Missing-Area Reconstruction in Multispectral Images Under a Compressive Sensing Perspective

Luca Lorenzi, *Student Member, IEEE*, Farid Melgani, *Senior Member, IEEE*, and Grégoire Mercier, *Senior Member, IEEE*

***Abstract*—The intent of this paper is to propose new methods for the reconstruction of areas obscured by clouds. They are based on compressive sensing (CS) theory, which allows finding sparse signal representations in underdetermined linear equation systems. In particular, two common CS solutions are adopted for our reconstruction problem: the basis pursuit and the orthogonal matching pursuit methods. A novel alternative CS solution is also proposed through a formulation within a multiobjective genetic optimization scheme. To illustrate the performances of the proposed methods, a thorough experimental analysis on FORMOsa SATellite-2 and Satellite Pour l'Observation de la Terre-5 multispectral images is reported and discussed. It includes a detailed simulation study that aims at assessing the accuracy of the methods in different qualitative and quantitative cloud-contamination conditions. Compared with state-of-the-art techniques for cloud removal, the proposed methods show a clear superiority, which makes them a promising tool in cleaning images in the presence of clouds.**

***Index Terms*—Cloud removal, compressive sensing (CS), genetic algorithm (GA), image reconstruction, missing data, sparse regression, sparse representation.**

I. Introduction

DEPENDING on the end-user requirements, clouds in remotely sensed imagery may or may not represent an unwanted source of noise. In case they are viewed as a noise source, several methodologies have been developed in the past in order to cope with this problem. Generally, the common approach first detects the contaminated regions and, in a second instance, attempts to remove the clouds by substituting them with cloud-free estimations. In this paper, we will focus on the second part of the approach. In the related literature, one of the first proposed techniques produces a cloud-free image mosaic by compositing several cloudy optical satellite images acquired from the same area [1]. Other techniques reconstruct missing areas by exploiting the inpainting approach. For instance, a recent work suggests to enrich the region (patch) search process by including local image properties or by isometric transformations or to reformulate the problem under a multiresolution processing scheme [2]. Another synthesis technique was proposed by Maalouf *et al.* [3], where information is propagated by a technique based on the bandelet transform, a special case of the wavelet transform. Accurate reconstructions may be obtained through prediction techniques if temporal information is available. A first work exploiting the temporal prediction principle was presented in [4], where a least squares linear prediction with an escalator structure is implemented. Unsupervised contextual prediction was also adopted in [5], where local spectrotemporal relationships are used to predict missing data through linear or nonlinear regression. Tseng *et al.* [6] proposed a method to correct radiometric inconsistencies of cloud-contaminated images and their corresponding temporal images by generating a cloud-free mosaic image for a multitemporal SPOT data set. In order to ameliorate the transition between two mosaic parts, a wavelet-based fusion is adopted. More recently, a cloud removal method based on information cloning has been developed [7]. The authors proposed to clone cloud-free information from a set of multitemporal images by adopting a patch-based reconstruction method formulated as a Poisson equation and solved using a global optimization process.

Recently, compressive sensing (CS), also known as compressive sampling or sparse representation, has been introduced by Donoho [8] and Candès *et al.* [9]. CS theory aims at recovering an unknown sparse signal from a small set of linear projections. By exploiting this new and important result, it is possible to obtain equivalent or better representations by using less information compared with traditional methods (i.e., lower sampling rate or smaller data size) [10]. CS has been proved to be a powerful tool for several applications, such as acquisition, representation, regularization in inverse problems, feature extraction, and compression of high-dimensional signals, and applied in different research fields: signal processing, object recognition, data mining, and bioinformatics [11]. In these fields, CS has been adopted to cope with several tasks: face recognition [12], image superresolution [13], segmentation [14], denoising [15], inpainting [16], and classification [17]. Note that images are a special case of signals which hold a natural sparse representation, with respect to fixed bases, also called dictionary (i.e., Fourier and wavelet transforms) [18]. In the literature, a common dictionary choice is the one proposed in [11], the k-means singular-value decomposition. It adapts

Manuscript received February 28, 2012; revised September 26, 2012; accepted October 24, 2012. Date of publication January 29, 2013; date of current version June 20, 2013.

L. Lorenzi is with the Signal Processing and Recognition Laboratory, Department of Information Engineering and Computer Science, University of Trento, 38123 Trento, Italy, and also with the Image and Information Processing Department (ITI), Telecom Bretagne, Institut Telecom, 29238 Brest, France (e-mail: lorenzi@disi.unitn.it; luca.lorenzi@telecom-bretagne.eu).

F. Melgani is with the Signal Processing and Recognition Laboratory, Department of Information Engineering and Computer Science, University of Trento, 38123 Trento, Italy (e-mail: melgani@disi.unitn.it).

G. Mercier is with the Image and Information Processing Department (ITI), Telecom Bretagne, Institut Telecom, 29238 Brest, France (e-mail: gregoire.mercier@telecom-bretagne.eu).

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/TGRS.2012.2227329

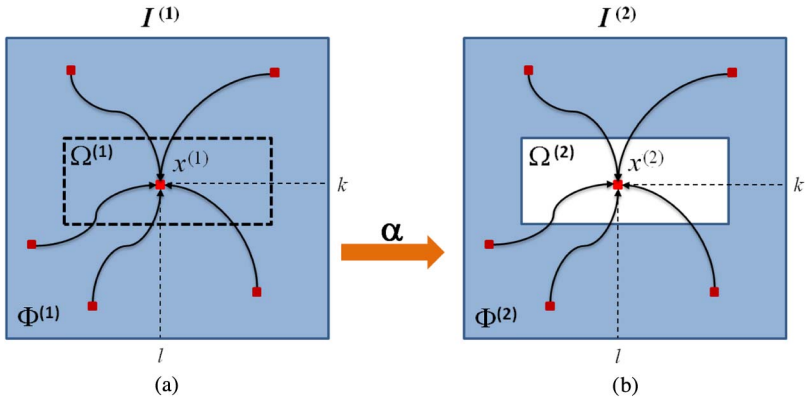


Fig. 1. Illustration of the reconstruction principle under a CS perspective.

dictionaries by iteratively alternating between sparse coding of exemplar based on the current dictionary and a process of updating the dictionary atoms to better fit the data. However, to simplify the selection of a proper dictionary, in some cases, the training pixels or patches (which are supposed to be known) are collected directly from the image of interest [12], [16] or by random sampling the training area [19]. More recently, CS theory has been applied also in the remote sensing (RS) field. For example, there exist specific applications such as the following.

1) For image pan sharpening as in [19], where CS theory ensures that a compressible signal can be correctly recovered from global linear sampled data.
2) For hyperspectral image classification [20], where two approaches have been proposed.
   a) In the first one, an explicit smoothing constraint is imposed by forcing the Laplacian vector to zero.
   b) The second approach relies on a joint sparsity model where hyperspectral pixels are simultaneously represented by linear combinations.
3) For superresolution restoration [21], where high-resolution image patches are recovered from the downsampled low-resolution ones.
4) For saving camera size, power consumption, and storage burden during the RS image acquisition process [22].

In this paper, we propose three novel methods to solve the problem of the reconstruction of missing data due to the presence of clouds. Given a cloud-free and a cloud-contaminated image, each of the missing measurements is recovered by applying the CS theory in which cloud-free pixels are exploited. In more detail, we first apply two of the most common CS methods which approximate the CS solution, namely, the basis pursuit (BP) and the orthogonal matching pursuit (MP) (OMP). Furthermore, we propose an alternative CS solution which exploits the search capabilities of genetic algorithms (GAs). Experimental results on two temporal multispectral images acquired by the FORMOSAT-2 and SPOT-5 optical satellites are reported and discussed. They include a thorough simulation to assess accurately the performances of the three proposed methods and other two cloud removal techniques in terms of reconstruction quality.

This paper is organized as follows. In the next section, we formulate the cloud-contaminated-image reconstruction problem. In Section III, we introduce the CS theory and describe two common CS methods, whereas the GA-based method is described in Section IV. The experimental results are illustrated in Section V. Section VI draws the conclusions.

II. Problem Formulation

Let us consider two multispectral (B-channel) images $I^{(1)}$ and $I^{(2)}$ acquired by an optical sensor at two different dates and registered over the same geographical area. Let us suppose that the two acquisitions are temporally close to each other, and thus, the images are characterized by a similar spatial structure. The objective of the proposed methods is to reconstruct any area contaminated by clouds. We remark that the cloud detection (i.e., finding the position of clouds in an image) is out of the scope of this paper. At this level, we make the hypothesis that image $I^{(2)}$ is obscured by the presence of clouds, and a cloud/noncloud binary classifier has been adopted to discriminate between these two classes. We will call this cloudy area in image $I^{(2)}$ as target region $\mathbf{\Omega}^{(2)}$ and the remaining part as source region $\mathbf{\Omega}^{(2)}$, subject to $I^{(2)} = \mathbf{\Omega}^{(2)} \cup \mathbf{\Phi}^{(2)}$ (following the classical notation in the inpainting literature). Image $I^{(1)} = \mathbf{\Omega}^{(1)} \cup \mathbf{\Phi}^{(1)}$ does not contain clouds since it is supposed cloud free. Our aim is to generate a new image $I^{(2)}$ without clouds.

We will assume that any pixel $x^{(1)} \in \mathbf{\Omega}^{(1)}$ can be expressed as a linear combination of pixels in region $\mathbf{\Phi}^{(1)}$ of $I^{(1)}$ [see Fig. 1(a)]. In other words, in $I^{(1)}$, we have

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

where $\boldsymbol{\alpha}$ is an unknown weight vector associated with the considered pixel $x^{(1)}$ and having the same dimension as the number of pixels belonging to $\mathbf{\Phi}^{(1)}$. The problem at this point is to infer $\boldsymbol{\alpha} = f(\mathbf{\Phi}^{(1)}, x^{(1)})$. Once $\boldsymbol{\alpha}$ is computed, if we assume that $I^{(1)}$ and $I^{(2)}$ are temporally close, so that the scene did not change in between the two observations, it will be possible to reuse the $\boldsymbol{\alpha}$ coefficients to reconstruct the spatially corresponding pixel in the missing area $\mathbf{\Omega}^{(2)}$, adopting the previous formulation for $I^{(2)}$, i.e., $\hat{x}^{(2)} = \mathbf{\Phi}^{(2)} \cdot \mathbf{\Phi}^{(2)} \cdot 2$ [see Fig. 1(b)]. In other words, for each pixel $x^{(1)} \in \mathbf{\Omega}^{(1)}$, we evaluate $\boldsymbol{\alpha}$, and in a second

moment, we reuse this weight vector to return an estimation of $x^{(2)} \in \mathbf{\Omega}^{(2)}$

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

where $f(\cdot)$ represents an estimation function. Finally, we note that, different from those of reference techniques as the one presented in [5], all image channels are processed simultaneously. In the next two sections, different methods are proposed to solve the issue of the estimation of the weight vectors.

III. Reconstruction via CS

A. Generalities on CS

CS is a useful way to obtain a sparse representation of a signal. It relies on the idea to exploit redundancy (if any) in the signals [8], [9]. Usually, signals like images are sparse, as they contain, in some representation domain, many coefficients close to or equal to zero. CS starts taking a weighted linear combination of pixels in a basis in which the signal is assumed to be sparse. The fundamental of the CS theory is the ability to recover with relatively few measurements $x = D \cdot \boldsymbol{\alpha}$ by solving the following L_0-minimization problem:

$$\min \|\boldsymbol{\alpha}\|_0 \qquad \text{subject to } x = D \cdot \boldsymbol{\alpha} \tag{3}$$

where D is a dictionary with a certain number of atoms, 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 (3) represents a nondeterministic polynomial-time hard problem, which means that it is computationally infeasible to solve. Following the discussion of Candès and Tao [23], it is possible to simplify the evaluation of (3) in a relatively easy linear programming (LP) solution. They demonstrate that, under some reasonable assumptions, minimizing the L_1 norm is equivalent to minimizing the L_0 norm, which is defined as $\|\boldsymbol{\alpha}\|_1 = \sum_i |\alpha_i|$. Accordingly, it is possible to rewrite (3) as

$$\min \|\boldsymbol{\alpha}\|_1 \qquad \text{subject to } x = D \cdot \boldsymbol{\alpha}. \tag{4}$$

In the literature, there exist several algorithms for solving optimization problems similar to the one expressed in (4). In the next sections, we briefly introduce two of them, which represent the most common solutions from the literature.

B. CS Solutions

1) BP: A well-known solution for problem (4) is the BP principle [23], [24]. 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 a convex nonquadratic optimization technique, solving problem (4). It can be reformulated as an LP problem and solved using modern interior-point methods, simplex methods, or other techniques, such as homotopy techniques [25]. Given that, it is possible to rewrite the L_1 norm in (4) as

$$\|\boldsymbol{\alpha}\|_1 = \sum_i |\alpha_i| = \sum_i u_i + v_i,$$
$$\text{where } \begin{cases} \alpha_i = u_i & v_i = 0, \quad \text{if } \alpha_i \geq 0 \\ \alpha_i = -v_i & u_i = 0, \quad \text{if } \alpha_i \leq 0. \end{cases} \tag{5}$$

Substituting it in (4), it allows performing a linear minimization problem. Note that, if the original signal x is sufficiently sparse, the recovery via BP is provably exact.

2) OMP: One of the easiest and fastest alternative techniques is the OMP, an improved version of the 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 [26], [27]. The algorithm approximates the signal x considering these two decompositions [24]

$$x = \sum_{d \in D} \alpha_d \varphi_d = \sum_{i=1}^{m} \alpha_{d_i} \varphi_{d_i} + R^{(m)} \tag{6}$$

where dictionary D is a collection of atom vectors $\{\varphi_d\}_{d \in D}$ and $R^{(m)}$ is a residual. Starting from an initial approximation $x^{(0)} = 0$ and residual $R^{(0)} = x$, it builds up a sequence of sparse approximations stepwise. At stage k, it identifies the dictionary atom that best correlates with the residual and then adds to the current approximation a scalar multiple of that atom, so that $x^{(k)} = x^{(k-1)} + \alpha_k \varphi_{d_k}$, where $\alpha_k = \langle R^{(k-1)}, \varphi_{d_k} \rangle$ and $R^{(k)} = x - x^{(k)}$. After m steps, one has a representation of the form (6), with residual $R = R^{(m)}$, where the original signal x is decomposed into a sum of dictionary elements that are chosen to best match its residues.

Unfortunately, the convergence speed of this algorithm is not fast. To overcome this drawback, an improved solution called OMP was developed. Different 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 [26], [27].

3) BP Versus OMP: From the literature [28], [29], it comes out that the BP and OMP algorithms provide, in general, good performances in reconstruction problems. Nonetheless, BP is considered more powerful than OMP, since it can recover with high probability all sparse signals and is more stable. On the contrary, OMP results attractive for its fast convergence and in its ease of implementation.

IV. Genetic Algorithm

A. General Concepts on GA

GAs are a part of evolutionary computation which solves optimization problems by mimicking the principles of biological evolution [30], [31]. A genetic optimization algorithm performs a search by regenerating a population of candidate solutions (or individuals) represented by chromosomes. From

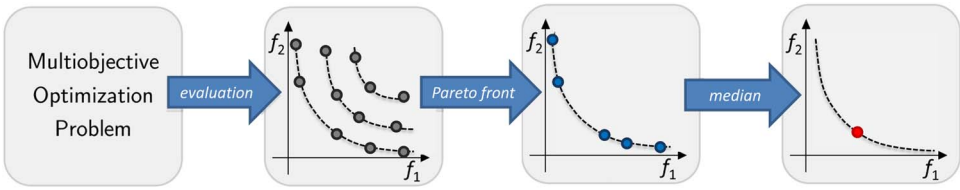


Fig. 2. Illustration of a multiobjective genetic procedure.

one generation to the next, the population is improved following biological rules, adopting deterministic and nondeterministic genetic operators. In general, a common 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. Evaluating the fitness function allows keeping or discarding chromosomes, by using a proper rule based on the principle that, the better the fitness, the higher the chance of being selected. 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 (e.g., maximum number of generations or fitness value limit).

Several multiobjective GA-based approaches have been proposed in the literature [32], such as Strength Pareto Evolutionary Algorithm-II [33], Pareto Archieved Evolution Strategy [34], and Non-Dominated Sorting Genetic Algorithm-II (NSGA-II) [35]. In this paper, we will adopt the nondominated sorting solution (NSGA-II) for its low computational requirements, its aptitude to distribute uniformly the optimal solutions along the Pareto front [35], and its successful application to different RS problems [36]–[38]. NSGA-II is based on the concept of nondominance. A solution s_1 is said to dominate another solution s_2 if s_1 is not worse than s_2 in all objectives and better than s_2 in at least one objective. A solution is said to be nondominated if it is not dominated by any other solution. Fig. 2 shows a multiobjective GA chromosome selection problem with two fitness functions f_1 and f_2. As GA does, NSGA-II starts by generating a random parent population. Individuals (chromosomes) selected through a crowded tournament selection undergo crossover and mutation operations to form an offspring population. Both offspring population and parent population are then combined and sorted into fronts of decreasing dominance (rank). After the sorting process, the new population is filled with solutions of different fronts starting from the best one. If a front can only partially fill the next generation, crowded tournament selection is applied again to ensure diversity. Once the next generation population has been completely filled, the algorithm loops back to create a new offspring population, and the process continues up to convergence.

B. GA Setup

The design of a multiobjective GA optimization relies upon two components, namely, the chromosome structures and the fitness functions, which encode the considered optimization problem and show the direction to obtain the best solution, respectively.

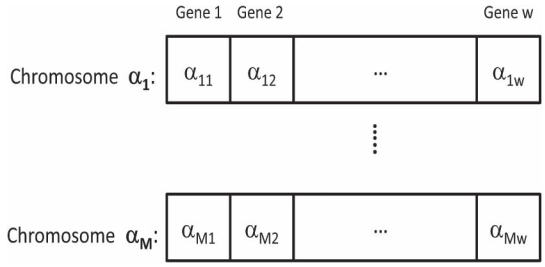


Fig. 3. Adopted chromosome structure.

Concerning the first component, we will consider a population of M chromosomes $\alpha_m \in R, m \in \{1, 2, \ldots, M\}$, where each chromosome is a real vector composed of genes corresponding to the weight vector $\boldsymbol{\alpha}$ defined earlier in the previous sections (see Fig. 3). The length w of the chromosome is thus equal to the one of the dictionary D. Chromosomes can be randomly initialized, or to obtain a faster and better convergence, it could be envisioned to add *a priori* information coming from more simple CS techniques, i.e., the OMP and BP algorithms. In other words, one could exploit the OMP and BP solutions to generate an initial population by perturbing these solutions.

Regarding the fitness function, we will investigate separately and jointly two fitness functions, i.e., those defining the optimization problem in (3). The first one aims at maximizing the sparsity level by minimizing the L_0 norm of the weight vector $\boldsymbol{\alpha}$, which corresponds to minimizing the number of almost nonzero coefficients in $\boldsymbol{\alpha}$

$$f_1 = \min \|\boldsymbol{\alpha}\|_0. \tag{7}$$

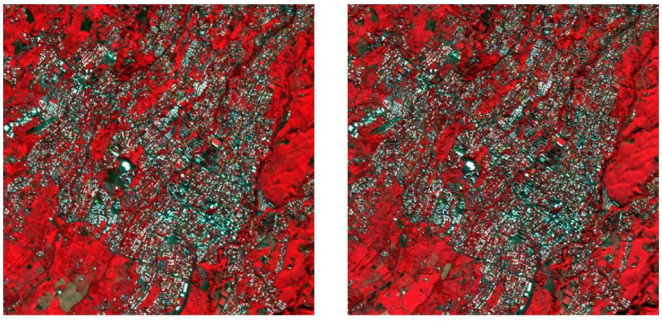
An almost nonzero coefficient is a coefficient exhibiting a value less than a predefined small threshold value (th). The second fitness function is derived from the constraint in (3). It points to a perfect reconstruction of the considered pixel (at position $[k, l]$). In other words, it is expressed as

$$f_2 = \min \|D\boldsymbol{\alpha} - x\|^2. \tag{8}$$

NSGA-II returns several optimal (nondominated) solutions along the Pareto front. Since a single solution has to be selected from the nondominated set, different strategies can be adopted. In this study, we suggest to choose the median solution as typically performed in the literature (see the last step in Fig. 2). In such a way, we expect to get a compromise solution with respect to what could be obtained by OMP and BP, i.e., a tradeoff between reconstruction model sparseness and reconstruction error.

(a) (b)

Fig. 4. Data set 1: FORMOSAT-2 images acquired in the Arcachon area on (a) June 24 and (b) July 16, 2009.



(a) (b)

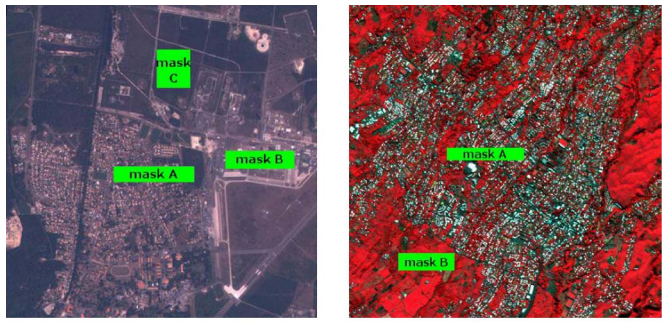
Fig. 5. Data set 2: SPOT-5 images acquired in the Reunion island on (a) May 2 and (b) June 18, 2008.

V. Experimental Results

A. Data Set Description and Setup

The first data set comes from the Taiwanese optical high-resolution FORMOSAT-2 satellite, which permits acquiring repeat imagery of an area of interest every day, from the same angle and under the same light conditions [39]. These images represent part of the *Arcachon* basin in the south region of Aquitaine, in France. The images are composed of 400×400 pixels and four spectral bands (blue, green, red, and near infrared) with a pixel spacing of 8 m. They were acquired on June 24 and July 16, 2009, respectively (see Fig. 4). The second data set comes from the SPOT-5 French satellite, whose images represent part of the *Reunion* island [40]. The images are characterized by a size of 450×450 pixels, four spectral bands (blue, green, red, and near infrared), and a pixel spacing of 10 m and were taken on May 2 and June 18, 2008, respectively (see Fig. 5). The two data sets present thus several differences among which are the sensor type, the pixel spacing, and the land covers. Indeed, the first one presents more vegetation areas than urban areas, while the second one exhibits an opposite composition. For the purpose of comparison, we implemented two other methods developed to reconstruct cloudy areas in RS images. One consists in a recent work exploiting a multiresolution inpainting (MRI) [2], whereas the second method estimates a missing pixel by contextual multiple linear prediction (CMLP) [5].

In order to make it possible to quantify the reconstruction accuracy of the methods, the experimental procedure adopted consisted of the following: 1) to consider a cloud-free image, e.g., $I^{(1)}$; 2) to simulate the presence of clouds by partly obscuring the other image, e.g., $I^{(2)}$; and 3) to compare the reconstructed image with the original cloud-free image. The simulation study is aimed at understanding the sensitivity of the five investigated methods (i.e., the MRI, CMLP, OMP, BP, and GA methods) to two aspects, which are the following: 1) the kind of ground covers obscured and 2) the size of the contaminated area. In order to obtain a detailed assessment of the reconstruction quality, we adopt the well-known peak signal-to-noise ratio (PSNR) measure [41]. 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 $\mathbf{\Phi}$. For the GA setup, the parameters adopted by our experiment were as follows: chromosome number $M = 50$, threshold value $\mathrm{th} = 10^{-4}$, probability of crossover $P_c = 0.8$, and probability of mutation $P_m = 0.005$.



(a) (b)

Fig. 6. Masks adopted to simulate the contamination of different ground covers.

B. Results

1) Contamination of Different Ground Covers: Fig. 6 shows different masks whose positions were selected in a way to simulate the obscuration of different kinds of ground cover. In particular, for data set 1, Fig. 6(a) shows mask A covering a region that includes mainly an urban area, mask B obscuring an industrial zone, and mask C covering a vegetation area. For data set 2, Fig. 6(b) shows mask A covering mainly a rural area and mask B covering a vegetation region. The experiments were carried out by considering each mask at a time, where each mask is composed of around 2000 pixels and the dictionary is composed of 300 pixels.

Tables I and II report for the two data sets 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 a vegetation region. This is the case for mask C in data set 1 and mask B in data set 2. The 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 our implementations, the OMP algorithm produces very sparse reconstruction

TABLE I
QUANTITATIVE RESULTS OBTAINED IN THE FIRST SIMULATION EXPERIMENTS FOR DATA SET 1

Method	Mask A				Mask B				Mask C			
	PSNR		Complexity	time [s]	PSNR		Complexity	time [s]	PSNR		Complexity	time [s]
	I_1	I_2			I_1	I_2			I_1	I_2		
MRI	-	22.54	-	2856	-	16.05	-	2517	-	33.77	-	2898
CMLP	-	20.99	1	1	-	20.11	1	1	-	24.05	1	1
OMP	39.41	**23.96**	3	4	36.33	20.60	3	4	44.28	31.97	3	4
BP	80.59	22.22	294	66	77.10	**24.74**	168	59	98.53	30.67	301	60
GA	42.09	23.78	148	68621	43.38	23.15	95	26312	45.62	**32.01**	138	43193

TABLE II
QUANTITATIVE RESULTS OBTAINED IN THE FIRST SIMULATION EXPERIMENTS FOR DATA SET 2

Method	Mask A				Mask B			
	PSNR		Complexity	time [s]	PSNR		Complexity	time [s]
	I_1	I_2			I_1	I_2		
MRI	-	24.27	-	2995	-	29.54	-	3614
CMLP	-	24.61	1	1	-	27.69	1	1
OMP	46.53	26.36	3	5	54.49	30.43	3	5
BP	86.22	26.45	338	61	99.62	**31.63**	365	91
GA	50.70	**26.72**	173	69231	56.30	31.28	201	38475

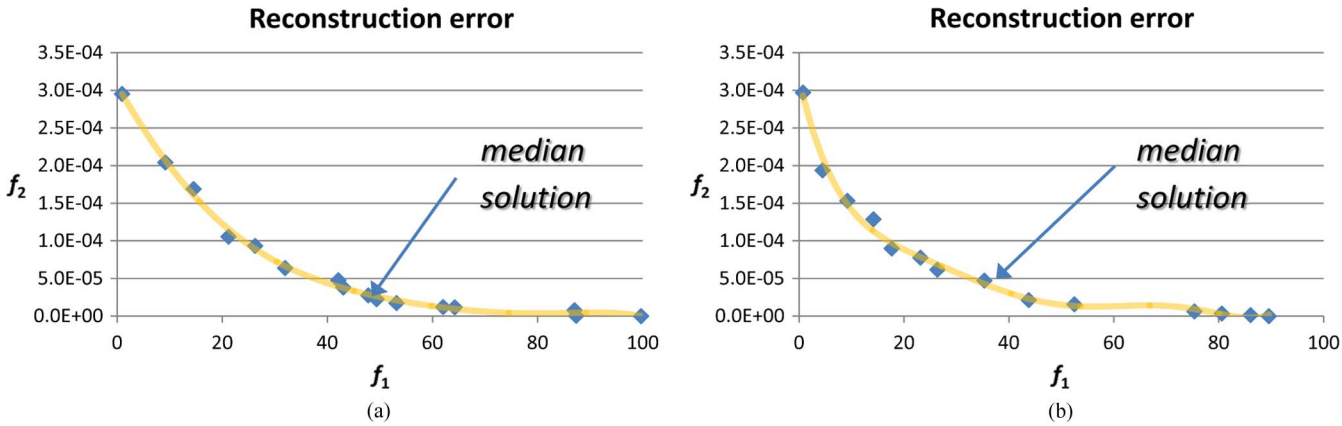
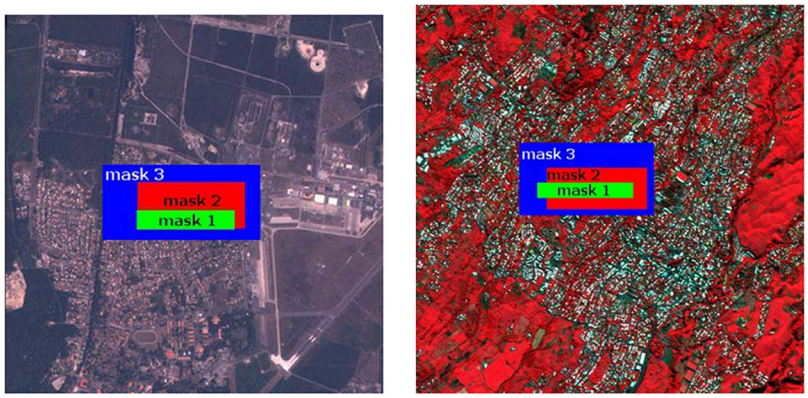


Fig. 7. Pareto fronts obtained at convergence for (a) data set 1 and (b) data set 2. Note that f_1 is represented in percentage.

solutions (around three 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, the 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. Finally, 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 than OMP (see the model complexity columns in Tables I and II). Its reconstruction error is almost all the time the best or the second best. Fig. 7 shows the Pareto fronts obtained at convergence for mask A in both data sets and the corresponding median solutions. In general, it can be observed that satisfactory results were achieved with the CS solutions despite the complexity of the faced problem.

2) Contamination With Different Sizes: Another important test for the five methods consists of assessing their performances by varying the amount of missing data. Fig. 8 shows the three different masks adopted to simulate different increasing cloud cover sizes. In particular, mask 1 is the same as 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 three and six, obtaining masks covering around 6000 and 12 000 pixels, respectively. Also, in these experiments, the adopted dictionaries are composed of 300 pixels belonging to the Φ region. Tables III and IV report for the two data sets the results achieved by the different reconstruction techniques and by varying the amount of missing data.



(a) (b)

Fig. 8. Masks adopted to simulate the different sizes of contamination.

TABLE III
QUANTITATIVE RESULTS OBTAINED IN THE SECOND SIMULATION EXPERIMENTS FOR DATA SET 1

Method	Mask 1				Mask 2				Mask 3			
	PSNR		Complexity	time [s]	PSNR		Complexity	time [s]	PSNR		Complexity	time [s]
	I_1	I_2			I_1	I_2			I_1	I_2		
MRI	-	22.54	-	2856	-	21.35	-	6938	-	19.63	-	14774
CMLP	-	20.99	1	1	-	21.13	1	1	-	20.83	1	2
OMP	39.41	**23.96**	3	4	42.45	23.21	7	6	42.00	**25.01**	3	19
BP	80.59	22.22	294	66	80.18	22.89	277	145	79.53	21.47	265	865
GA	42.09	23.78	148	68621	45.46	**23.85**	140	99072	45.13	23.03	149	275394

TABLE IV
QUANTITATIVE RESULTS OBTAINED IN THE SECOND SIMULATION EXPERIMENTS FOR DATA SET 2

Method	Mask 1				Mask 2				Mask 3			
	PSNR		Complexity	time [s]	PSNR		Complexity	time [s]	PSNR		Complexity	time [s]
	I_1	I_2			I_1	I_2			I_1	I_2		
MRI	-	24.27	-	2995	-	22.85	-	10176	-	23.82	-	22353
CMLP	-	24.61	1	1	-	24.43	1	2	-	25.46	1	2
OMP	46.53	26.36	3	5	46.89	26.42	3	16	47.49	27.39	3	21
BP	86.22	26.45	338	61	87.49	26.82	332	143	86.60	**28.25**	329	972
GA	50.70	**26.72**	173	69231	50.63	**27.10**	168	103342	51.14	28.15	170	259459

From a quantitative viewpoint, in terms of PSNR, we have results similar to those 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, our implementations return better results in terms of PSNR in all the simulations and present the advantage of not depending on the size of the missing area. The best solution in these experiments in terms of PSNR comes from GA, which outperforms all other methods in three cases, and in the other three, it is the second best choice. About the computation time, as expected, it increases as the amount of missing data increases. The 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.

Figs. 9 and 10 show the qualitative reconstruction results in red, green, and blue composites obtained in the most critical reconstruction scenario, i.e., the largest simulated cloud mask 3, for both data sets and for all reconstruction methods. As mentioned before, MRI reconstruction exhibits the worst reconstruction case. The CMLP method is capable of obtaining a good reconstruction compared with MRI. Regarding the CS reconstruction techniques (OMP, BP, and GA), good reconstructions are obtained, particularly for data set 2, where it is not simple to find significant differences comparing the reconstructions with the original (cloud-free) images.

3) Reconstruction Impact on Image Classification: Since the generation of classification maps represents one of the most widespread applications of RS images, it was also worth to evaluate the quality of the reconstruction process in terms of classification error. The latter was computed first by generating a classification map of the original images that

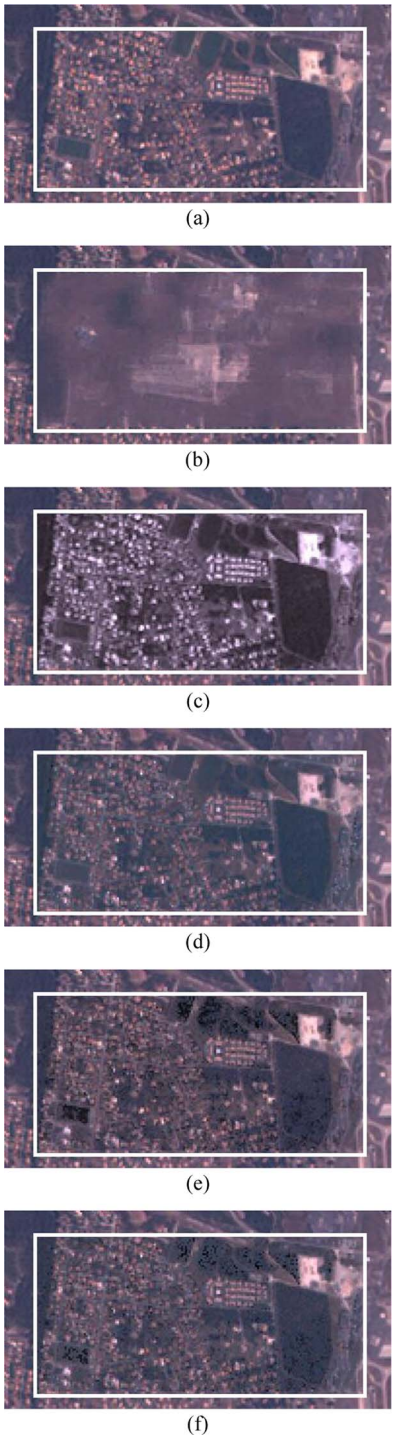


Fig. 9. Data set 1. (a) Color composite (bands 1, 2, and 3) of the original image. Color composites (bands 1, 2, and 3) of the same image reconstructed after the contamination with the largest simulated mask 3 by the (b) MRI, (c) CMLP, (d) OMP, (e) BP, and (f) GA methods.

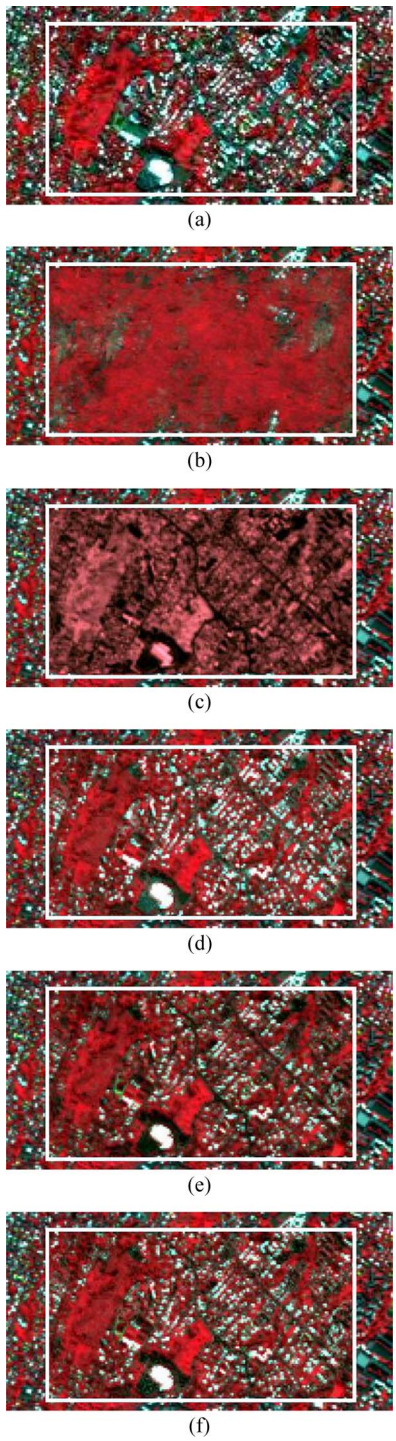


Fig. 10. Data set 2. (a) Color composite (bands 1, 2, and 3) of the original image. Color composites (bands 1, 2, and 3) of the same image reconstructed after the contamination with the largest simulated mask 3 by the (b) MRI, (c) CMLP, (d) OMP, (e) BP, and (f) GA methods.

served as reference classification maps by means of the popular k-means classifier. Then, each reconstructed image was given as input to the k-means classifier to provide a reconstruction classification map. For each reconstruction method, it was thus possible to evaluate the overall number of classification errors (OE) inside the reconstructed cloud-contaminated area by a simple comparison of both the reconstruction and the original classification maps. We repeated this exercise with different numbers of clusters (from three to seven clusters). The results confirm what was previously observed, i.e., the CS methods behave better than the traditional ones. We show in Fig. 11 the clustering results (with $k = 5$) obtained for the FORMOSAT-2 image with mask A as an example. Fig. 12 shows in more detail the differences between the various methods at the level of the reconstructed area (i.e., the mask). The best classification is achieved from the reconstruction with OMP (OE $= 6.3\%$), followed by GA (OE $= 8.7\%$), BP (OE $= 19.6\%$), MRI (OE $= 25.2\%$), and CMLP (OE $= 29.4\%$).

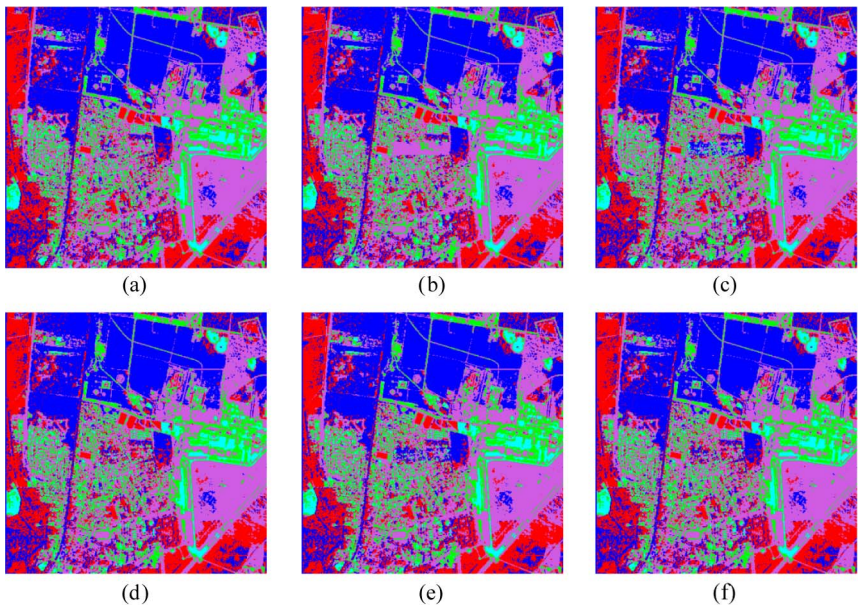


Fig. 11. (a) Unsupervised classification map obtained by the k-means algorithm $(k = 5)$ for the original FORMOSAT-2 image. Unsupervised classification maps obtained by the k-means algorithm $(k = 5)$ for the same image reconstructed after contamination with mask A by the (b) MRI, (c) CMLP, (d) OMP, (e) BP, and (f) GA methods.

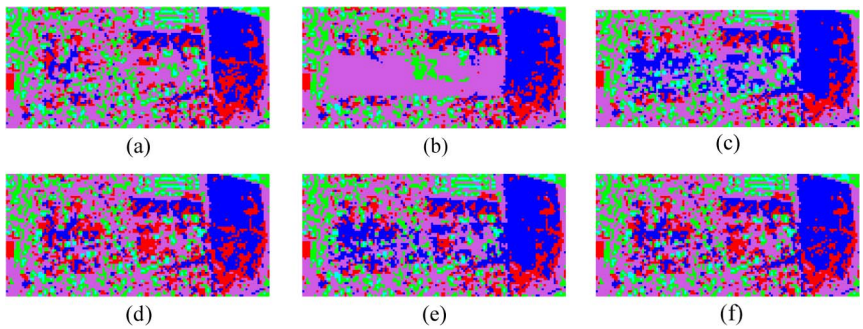


Fig. 12. Zoom of Fig. 11 on the mask-A area. (a) Original classification. Classification after reconstruction with the (b) MRI, (c) CMLP, (d) OMP, (e) BP, and (f) GA methods.

VI. Conclusion

This paper deals with the complex and important problem of removal of clouds from multispectral images. In particular, three novel methods have been proposed, which, given a contaminated image and a cloud-free image, allow reconstructing the missing measurements through the CS approach. First, we have shown how two common CS solutions, namely, the OMP and BP algorithms, can be formulated for a cloud-contaminated-image reconstruction problem. Then, we have proposed a solution for solving the CS problem under an L_0-norm perspective, exploiting the capabilities of GAs. The main properties of the proposed methods are as follows: 1) They rely on the assumptions that spectral nonstationarity is allowed while the spatial structure of the image should be almost identical between the two images; 2) they are not ground cover dependent; 3) they are unsupervised; and 4) different from CMLP, they need just one reference cloud-free image, and the reconstruction of each pixel is performed in all spectral bands simultaneously.

The experimental results point out the superiority of the proposed methods compared to two reference methods for cloud removal. Comparing our proposed methods, all three exhibit good results in the reconstruction of missing areas. OMP has the advantage of being sparser and significantly faster than BP and GA, but it is the less robust method. Indeed, since the reconstruction of each pixel depends typically on three coefficients and, thus, three other pixels of the image, it is enough that one of them is missing (covered by a cloud) to render the reconstruction model inaccurate. This problem is much less important to BP as it is much less sparse than OMP. GA represents a good compromise between the OMP and BP methods, mainly because it is more robust than OMP and more sparse than BP. Another empirical conclusion is that the kind of ground cover obscured may be an important factor to take into consideration for the reconstruction, while the size of the contaminated area only marginally affects the performance of the proposed reconstruction methods, which depends more on the information available outside the missing area. In other words, if a ground cover is contaminated and it is not represented outside of the contaminated area, the reconstruction process will not deal suitable with such a situation.

Finally, in order to improve the accuracy of the reconstruction process, different aspects of the methods deserve to be investigated in future research studies. For example, in almost all the reconstructed images, we note an unwanted presence of a slight "curtain effect" because of the presence of an estimation bias. A possible solution to this problem could be a cluster-oriented sparse reconstruction. Note that, in general, it is difficult to obtain pairs of images where one of them is completely cloud free. By adopting the implemented CS solutions, it will be possible to exploit and reconstruct two cloud-contaminated images only if they do not convey clouds in the same region.

ACKNOWLEDGMENT

The authors would like to thank CNES for making available online the FORMOSAT-2 and SPOT-5 images (http://kalideos.cnes.fr) and D. Donoho and Y. Tsaig for the SparseLab toolbox used in this research (http://sparselab.stanford.edu/).

REFERENCES

[1] S. C. Liew, M. Li, and L. K. Kwoh, "Automated production of cloud-free and cloud-shadow image mosaics from cloudy satellite imagery," in *Proc. 20th ISPRS Congr.*, Istanbul, Turkey, Jul. 12–13, 2004, pp. 523–530.

[2] L. Lorenzi, F. Melgani, and G. Mercier, "Inpainting strategies for reconstruction of missing data in VHR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 5, pp. 914–918, Sep. 2011.

[3] A. Maalouf, P. Carre, B. Augereau, and C. Fernandez-Maloigne, "A bandelet-based inpainting technique for clouds removal from remotely sensed images," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2363–2371, Jul. 2009.

[4] S. Lee and M. M. Crawford, "An adaptive reconstruction system for spatially correlated multispectral multitemporal images," *IEEE Trans. Geosci. Remote Sens.*, vol. 29, no. 4, pp. 494–508, Jul. 1991.

[5] F. Melgani, "Contextual reconstruction of cloud-contaminated multitemporal multispectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 2, pp. 442–455, Feb. 2006.

[6] D.-C. Tseng, H.-T. Tseng, and C.-L. Chien, "Automatic cloud removal from multi-temporal SPOT images," *Appl. Math. Comput.*, vol. 205, no. 2, pp. 584–600, Nov. 2008.

[7] C.-H. Lin, P.-H. Tsai, K.-H. Lai, and J.-Y. Chen, "Cloud removal from multitemporal satellite images using information cloning," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 232–241, Jan. 2013.

[8] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[9] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principle—Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[10] L. Jiying, Z. Jubo, Y. Fengxia, and Z. Zenghui, "Theoretical frameworks of remote sensing systems based on compressive sensing," *Int. Archives Photogramm. Remote Sens.*, vol. 37, pp. 77–81, Jul. 2010.

[11] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[12] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[13] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution via sparse representations," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.

[14] S. Rao, R. Tron, R. Vidal, and Y. Ma, "Motion segmentation via robust subspace separation in the presence of outlying, incomplete, corrupted trajectories," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 10, pp. 1832–1845, Oct. 2010.

[15] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *IEEE Trans. Image Process.*, vol. 17, no. 1, pp. 53–69, Jan. 2008.

[16] B. Shen, W. Hu, Y. Zhang, and Y.-J. Zhang, "Image inpainting via sparse representation," in *Proc. IEEE ICASSP*, 2009, pp. 697–700.

[17] A. Quattoni, M. Collins, and T. Darrell, "Transfer learning for image classification with sparse prototype representation," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[18] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *Proc. IEEE*, vol. 98, no. 6, pp. 1031–1044, Jun. 2010.

[19] S. Li and B. Yang, "A new pan-sharpening method using a compressed sensing technique," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 738–746, Feb. 2011.

[20] Y. Chen, N. M Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, Oct. 2011.

[21] S. Yang, F. Sun, M. Wang, Z. Liu, and L. Jiao, "Novel super resolution restoration of remote sensing images based on compressive sensing and example patches-aided dictionary learning," in *Proc. Int. Workshop Multi-Platform/Multi-Sens. Remote Sens. Mapping*, Jan. 2011, pp. 1–6.

[22] J. Ma, "Single-pixel remote sensing," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 2, pp. 199–203, Apr. 2009.

[23] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[24] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, Aug. 1998.

[25] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modelling of signals and images," *SIAM Rev.*, vol. 51, no. 1, pp. 34–81, Feb. 2009.

[26] N. Wang and Y. Wang, "An image reconstruction algorithm based on compressive sensing using conjugate gradient," in *Proc. IUCS*, Oct. 2010, pp. 374–377.

[27] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decompositions," in *Proc. 27th Asilomar Conf. Signals., Syst. Comput.*, 1993, pp. 40–44.

[28] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[29] S. Kunis and H. Rauhut, "Random sampling of sparse trigonometric polynomials, II. Orthogonal matching pursuit versus basis pursuit," *Found. Comput. Math.*, vol. 8, no. 6, pp. 737–763, Nov. 2008.

[30] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*. Reading, MA: Addison-Wesley, 1989.

[31] L. Chambers, *The Practical Handbook of Genetic Algorithms*. New York: Chapman & Hall, 2001.

[32] K. Deb, *Multi-Objective Optimization Using Evolutionary Algorithms*. Chichester, U.K.: Wiley, 2001.

[33] E. Zitzler, M. Laumanns, and L. Thiele, "SPEA2: Improving the strength Pareto evolutionary algorithm," *Evolutionary Methods for Design, Optimization and Control With Applications to Industrial Problems*, no. 103, pp. 95–100, 2002.

[34] J. Knowles and D. Corne, "The Pareto archived evolution strategy: A new baseline algorithm for Pareto multiobjective optimisation," in *Proc. CEC*, 1999, vol. 1, pp. 98–105.

[35] N. Srinivas and K. Deb, "Multiobjective function optimization using nondominated sorting genetic algorithms," *Evol. Comput.*, vol. 2, no. 3, pp. 221–248, 1995.

[36] N. Ghoggali, F. Melgani, and Y. Bazi, "A multiobjective genetic SVM approach for classification problems with limited training samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 6, pp. 1707–1718, Jun. 2009.

[37] N. Ghoggali and F. Melgani, "Genetic SVM approach to semisupervised multitemporal classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 212–216, Apr. 2008.

[38] E. Pasolli, F. Melgani, and M. Donelli, "Automatic analysis of GPR images: A pattern-recognition approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2206–2217, Jul. 2009.

[39] C.-C. Liu, "Processing of FORMOSAT-2 daily revisit imagery for site surveillance," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 11, pp. 3206–3214, Nov. 2006.

[40] A. Baudoin, "Mission analysis for SPOT 5," in *Proc. IEEE IGARSS*, Aug. 1993, vol. 3, p. 1084.

[41] A. K. Jain, *Fundamentals of Digital Image Processing*. New York: Prentice-Hall, 1988.

Luca Lorenzi (S'08) received the M.Sc. degree in telecommunications engineering from the University of Trento, Trento, Italy, in 2008, where he is currently working toward the Ph.D. degree in information and communication technologies.

He is currently with the Signal Processing and Recognition Laboratory, Department of Information Engineering and Computer Science, University of Trento. He is also with the Image and Information Processing Department (ITI), Telecom Bretagne, Institut Telecom, Brest, France. His research interests include processing and recognition techniques applied to remote sensing images (classification, regression, multitemporal analysis, and machine learning).

Farid Melgani (M'04–SM'06) received the State Engineering degree in electronics from the University of Batna, Batna, Algeria, in 1994, the M.Sc. degree in electrical engineering from the University of Baghdad, Baghdad, Iraq, in 1999, and the Ph.D. degree in electronic and computer engineering from the University of Genoa, Genoa, Italy, in 2003.

From 1999 to 2002, he cooperated with the Signal Processing and Telecommunications Group, Department of Biophysical and Electronic Engineering (DIBE), University of Genoa. Since 2002, he has been with the University of Trento, Trento, Italy, where he was an Assistant Professor and is currently an Associate Professor of telecommunications and where he has taught pattern recognition, machine learning, radar remote sensing (RS) systems, and digital transmission. He is also the Head of the Signal Processing and Recognition Laboratory, Department of Information Engineering and Computer Science, University of Trento. He is a coauthor of more than 130 scientific publications and is a Referee for several international journals. His research interests are in the area of processing, pattern recognition, and machine learning techniques applied to RS and biomedical signals/images (classification, regression, multitemporal analysis, and data fusion).

Dr. Melgani has served on the scientific committees of several international conferences and is an Associate Editor of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS.

Grégoire Mercier (M'02–SM'07) was born in France in 1971. He received the Dipl. Ing. degree from the Institut National des Télécommunications, Évry, France, in 1993 and the Ph.D. degree and the Habilitation à Diriger des Recherches from the University of Rennes 1, Rennes, France, in 1999 and 2007, respectively.

Since 1999, he has been with Telecom Bretagne, Institut Telecom, Brest, France, where he was an Associate Professor with the Image and Information Processing Department (ITI) and has been a Professor since 2010. From March to May 2006, he was a Visiting Researcher with the Department of Biophysical and Electronic Engineering (DIBE), University of Genoa, Genoa, Italy, where he developed a change detection technique for heterogeneous data. From April to June 2007, he was also a Visiting Researcher with the Centre National d'Etudes Spatiales, Toulouse, France, to take part of the Orfeo Toolbox development. His research interests are in remote sensing image compression and segmentation, particularly in hyperspectral and synthetic aperture radar. His research is dedicated to change detection and combating pollution.

Dr. Mercier is an Associate Editor of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. He is the President of the French Chapter of the IEEE Geoscience and Remote Sensing Society.