulation can be adapted to the investigated reconstruction problem as described in Sect. 2 of this Chapter. More details are provided in [7].

Several algorithms have been proposed in the literature to solve the optimization problem expressed in (13). In the following, first, we show how two common CS solutions, the orthogonal matching pursuit (OMP) and the basis pursuit (BP) algorithms, can be formulated for a cloud-contaminated image reconstruction problem. Then, we illustrate an alternative solution for solving the CS problem under a L_0-norm perspective, by exploiting the capabilities of genetic algorithms.

4.1 CS Solutions

The basis pursuit (BP) principle represents a traditional solution to solve the CS problem [11]. It suggests a convexification of the problem by using the L_1-norm instead of L_0. This means that the best approximation of the problem becomes equal to a support minimization problem. BP finds signal representations in overcomplete dictionaries by convex, nonquadratic optimization technique. It can be reformulated as a linear programming (LP) problem, and solved using modern interior-point methods, simplex methods, or other techniques, such as homotopy techniques. Note that, if the original signal $\mathbf{x}$ is sufficiently sparse, the recovery via BP is provably exact.

One of the easiest and fastest alternative technique is the orthogonal matching pursuit (OMP) [12], an improved version of the matching pursuit (MP) method. MP finds the atom that has the highest correlation with the signal. It subtracts off the correlated part from the signal and then iterates the procedure on the resulting residual signal. Unfortunately the convergence speed of this algorithm is not fast. To overcome this drawback, an improved solution called orthogonal MP (OMP) was developed. Differently from MP, OMP updates the coefficients of the selected atoms at each iteration so that the resulting residual vectors are orthogonal to the subspace spanned by the selected atoms. When stopped after only few iterations, it generally yields a satisfactory approximation, using only few atoms.

4.2 Genetic Algorithm-Based CS Solution

A third CS strategy that we investigate in this part of the chapter is based on genetic algorithms (GAs) [13]. A genetic optimization algorithm performs a search by regenerating a population of candidate solutions (or individuals) represented by chromo-

somes. From one generation to the next, the population is improved following biological rules, adopting deterministic and nondeterministic genetic operators. In general, GA involves the following steps. First, an initial population of chromosomes is randomly generated. Then, the goodness of each chromosome is evaluated according to a predefined fitness function representing the aim of the optimization. Once the selection of the best chromosomes is done, the next step is devoted to the reproduction of a new population. This is done by genetic operators such as crossover and mutation operators. All these steps are iterated until some predefined condition is satisfied.

Several multiobjective GA-based approaches have been proposed in the literature. We adopt the nondominated sorting solution (NSGA-II) [14] for its low computational requirements, its aptitude to distribute uniformly the optimal solutions along the Pareto front and its successful application to different remote sensing problems. The design of a multiobjective GA optimization relies upon two components, the chromosome structure and the fitness functions, which encode the considered optimization problem and show the direction to obtain the best solution, respectively.

Concerning the first component, we consider a population of M chromosomes $\boldsymbol{\alpha}_m \in \Re$, $m \in \{1, 2, \ldots, M\}$, where each chromosome is a real vector composed of genes corresponding to the weight vector $\boldsymbol{\alpha}$ defined above in the previous sections. The length w of the chromosome is thus equal to the one of the dictionary D.

Regarding the fitness function, we investigate separately and jointly two fitness functions, i.e., those defining the optimization problem in (13). The first one aims at maximizing the sparsity level by minimizing the L_0-norm of the weight vector $\boldsymbol{\alpha}$, which corresponds to minimize the number of almost nonzero coefficients in $\boldsymbol{\alpha}$:

$$f_1 = \min \|\boldsymbol{\alpha}\|_0. \tag{15}$$

An almost nonzero coefficient is a coefficient exhibiting a value less than a predefined small threshold value. The second fitness function is derived from the constraint in [13] and points to a perfect reconstruction of the considered pixel:

$$f_2 = \min \|D\boldsymbol{\alpha} - \mathbf{x}\|_0. \tag{16}$$

5 Experimental Results

A. Dataset description and experimental design

The two investigated reconstruction approaches have been evaluated experimentally on a real multitemporal multispectral remote sensing image. The dataset was acquired by the Taiwanese optical high resolution FORMOSAT-2 satellite and represent part of the Arcachon basin in the south region of Aquitaine, in France. The images were acquired on the 24th of June and the 16th of July 2009, respectively, and are composed of 400×400 pixels, 4 spectral bands (blue, green, red and near-infrared) with a spatial resolution of 8 m (see Fig. 1).

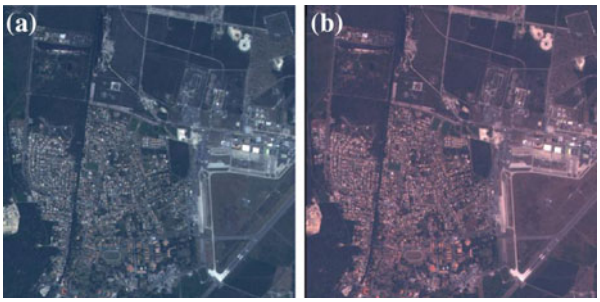

Fig. 1 FORMOSAT-2 images acquired in the Arcachon area on **a** the 24th of June and **b** the 16th of July, 2009

In order to quantify the reconstruction accuracy, the experiments were done in this way: (1) a cloud-free image $I^{(1)}$ is considered; (2) the presence of clouds is simulated by partly obscuring the other image $I^{(2)}$; and (3) the reconstructed image is compared with the original cloud-free image. The simulation study aims at understanding the sensitivity of the four investigated methods (i.e., the CMLP, the OMP, the BP, and the GA) to two aspects: (1) the kind of ground covers obscured and (2) the size of the contaminated area. In terms of comparisons, we considered also a recent work based on multiresolution inpainting (MRI). In order to obtain a detailed assessment of the reconstruction quality, we adopt the well-known peak signal-to-noise ratio (PSNR) measure. Other quantitative criteria are the computation time (in seconds) and the model complexity, namely the number of coefficients required for the reconstruction model. Regarding the dictionaries, we collected directly training samples from the image, by sampling pixels in the source region Φ.

B. Simulation experimental results

Figure 2a shows different masks whose positions were selected in a way to simulate the obscuration of different kinds of ground cover. In particular, mask A covers a region that includes mainly a urban area, mask B obscures an industrial zone, and

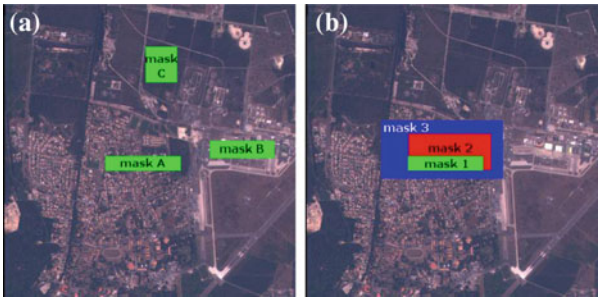

Fig. 2 Masks adopted to simulate **a** the contamination of different ground covers and **b** the different sizes of contamination

Table 1 Quantitative results obtained in the first simulation experiments

	Method	MRI	CMLP	OMP	BP	GA
Mask A	PSNR l_1	–	–	39.41	80.59	42.09
	PSNR l_2	22.54	20.99	23.96	22.22	23.78
	Complexity	–	1	3	294	148
	Time [s]	2856	1	4	66	68621
Mask B	PSNR l_1	–	–	36.33	77.10	43.38
	PSNR l_2	16.05	20.11	20.60	24.74	23.15
	Complexity	–	1	3	168	95
	Time [s]	2517	1	4	59	26312
Mask C	PSNR l_1	–	–	44.28	98.53	45.62
	PSNR l_2	33.77	24.05	31.97	30.67	32.01
	Complexity	–	1	3	301	138
	Time [s]	2898	1	4	60	43193

mask C covers a vegetation area. The experiments were carried out by considering each mask at a time, where each mask is composed by around 2000 pixels and the dictionary by 300 pixels.

Table 1 reports the results of the different reconstruction techniques over different obscured land covers. In greater detail, MRI generally reconstructs the missing data with a good PSNR level, but the corresponding reconstructed images appear visually of poor quality since it does not capture satisfactorily the textural properties of the missing areas. In general, MRI can return visually satisfactory results only when the missing area refers to a uniform region such as vegetation region. CMLP method provides generally satisfactory results in terms of reconstruction error and computation time. To obtain better results, it would need more than two temporal images. Coming now to the CS-based implementations, OMP algorithm produces very sparse reconstruction solutions (around 3 nonzero coefficients). On the one hand, this may be an advantage in terms of computation time. On the other hand, OMP is potentially subject to underfitting problems. On the contrary, BP algorithm may be subject to overfitting problems due to the fact that most of the time it selects a large number of weight coefficients (in general around 300 coefficients). Comparing OMP and BP in terms of computation time, the latter is far less efficient, whereas in terms of PSNR both methods return similar reconstruction values, outperforming CMLP and MRI. Lastly, GA can be viewed as a compromise between the two previous methods. Despite the very long time needed to estimate the reconstruction model, it results sparser than BP, but less parsimonious to OMP (see model complexity columns in Table 1). Its reconstruction error is almost all the time the best or the second best.

Another important test for the five methods consists of assessing their performances by varying the amount of missing data. Figure 2b illustrates the three different masks adopted to simulate different increasing cloud cover sizes. In particular, masks 1 is the same as the mask A adopted in the previous experiments, i.e., it covers about 2000 pixels. To build masks 2 and 3, we multiplied the previous size by 3 and

Table 2 Quantitative results obtained in the second simulation experiments

	Method	MRI	CMLP	OMP	BP	GA
Mask 1	PSNR l_1	–	–	39.41	80.59	42.09
	PSNR l_2	22.54	20.99	23.96	22.22	23.78
	Complexity	–	1	3	294	148
	Time [s]	2856	1	4	66	68621
Mask 2	PSNR l_1	–	–	42.45	80.18	45.46
	PSNR l_2	21.35	21.13	23.21	22.89	23.85
	Complexity	–	1	7	277	140
	Time [s]	6938	1	6	145	99072
Mask 3	PSNR l_1	–	–	42.00	79.53	45.13
	PSNR l_2	19.63	20.83	25.01	21.47	23.03
	Complexity	–	1	3	265	149
	Time [s]	14774	2	19	865	275394

by 6, respectively. Also in these experiments, the adopted dictionaries are composed of 300 pixels belonging to the Φ region. Table 2 reports the results achieved by the different reconstruction techniques and by varying the amount of missing data.

From a quantitative viewpoint, in terms of PSNR, we have similar results as in the previous experiments. MRI still presents problems in reconstructing satisfactorily complex textures. CMLP competes seriously with MRI in terms of computation time and PSNR. However to get higher PSNR values, one needs to resort to CS techniques. Indeed, the CS-based implementations return better results in term of PSNR in all the simulations and present the advantage for not depending on the size of the missing area. The best solution in these experiments in terms of PSNR comes from GA. About the computation time, as expected, it increases as the amount of missing data increases. Results from this viewpoint underline the main weakness of the GA solution, i.e., its expensive computational needs. Regarding the model complexity, we got in these experiments results in line with those of the previous series of experiments.

Figure 3 shows qualitative reconstruction results in RGB composites obtained in the most critical reconstruction scenario (i.e., mask 3). As mentioned before, MRI reconstruction exhibits the worst reconstruction case. CMLP method is capable to obtain a good reconstruction compared with MRI. Regarding the CS reconstruction techniques (OMP, BP, and GA), good reconstructions are obtained.

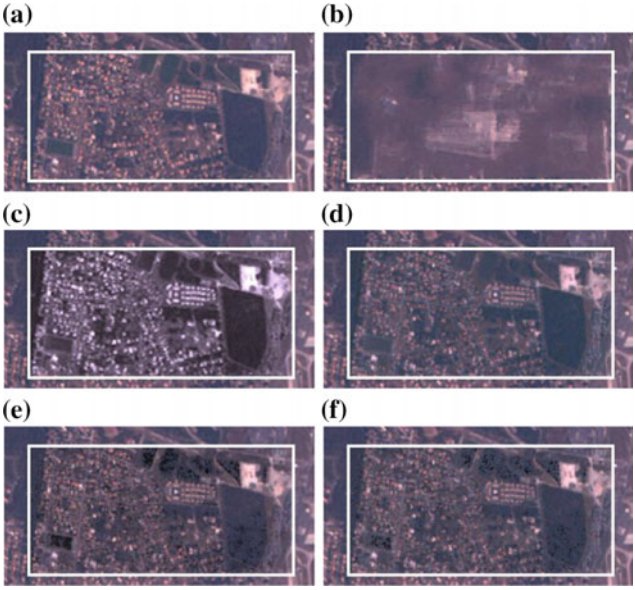

Fig. 3 Color composite (**a**) of the original image and the same image reconstructed after the contamination with the largest simulated mask 3 by **b** MRI, **c** CMLP, **d** OMP, **e** BP, and **f** GA methods

6 Conclusion

In this chapter we have dealt with the problem of the removal of clouds from sequences of multitemporal multispectral optical images. Given a contaminated image of a sequence, each area of missing measurements is reconstructed. Two main approaches have been investigated, the first one based on a contextual prediction system (CMLP) and the second one based on compressive sensing (CS), for which three different solutions based on basis pursuit (BP), orthogonal matching pursuit (OMP) and genetic algorithm (GA) have been considered.

The experimental results stress that the kind of ground cover obscured and the size of the contaminated area only marginally affect the performance of the first reconstruction approach. The latter depends more directly on the representativeness of the samples extracted from the spatial neighborhood of the contaminated area and used to train the predictors. The results point out also the superiority of the CS approach. Comparing the CS solutions, OMP has the advantage to be sparser and significantly faster than BP and GA, but it is the less robust method. GA can be seen as a good compromise between OMP and BP methods, mainly because it is more robust than OMP and more sparse than BP.

References

1. Cihlar, J.: Remote sensing of global change: an opportunity for Canada. In: Proceedings of 11th Canadian Symposium on Remote Sensing, pp. 39–48. Waterloo, Canada (1987)
2. Vasudevan, B.G., Gohil, B.S., Agarwal, V.K.: Backpropagation neural-network-based retrieval of atmospheric water vapor and cloud liquid water from IRS-P4 MSMR. IEEE Trans. Geosc. Remote Sens. **42**, 985–990 (2004)
3. Banham, M.R., Katsaggelos, A.K.: Digital image restoration. IEEE Sign. Process. Mag. **14**, 24–41 (1997)
4. Tseng, D.-C., Tseng, H.-T., Chien, C.-L.: Automatic cloud removal from multi-temporal SPOT images. Appl. Math. Comput. **205**(2), 584–600 (2008)
5. Lin, C.-H., Tsai, P.-H., Lai, K.-H., Chen, J.-Y.: Cloud removal from multitemporal satellite images using information cloning. IEEE Trans. Geosc. Remote Sens. **51**(1), 232–241 (2013)
6. Melgani, F.: Contextual reconstruction of cloud-contaminated multitemporal multispectral images. IEEE Trans. Geosc. Remote Sens. **44**(2), 442–455 (2006)
7. Lorenzi, L., Melgani, F., Mercier, G.: Missing area reconstruction in multispectral images under a compressive sensing perspective. IEEE Trans. Geosc. Remote Sens. **51**(7), 3998–4008 (2013)
8. Moon, T.K.: The expectation-maximization algorithm. Sig. Process. Mag. **13**, 47–60 (1996)
9. Rissanen, J.: Stochastic Complexity in Statistical Enquiry. World Scientific, Singapore (1989)
10. Donoho, D.L.: Compressed sensing. IEEE Trans. Inf. Theor. **52**(4), 1289–1306 (2006)
11. Chen, S.S., Donoho, D.L., Saunders, M.A.: Atomic decomposition by basis pursuit. SIAM J. Sci. Comp. **20**, 33–61 (1999)
12. Pati, Y.C., Rezaiifar, R., Krishnaprasad, P.S.: Orthogonal matching pursuit: recursive function approximation with applications to wavelet decompositions. In: Proceedings of 27th Asilomar Conferences on Signal System and Computers (1993)
13. Deb, K.: Multi-Objective Optimization Using Evolutionary Algorithms. Wiley, Chichester (2001)
14. Srinivas, N., Deb, K.: Multiobjective function optimization using nondominated sorting genetic algorithms. Evol. Comput. **2**(3), 221–248 (1995)

View publication stats