

A Hyperheuristic Approach for Unsupervised Land-Cover Classification

Joao Papa Papa, Luciene Patrici Papa, Danilo Roberto Pereira, and Rodrigo Jose Pisani

Abstract—Unsupervised land-use/cover classification is of great interest, since it becomes even more difficult to obtain high-quality labeled data. Still considered one of the most used clustering techniques, the well-known k -means plays an important role in the pattern recognition community. Its simple formulation and good results in a number of applications have fostered the development of new variants and methodologies to address the problem of minimizing the distance from each dataset sample to its nearest centroid (mean). In this paper, we present a genetic programming-based hyperheuristic approach to combine different metaheuristic techniques used to enhance k -means effectiveness. The proposed approach is evaluated in four satellite and one radar image showing promising results, while outperforming each individual metaheuristic technique.

Index Terms—Hyperheuristic, k -means, land-cover classification.

I. INTRODUCTION

IN the last decades, advances in remote sensing image acquisition systems have moved in lockstep with the need for applications that make use of such sort of data. Land-use/cover recognition [1]–[4], target recognition [5]–[7], image classification [8], [9], and band selection in hyperspectral images [10], [11] are among the most pursued applications, just to name a few. The large amount of high-resolution content available by satellites also highlights the bottleneck that takes place when labeling data. Such process is skilled dependent, and it might be very prone to errors when dealing with manual annotation. Such shortcomings have fostered even more the research on semisupervised and unsupervised techniques, which may work well in some remote-sensing-oriented applications.

Considered a hallmark in the pattern recognition research field, the so-called k -means algorithm [12] has been consistently enhanced in the last decades. Given it does not make use of labeled data and it has a simple formulation, k -means is still one of the most used classification techniques up to date.

Manuscript received September 26, 2015; revised March 18, 2016; accepted April 08, 2016. Date of publication May 17, 2016; date of current version July 05, 2016. This work was supported in part by São Paulo Research Foundation (FAPESP) Grants #2013/20387-7 and #2014/16250-9, and in part by National Council for Scientific and Technological Development (CNPq) Grants #470571/2013-6, #487032/2012-8 and #306166/2014-3. (Corresponding author: Joao Papa Papa.)

J. P. Papa and D. R. Pereira are with the Department of Computing, São Paulo State University, Bauru, SP, CEP 17033-360 Brazil (e-mail: papa@fc.unesp.br; dpereira@ic.unicamp.br).

L. P. Papa is with the Department of Health, São Paulo State South-west College, Av. Prof. Clso Ferreira da Silva, 1001 18707-150 (e-mail: lucienepapa@yahoo.com.br).

R. J. Pisani is with the Institute of Natural Sciences, Federal University of Alfenas, Alfenas 37130-000 Brazil (e-mail: rodrigo.pisani@unifal-mg.edu.br).

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/JSTARS.2016.2557584

Roughly speaking, given a set of feature vectors (samples) extracted from a dataset, k -means tries to minimize the distance from each sample to its closest center (mean). Such process ends up clustering the data after some steps, being two samples from the same cluster more “connected” to its centroid than to any other in the dataset. As its main drawbacks, we can shed light the number of clusters required as an input, and the leaning of the naïve algorithm, which can get trapped from local optima, i.e., centroids that do not represent well the clusters.

The aforementioned scenario turns k -means algorithm more prone to be addressed by means of optimization techniques, mainly those based on nature- and evolutionary-oriented mechanisms. Actually, not only k -means but a number of other techniques have used the framework of metaheuristic-based optimization to cope with problems that somehow can be modeled as a task of finding decision variables that maximize/minimize some certain fitness function. Chen *et al.* [13], for instance, employed genetic algorithms (GAs) and neural networks to classify both land-use and landslide zones in eastern Taiwan, being the former used to compute the set of weights that combine some landslide incidence factors. Nakamura *et al.* [14] dealt with the task of band selection in hyperspectral imagery through nature-inspired techniques. Truly speaking, the idea is to model the problem of finding the most important bands as a feature selection task. Without loss of generality, both problems are the very same one when the brightness of each pixel is used to represent it.

Very recently, Goel *et al.* [15] tackled the problem of remote sensing image classification using some nature-inspired techniques, say that cuckoo search (CS) and artificial bee colony. Senthilnatha *et al.* [16] used GAs, particle swarm optimization (PSO), and firefly algorithm for the automatic image registering of multitemporal remote sensing data. In short, the idea is to perform image registration while minimizing some criterion function (mutual information in that case). The theory about artificial immune systems has been used to classify remote sensing data as well [17], in which a multiband image covering the area of northeastern part of Algiers was used for validation purposes.

Coming back to the k -means technique, Chandran and Nazeer [18] proposed to solve the problem of minimizing the distance from each dataset sample to its nearest centroid using the harmony search (HS), which is a metaheuristic optimization technique based on the way musicians create songs in order to obtain the best harmony. Forsati *et al.* [19] employed a similar approach, but in the context of web page clustering, while Lin *et al.* [20] proposed a hybrid approach concerning the task of k -means clustering and PSO. Later on, Kuo *et al.* [21] integrated k -means and artificial immune systems for dataset clustering, and

Saida *et al.* [22] employed the CS to optimize k -means aiming at classifying documents. Finally, a comprehensive study about the application of nature-inspired techniques to boost k -means was presented by Fong *et al.* [23].

Despite all aforementioned works aimed at enhancing k -means using metaheuristic techniques, there is a little concern about the application of hyperheuristic techniques for that purpose, as well as only a very few works attempted at dealing with k -means optimization in the context of land-use/cover classification. The term “hyperheuristics” was coined to address new algorithms designed to solve general problems by combining known metaheuristics, in such a way each technique may compensate the weaknesses of others [24]. In such context, Papa *et al.* [25] were one of the first that focused on the application of hyperheuristics to optimize k -means, being the proposed approach validated in the background of both satellite- and radar-based land-cover classification.¹ That work employed genetic programming to combine five variations of the HS algorithm with promising results. In this paper, we extend the work by Papa *et al.* [25] with a deeper experimental analysis, in which PSO, bat algorithm (BA), and firefly algorithm (FFA) are also considered together with HS and its variants for combination purposes through genetic programming. The results obtained in this paper outperformed the previous work by Papa *et al.* [25], thus emphasizing the benefits of the hyper-heuristic-based framework. To the best of our knowledge, that is the first time such approaches are combined with each other aiming at optimizing k -means.

The remainder of this paper is organized as follows. Section II presents some related-work regarding metaheuristic optimization techniques employed for similar problems. Section III and IV present the proposed approach and the experimental setup, respectively. Section V states conclusions and future works.

II. RELATED WORKS

Recently, Fong *et al.* [23] presented an interesting study about the application of nature-inspired techniques aiming at optimizing k -means. Their work compared naïve k -means against with FFA, CS, ant colony optimization, and BA, being the results of the optimization techniques more accurate than standard k -means in two out three datasets. Hruschka *et al.* [26] presented a survey about evolutionary algorithms applied for clustering problems, where the techniques were divided in those that require a fixed number of clusters beforehand, and the others that can handle a variable number of clusters. The authors also stated most works consider the quality of the partitioning process only, and they do not take care about the computational efficiency, which can be a problem in large-scale problems.

Naldi *et al.* [27] evaluated some efficiency issues in k -means when optimized with evolutionary techniques, showing it can be faster than standard approaches to find out the number of clusters when one does not know that in advance. Later on, Naldi and Campello [28] evaluated an evolutionary-based k -means in the context of distributed data, which means the technique

must be able to cluster the data partially, and then, to merge the results when the whole data is available. Luchi *et al.* [29] also considered the problem of clustering large datasets with k -means by presenting a GA-based approach for sampling the initial clusters' centroids. A similar approach was presented by Bousbaci and Kamel [30] either, but now using the PSO.

Very recently, Serapião *et al.* [31] evaluated the fish school search in the context of k -means, which demonstrated good results. Saida *et al.* [32] employed the CS for the very same purpose, i.e., to optimize k -means. Since CS uses Lévy flights to avoid local optimum, it is usually a good alternative for several optimization-related problems. Roughly speaking, the idea in using metaheuristics to guide k -means optimization is just to avoid it to get trapped from local optimum.

Other interesting approach concerns with the hybridization of metaheuristics [33]–[35], which aims at combining different optimization techniques into a single one in order to exploit the weakness of each one. Multiobjective-driven clustering is also another interesting idea, since we can optimize different fitness functions that are somehow conflicting to each other [36], [37]. Usually, the fitness functions may comprise the number of clusters versus some efficiency measure, for instance. We can also work with feature selection and, at the same time, to choose a suitable number of clusters.

III. OPTIMIZING k -MEANS THROUGH HYPERHEURISTICS

Essentially, the problem of estimating the centroids of k clusters can be modeled as an optimization task, in which we aim at minimizing the distance from each dataset sample to its nearest centroid. Therefore, any fitness function that somehow encodes such behavior can be employed. In this study, we used two metrics: 1) the ‘average distance of data points to cluster centroids’ (ADDC) and 2) the ‘average distance of data points between cluster’ (ADDBC). The ADDC and ADDBC are formulated as follows:

$$\text{ADDC} = \frac{1}{N} \sum_{j=1}^k \sum_{\forall \mathbf{x}_i \in c_j} D(\mathbf{c}_j, \mathbf{x}_i), \quad (1)$$

$$\text{ADDBC} = \frac{1}{N} \frac{1}{k} \sum_{j=1}^k \sum_{\forall \mathbf{x}_i \in c_j} D(\mathbf{c}_j, \bar{\mathbf{x}}_i) \quad (2)$$

where $D(\mathbf{c}_j, \mathbf{x}_i)$ stands for the distance between centroid $\mathbf{c}_j \in \mathbb{R}^n$ and sample \mathbf{x}_i , and N denotes the number of dataset samples that contributed for the metric computation.

Roughly speaking, ADDC represents the average internal homogeneity of the clusters and ADDBC maps a good intercluster separation. It is reasonable to note that a satisfactory clustering process generates a high intracluster homogeneity, as well as a good intercluster separation. In our experiments, we used both ADDC and a combination of these two metrics, as follows:

$$\mathbb{M}_1 = \text{ADDC} \quad (3)$$

and

$$\mathbb{M}_2 = \frac{k \text{ ADDC}}{\text{ADDBC}}. \quad (4)$$

¹This study was presented at IGARSS'2015.

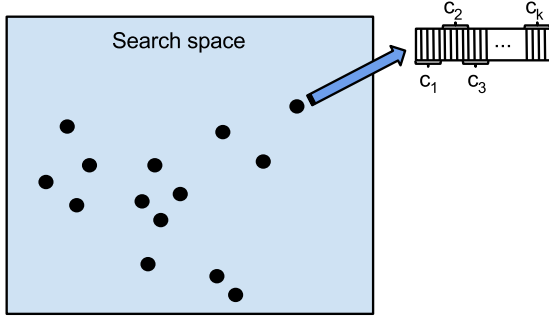


Fig. 1. Problem representation adopted in this study.

Given a problem with k clusters, each agent (e.g., harmonies, particles, bats, or fireflies) encodes a possible solution in $\mathbb{R}^{k \times n}$, as depicted in Fig. 1, where n stands for the number of features used to represent each dataset sample (pixels in our case). Although one has a continuous array, each cluster's centroid c_i stands for n positions of this data representation, which must be taken into account when feeding k -means algorithm.

Therefore, after placing all agents with random positions, the k -means algorithm is executed once for each agent using that positions as the starting point. Soon after, the metric M_1 (or M_2) is computed over the final clustered dataset to be used as the fitness function for each metaheuristic technique, which outputs the optimum/near-optimum possible solution with the best starting points for k -means.

The work proposed by Papa *et al.* [25] goes beyond that point by combining different solutions obtained through distinct metaheuristic techniques. Since each technique has its own weaknesses, the idea is to explore a higher level of optimization in order to improve each individual solution by means of the combination of all obtained solutions so far. Although such step can be performed by any optimization technique, we opted to employ genetic programming (GP) for the following two main reasons: 1) we did not use any metaheuristic technique that has been employed during the first step of optimization in order to avoid biases in this new procedure and 2) GP provides a more powerful combination process as a hyperheuristic technique, since it can apply a number of arithmetic operations for that purpose, instead of using movement-based equations to place agents from one position to another.

GP [38] is an evolutionary-based optimization algorithm that models each solution as an individual, which is usually represented as a tree composed of “function” and “terminal” nodes. The function nodes encode the arithmetic operators used over the terminal nodes in order to evaluate the trees [see Fig. 2(a)], and the terminal nodes represent constant values. At each iteration, specific operations over the current population are performed to design the next generation of individuals, being the most used ones: 1) mutation, 2) crossover, and 3) reproduction. Mutation and crossover aim at allowing a greater variability to the population of individuals, while reproduction tries to maintain the best ones to the next generation. In short, mutation operations change each individual without considering others, i.e., given a mutation point, we can simply generate a new random

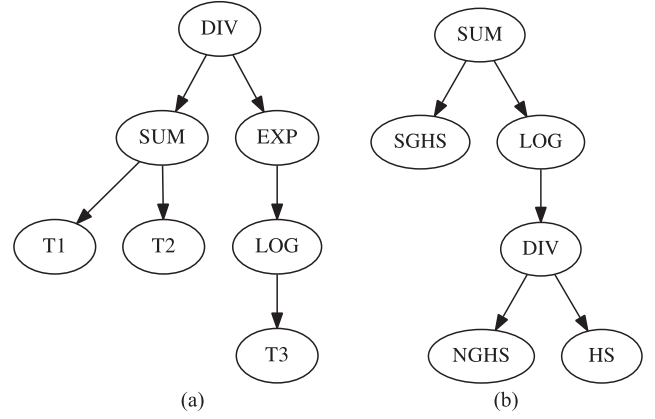


Fig. 2. GP trees: (a) example tree with terminal nodes “T1-T2-T3,” and (b) example tree extracted from the results obtained by Papa *et al.* [25]. The function node “SUM” stands for “summation,” “DIV” denotes the “division” operator, “EXP” stands for the “exponentiation,” and “LOG” corresponds to the “logarithm.”

subtree at that point, while crossover switch branches between two distinct trees.

The work by Papa *et al.* [25] employs the best result of each metaheuristic technique (i.e., HS, improved harmony search—IHS [39], global-best harmony search—GHS [40], novel global harmony search—NGHS [41], and self-adaptive global-best harmony search—SGHS [42]) to compose the set of terminal nodes [see Fig. 2(b)]. This means GP can use any technique from that set for combination purposes, and therefore, can decide which one will compose the best individual right after the evolutionary-oriented optimization process. In this study, we propose to employ not only HS-based metaheuristic techniques to compose the set of terminal nodes, but also other techniques, such as PSO [43], BA [44], and FFA [45]. The main assumption here concerns that exploring different mechanisms may allow us to obtain more accurate results, since we can count with more variable results. Such assumption has proven to be correct, since the results presented in this paper outperformed the ones discussed in the work by Papa *et al.* [25].

IV. METHODOLOGY AND EXPERIMENTS

In this section, we present the methodology employed to validate the proposed approach, which comprises the images, parameter setup, and experimental evaluation. Additionally, we also present the experimental results.

A. Methodology

1) *Image Dataset*: We used images obtained from CBERS-2B and Landsat 5 TM satellites covering the area of Itatinga, SP-Brazil, and images obtained from Ikonos-2 MS and Geosyde satellites covering the area of Duque de Caxias, RJ-Brazil. Additionally, we used two hyperspectral images obtained by AVIRIS sensor, say that Indian Pines over Indiana-USA, and Salinas-A over Salinas Valley, California-USA. Fig. 3 displays these

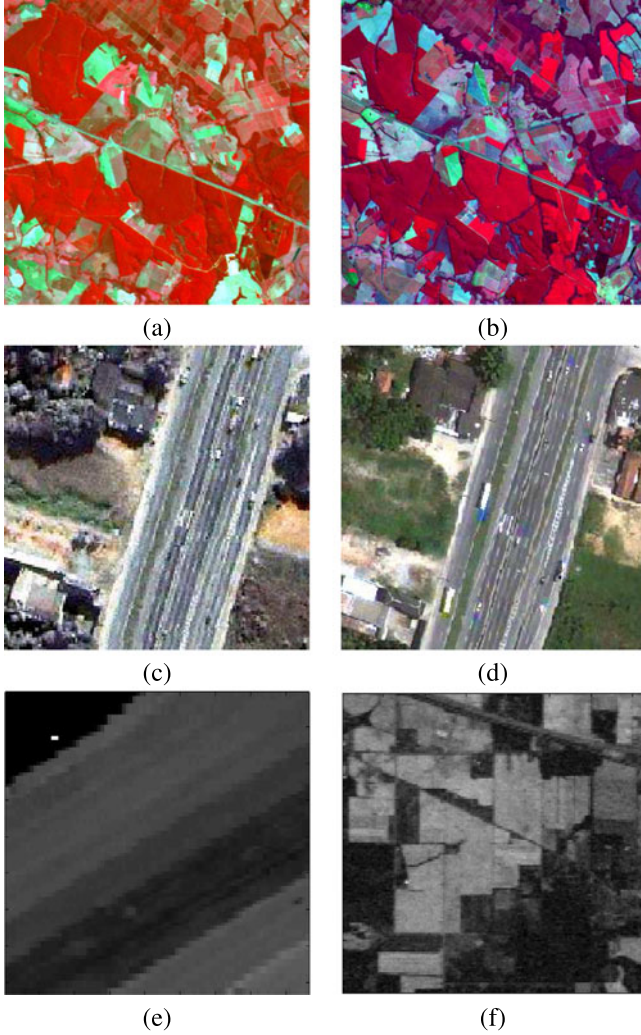


Fig. 3. Satellite images used in the experiments: covering the area of Itatinga, SP-Brazil by (a) CBERS-2B CCD (20 m) sensor (R2G3B4) and (b) Landsat 5 TM (30 m) sensor (R4G3B5), and covering the area of Duque de Caxias, RJ-Brazil by (c) Ikonos-2 MS sensor (R4G3B2), and (d) Geoeeye sensor (R5G4B3), (e) Salinas-A, and (f) Indian Pines. The CBERS-2B and Landsat 5 TM images have 526×492 pixels, Ikonos-2 MS and Geoeeye images have 258×250 and 268×250 pixels, respectively. Notice that Ikonos-2 MS and Geoeeye images were obtained through a fusion process between the corresponding images from MS (4 m) and PAN (1 m) sensors using the pan-sharpening method. The final image has a spatial resolution of 1m. Salinas-A image has 86×83 pixels and 204 bands, while Indian Pines image has 145×145 pixels and 200 bands.

images, being their respective ground-truth versions illustrated in Fig. 4.

2) *Parameter Setup*: Table I presents the parameters used for each metaheuristic technique, being such values empirically chosen. In regard to GP parameters, we employed the following configuration:

- 1) number of trees: 100;
- 2) number of iterations: 200;
- 3) tree generation: GROW [38] with max-depth equal to 5;
- 4) reproduction rate: 30%;
- 5) mutation rate: 30%;
- 6) crossover rate: 40%;

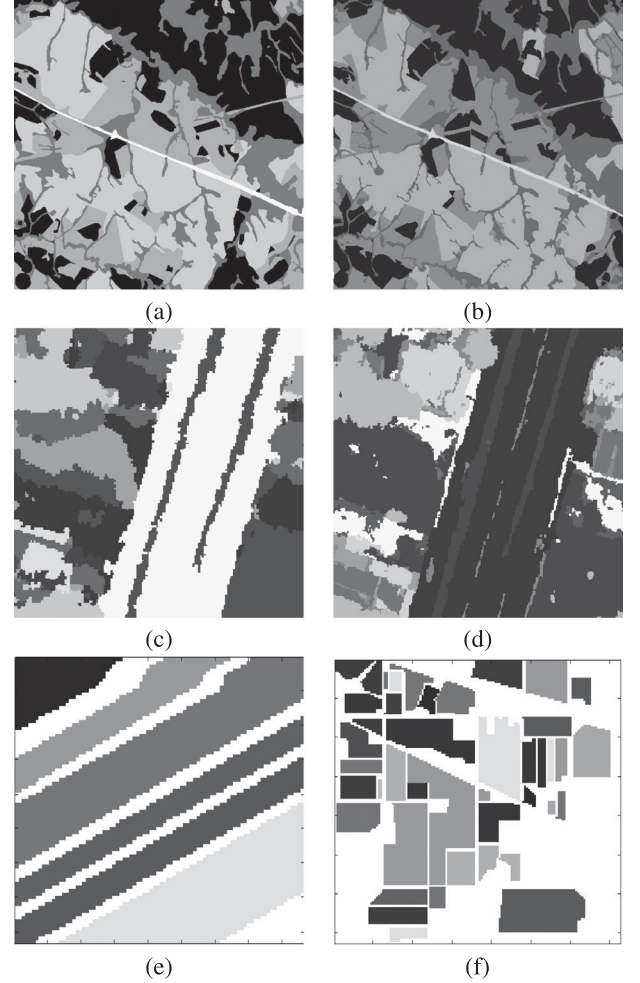


Fig. 4. Labeled images used in the experiments: (a) and (b) refer to the images displayed in Fig. 3(a) and (b), respectively, (c) and (d) stand for images displayed in Fig. 3(c) and (d), respectively, and (e) and (f) refer to the image displayed in Fig. 3(e) and (f).

TABLE I
PARAMETERS USED FOR THE METAHEURISTIC TECHNIQUES EMPLOYED IN THIS PAPER

Technique	Parameters
HS	HMCR= 0.9, PAR= 0.5, BW= 10
IHS	HMCR= 0.9, PAR _{min} = 0.1, PAR _{max} = 0.9 BW _{min} = 0.1, BW _{max} = 0.9
GHS	HMCR= 0.9, PAR= 0.5
NGHS	$p_m = 0.7$
SGHS	HMCR= 0.9, PAR= 0.5, $\sigma_{HMCR} = 0.01$ $\sigma_{HMCR} = 0.05$, $LP = 10.0$
PSO	$c_1 = c_2 = 2.0$, $w = 0.5$
BA	$q_{min} = 0.0$, $q_{max} = 1.0$, $\alpha = \gamma = 1.0$ $A^0 = 1.0$, $r^0 = 2$
FFA	$\psi = 0.5$, $\tau = \eta = 1.0$

- 7) function nodes: SUM (summation), DIV (division), MUL (multiplication), SUB (subtraction), and LOG (logarithm); and
- 8) terminal nodes: HS, IHS, GHS, NGHS, SGHS, PSO, BA, and FFA. We also used 1000 random generated numbers within the interval $[0, 0.1]$ to compose the terminal nodes.

TABLE II
AVERAGE \mathbb{M}_1 VALUE OVER THE TEST SET CONSIDERING EXPERIMENT I

	CBERS-2B	Geoeeye	Ikonos-2 MS	Indian	Landsat-5 TM	Salinas
<i>k</i> -means	30.43 \pm 9.66	17.13 \pm 3.11	21.20 \pm 2.42	2878.06 \pm 10.42	25.66 \pm 1.04	1488.65 \pm 7.54
BA	22.10 \pm 1.67	13.65 \pm 0.78	15.74 \pm 0.92	2452.91 \pm 4.44	22.90 \pm 0.89	1074.24 \pm 6.37
FFA	22.63 \pm 0.57	12.98 \pm 0.55	15.55 \pm 1.22	2428.90 \pm 7.78	23.13 \pm 1.07	1370.97 \pm 361.93
GHS	24.36 \pm 0.51	15.97 \pm 0.70	16.27 \pm 0.54	2422.62 \pm 5.11	25.78 \pm 2.24	1513.81 \pm 19.08
GP	22.13 \pm 1.51	13.40 \pm 1.54	15.23 \pm 1.18	2390.63 \pm 1.57	23.04 \pm 1.49	1051.89 \pm 11.18
HS	24.45 \pm 1.73	16.25 \pm 0.94	17.56 \pm 0.63	2381.45 \pm 4.36	26.26 \pm 0.78	1484.89 \pm 16.14
NGHS	27.59 \pm 2.83	17.09 \pm 2.77	18.40 \pm 1.10	2423.43 \pm 4.68	28.70 \pm 3.29	1304.78 \pm 8.34
PSO	31.33 \pm 15.85	13.74 \pm 2.52	15.16 \pm 1.27	2582.73 \pm 115.48	34.11 \pm 14.28	1565.91 \pm 348.16
SGHS	21.92 \pm 1.91	13.23 \pm 1.40	15.10 \pm 1.08	2378.88 \pm 5.32	23.27 \pm 0.77	1138.26 \pm 9.83

TABLE III
AVERAGE \mathbb{M}_2 VALUE OVER THE TEST SET CONSIDERING EXPERIMENT I

	CBERS-2B	Geoeeye	Ikonos-2 MS	Indian	Landsat-5 TM	Salinas
<i>k</i> -means	14.50 \pm 4.23	23.40 \pm 7.04	23.67 \pm 6.86	102.01 \pm 1.40	18.49 \pm 3.32	8.79 \pm 0.13
BA	10.81 \pm 1.26	20.89 \pm 11.31	27.08 \pm 17.25	91.32 \pm 1.69	14.54 \pm 1.11	8.01 \pm 2.93
FFA	10.39 \pm 0.74	17.93 \pm 1.57	15.98 \pm 1.19	85.59 \pm 3.47	14.83 \pm 0.88	9.15 \pm 1.71
GHS	10.76 \pm 1.14	15.97 \pm 1.40	16.18 \pm 0.84	88.54 \pm 4.10	15.43 \pm 1.56	8.51 \pm 1.61
GP	10.57 \pm 1.05	15.46 \pm 1.26	14.68 \pm 1.18	88.62 \pm 3.71	14.97 \pm 1.12	7.46 \pm 0.70
HS	12.69 \pm 1.40	25.47 \pm 2.38	25.33 \pm 3.27	92.80 \pm 4.06	17.74 \pm 1.24	13.49 \pm 4.43
NGHS	10.62 \pm 1.18	18.13 \pm 1.52	16.16 \pm 1.94	86.76 \pm 2.05	15.79 \pm 1.51	8.02 \pm 0.86
PSO	11.53 \pm 1.82	22.87 \pm 5.21	22.00 \pm 3.61	95.73 \pm 5.59	18.75 \pm 2.01	12.02 \pm 3.71
SGHS	10.34 \pm 0.99	15.33 \pm 1.04	14.84 \pm 0.83	90.13 \pm 3.37	14.09 \pm 1.11	8.43 \pm 0.68

TABLE IV
AVERAGE ACCURACY (%) OVER THE TEST SET CONSIDERING EXPERIMENT I AND \mathbb{M}_1

	CBERS-2B	Geoeeye	Ikonos-2 MS	Indian	Landsat-5 TM	Salinas
<i>k</i> -means	62.13 \pm 3.16	57.86 \pm 2.31	53.61 \pm 1.61	50.01 \pm 0.11	60.30 \pm 2.15	50.41 \pm 0.19
BA	59.44 \pm 4.56	59.96 \pm 1.94	56.82 \pm 2.36	50.86 \pm 0.21	64.21 \pm 2.97	49.83 \pm 0.18
FFA	60.73 \pm 2.00	59.57 \pm 3.26	57.33 \pm 1.88	50.44 \pm 0.45	60.75 \pm 3.37	50.28 \pm 0.28
GHS	60.01 \pm 1.17	59.62 \pm 2.36	55.55 \pm 3.09	50.56 \pm 0.54	59.44 \pm 4.43	50.31 \pm 0.30
GP	62.50 \pm 3.18	58.74 \pm 3.15	57.07 \pm 2.73	50.94 \pm 0.39	62.81 \pm 1.82	50.07 \pm 0.16
HS	61.28 \pm 5.33	57.23 \pm 0.54	57.56 \pm 1.13	51.08 \pm 0.35	60.49 \pm 2.87	50.45 \pm 0.52
NGHS	59.40 \pm 4.05	57.78 \pm 1.26	57.28 \pm 1.82	51.20 \pm 0.48	58.34 \pm 4.10	49.91 \pm 0.17
PSO	60.52 \pm 2.49	60.66 \pm 2.28	56.63 \pm 1.95	50.61 \pm 0.42	59.65 \pm 3.79	50.51 \pm 0.34
SGHS	62.31 \pm 5.24	54.03 \pm 8.22	57.26 \pm 5.85	50.90 \pm 1.64	64.52 \pm 3.34	50.89 \pm 0.14

In regard to HS-based approaches, harmony memory considering rate (HMCr) controls the probability of creating a new harmony from the solutions obtained previously, and pitch adjusting rate (PAR) is a probability that decides whether a possible new solution will be adjusted by some small step-size or not. The variants of it, i.e., IHS, GHS, NGHS, and SGHS, employ similar criteria, but with slight modifications. The other techniques employ different parameters, but they are related to the very same meaning, i.e., to control the amount of moving from a previous solution to a new one, and to allow small perturbations to escape from local optima. The reader can refer to the main references for each technique for a more detailed explanation about the rationale of each parameter.

The experiments also considered two different performance evaluators as the fitness functions: \mathbb{M}_1 (1) and \mathbb{M}_2 (2). The idea is to assess the robustness of the proposed approach under different performance evaluators. In order to fulfil this purpose, we used 50% of the entire image to compose the training set, and the remaining 50% to be part of the test set. Thus, the final starting positions (centroids) obtained by these techniques are then used to initialize *k*-means over the test set. It is worth mentioning, we performed a cross-validation approach with 20 executions for statistical purposes.² Finally, each dataset sample (pixel) is described by its brightness value, i.e., a 3-D feature vector regarding CBERS-2B, Landsat 5 TM, Ikonos-2 MS and Geoeeye

²We employed a twofold cross validation, i.e., the so-called holdout method.

TABLE V
AVERAGE ACCURACY (%) OVER THE TEST SET CONSIDERING EXPERIMENT I AND \mathbb{M}_2

	CBERS-2B	Geoeeye	Ikonos-2 MS	Indian	Landsat-5 TM	Salinas
k-means	60.82 \pm 3.89	55.26 \pm 2.73	59.05 \pm 1.82	50.11 \pm 0.57	59.33 \pm 2.34	50.00 \pm 0.00
BA	61.63 \pm 3.76	58.10 \pm 3.50	57.01 \pm 1.25	50.61 \pm 0.43	63.03 \pm 2.64	50.14 \pm 0.27
FFA	60.12 \pm 3.33	60.84 \pm 2.14	55.10 \pm 2.61	50.61 \pm 0.37	63.38 \pm 3.23	50.33 \pm 0.79
GHS	60.44 \pm 3.98	61.24 \pm 2.29	57.67 \pm 2.14	51.17 \pm 0.59	60.30 \pm 4.69	50.24 \pm 0.13
GP	62.55 \pm 2.19	58.63 \pm 4.21	55.96 \pm 3.37	50.75 \pm 0.24	62.71 \pm 2.70	50.33 \pm 0.19
HS	61.28 \pm 3.13	54.98 \pm 2.86	54.66 \pm 2.81	50.71 \pm 0.29	56.62 \pm 3.19	50.20 \pm 0.45
NGHS	62.83 \pm 1.35	57.82 \pm 2.02	57.17 \pm 2.99	51.02 \pm 0.26	63.17 \pm 4.40	50.30 \pm 0.50
PSO	61.92 \pm 1.74	55.87 \pm 2.08	58.17 \pm 1.18	50.42 \pm 0.27	60.53 \pm 2.75	50.42 \pm 0.17
SGHS	62.81 \pm 1.12	60.64 \pm 2.49	56.12 \pm 2.66	50.92 \pm 0.24	63.36 \pm 3.53	50.35 \pm 0.36

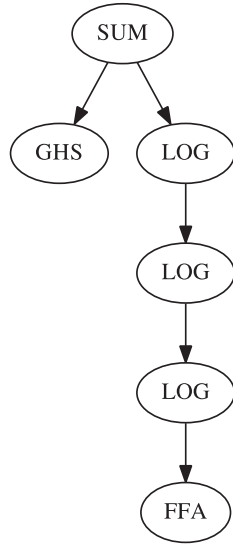


Fig. 5. Example of a complex best tree generated during CBERS-2B experiment using \mathbb{M}_1 .

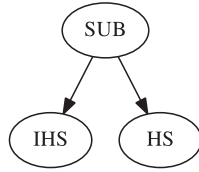


Fig. 6. Example of a simple best tree generated during CBERS-2B experiment using \mathbb{M}_1 .

images, and around a 200-dimensional feature vector regarding the hyperspectral images provided by AVIRIS (Salinas-A and Indian Pines).

B. Experimental Results

In this section, we present the experimental results to evaluate the proposed approach based on hyperheuristics and GP. The experimental section is divided in two rounds: 1) in the first one (see Section IV-B1), we evaluated all aforementioned meta-heuristic techniques, i.e., HS, IHS, GHS, NGHS, SGHS, PSO, BA, and FFA and 2) in the second round (see Section IV-B2),

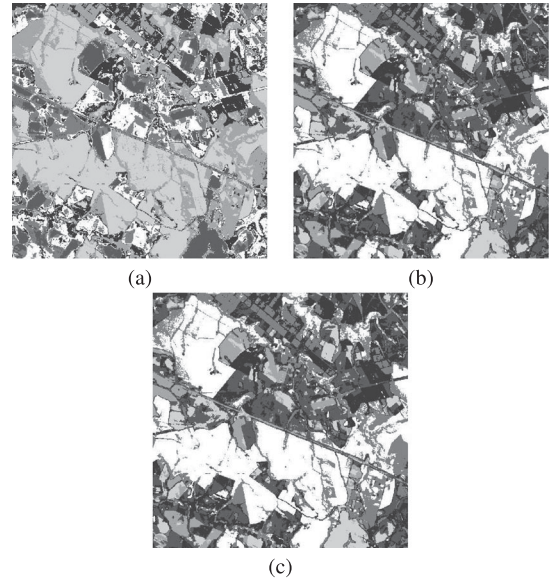


Fig. 7. Classified images regarding CBERS-2B satellite by (a) naive k -means, (b) SGHS, and (c) GP.

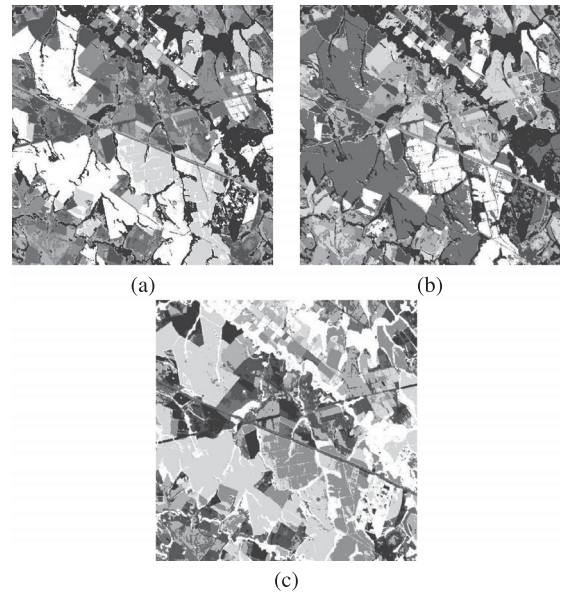


Fig. 8. Classified images regarding Landsat 5 TM satellite by (a) naive k -means, (b) SGHS, and (c) GP.

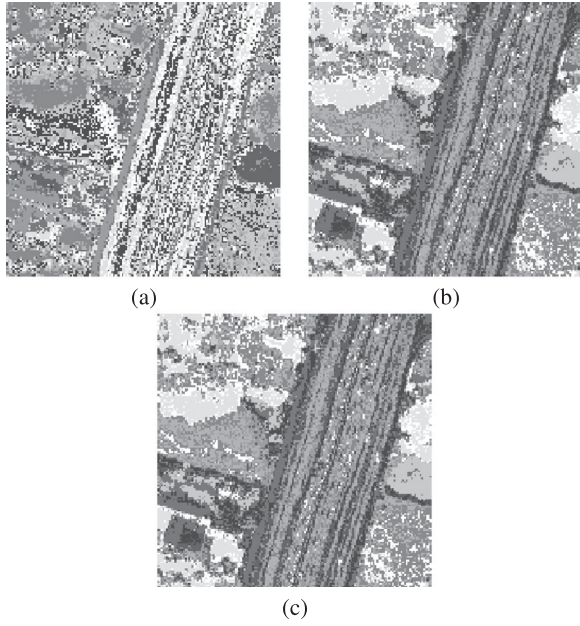


Fig. 9. Classified images regarding Ikonos-2 MS satellite by (a) naïve k -means, (b) SGHS, and (c) GP.

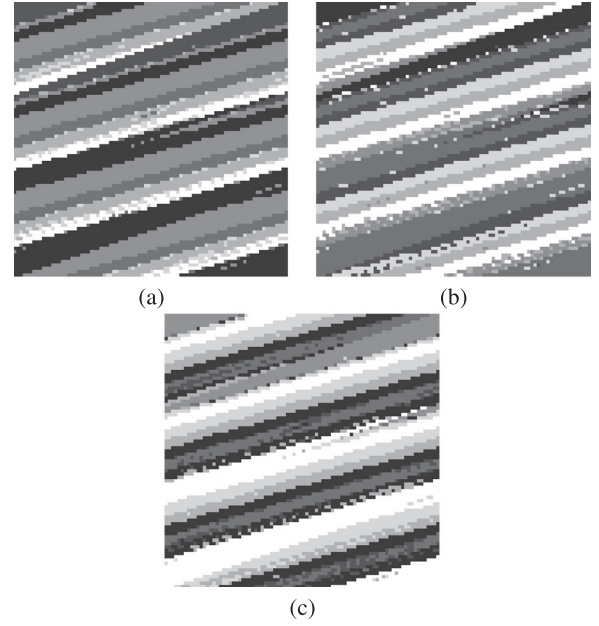


Fig. 11. Classified images regarding Salinas-A by (a) naïve k -means, (b) HS, and (c) GP.

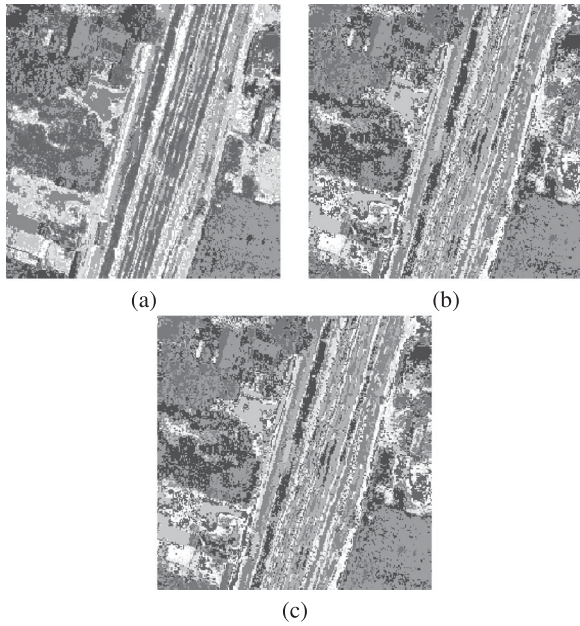


Fig. 10. Classified images regarding Geoeeye satellite by (a) naïve k -means, (b) SGHS, and (c) GP.

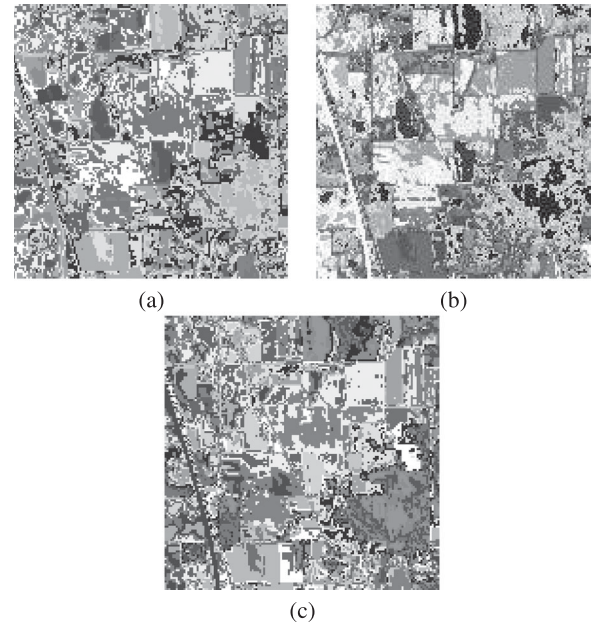


Fig. 12. Classified images regarding Indian Pines by (a) naïve k -means, (b) IHS, and (c) GP.

we removed SGHS from the set of possible choices, since it obtained the best individual result and may bias GP during the optimization process.

1) *Experiment I:* In this section, we present the experiments using the aforementioned metaheuristic techniques. Table II and III present the average of the \mathbb{M}_1 and \mathbb{M}_2 values, respectively. The values in bold stand for similar techniques according to the Wilcoxon signed rank statistical test [46]. The proposed approach based on GP obtained the best results for five images considering the metric \mathbb{M}_1 . In one out five situations (Salinas-A), GP was the technique that achieved the top results, being

followed by SGHS, BA, FFA, and HS. That results are interesting, since we can observe GP was able to improve the results of the best individual technique for some situations.

In regard to measure \mathbb{M}_2 , GP obtained the most accurate results for four out six images once again (except for Indian Pines and Landsat-5M), being the sole best technique in two out six images (Ikonos-2 MS and Salinas). Roughly speaking, the rationale of (4) is to both minimize ADDC (intracluster distances) and to maximize ADDBC (intercluster distances), i.e., we aim at minimizing $1/\text{ADDBC}$. That is the reason why we use $(k \text{ ADDC})/\text{ADDBC}$. In fact, the behavior of GP was

TABLE VI
MEAN EXECUTION TIMES [S] OBTAINED BY DIFFERENT OPTIMIZATION APPROACHES

	CBERS-2B	Geoeye	Ikonos-2 MS	Landsat-5 TM	Salinas-A	Indian Pines
<i>k</i> -means	21.52	13.88	6.74	15.66	11.94	158.02
BA	18995.18	11114.44	10643.2	28418.16	25934.00	76277.6
FFA	791.21	286.11	284.28	904.79	804.69	3445.34
GHS	75.10	40.76	45.39	91.56	66.01	318.14
GP	5412.14	2516.50	2518.28	8662.74	8363.18	39012.10
HS	160.3	80.66	99.63	220.27	205.55	729.72
IHS	82.58	41.10	47.49	94.99	82.15	319.46
NGHS	84.73	39.53	35.75	81.00	68.25	309.15
PSO	733.95	268.50	315.88	795.87	862.74	4757.77
SGHS	82.71	39.62	37.67	61.42	71.56	315.27

TABLE VII
AVERAGE \mathbb{M}_1 OVER THE TEST SET CONSIDERING EXPERIMENT II

	CBERS-2B	Geoeye	Ikonos-2 MS	Indian	Landsat-5 TM	Salinas
GP	22.83 ± 1.01	13.60 ± 1.77	15.30 ± 0.98	2419.08 ± 9.01	23.21 ± 2.08	1059.81 ± 19.37

TABLE VIII
AVERAGE \mathbb{M}_2 OVER THE TEST SET CONSIDERING EXPERIMENT II

	CBERS-2B	Geoeye	Ikonos-2 MS	Indian	Landsat-5 TM	Salinas
GP	10.67 ± 0.96	16.06 ± 0.77	14.88 ± 1.01	89.37 ± 4.07	14.86 ± 1.22	7.60 ± 1.07

quite similar considering both techniques, thus showing it can be robust to different performance measures.

We employed a specific methodology to estimate the accuracy of the clustering process, since we have the ground-truth images. However, it is important to highlight the accuracy was not used as a fitness function, and then, we cannot guarantee GP will obtain the best results ever. For each center computed by *k*-means, we used its true label (or the true label of its nearest sample, since the center may not be a training sample) to classify all samples that fall in the same cluster. In regard to the accuracy measure, we used the standard approach, i.e., the ratio between the number of correctly classified samples and the total number of testing samples. Table IV and V present the average accuracy and standard deviation for each optimization method considering \mathbb{M}_1 and \mathbb{M}_2 , respectively. Notice GP obtained the best results for several situations, although the accuracy was not used as a fitness function.

Fig. 5 depicts the best GP tree selected from the ninth iteration of the cross-validation procedure concerning CBERS-2B image and performance measure \mathbb{M}_1 . Notice this tree combines very accurate isolated metaheuristics, such as FFA, as well as not so good ones, such as GHS. Such results are very interesting, and they confirm what we are looking for when using GP, i.e., to obtain more diverse results that can somehow boost only the most accurate techniques, but also others. The expression encoded by the tree displayed in Fig. 5 is given by

$$F = \text{GHS} + \text{LOG}(\text{LOG}(\text{LOG}(\text{FFA}))) \quad (5)$$

where *F* means the final result. In the context of this study, suppose we have a single-band image with two land-cover classes

to be classified, i.e., we have two clusters $c_1, c_2 \in \mathcal{R}$ only, being $k = 2$, the input to *k*-means algorithm. As such, each cluster will be represented by a real value in the feature space, being the best solution the one that minimizes \mathbb{M}_1 or \mathbb{M}_2 .

Let $C^{\text{FFA}} = \{c_1^{\text{FFA}}, c_2^{\text{FFA}}\}$ and $C^{\text{GHS}} = \{c_1^{\text{GHS}}, c_2^{\text{GHS}}\}$ be the best solutions found by FFA and GHS, respectively. This means, for instance, C^{FFA} stands for the set of initial clusters' coordinates to be used as input to *k*-means in such a way \mathbb{M}_1 (or \mathbb{M}_2) is minimized. For the sake of simplification, we can rewrite (5) as follows:

$$C^* = C^{\text{GHS}} + \text{LOG}(\text{LOG}(\text{LOG}(C^{\text{FFA}}))) \quad (6)$$

where C^* stands for the best solution found by GP. Roughly speaking, (6) is just an expression that can be solved using pointwise-driven operations.

Obviously, we have also some simpler trees, as the one presented in Fig. 6, which was obtained during the seventh iteration of the very same cross-validation procedure concerning CBERS-2B image. This tree was considered the best at that iteration, but it does not contain any technique that obtained outstanding results when we consider them individually (see Table II). GP has the interesting ability to design possible solutions (trees) with very distinct shapes, but at the price of a higher computational cost. Such aspects motivated us to employ GP as a hyperheuristic approach, and the other techniques to compose the terminal nodes.

Figs. 7–12 depict some images classified with traditional *k*-means, as well as their enhanced version with centroids computed by SGHS and GP, considering CBERS-2B, Landsat

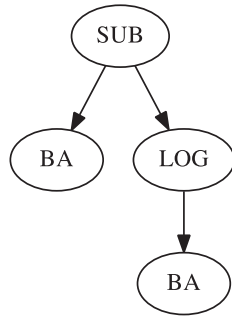


Fig. 13. Example of a best tree generated during Salinas-A experiment II using \mathbb{M}_2 .

5 TM, Ikonos-2 MS, Geoeye, Salinas-A, and Indian Pines, respectively. Clearly, the results using metaheuristic techniques and GP were much more accurate than naïve k -means when taking into account the ground-truth images (see Fig. 4).

Additionally, we also computed the mean execution times (seconds) for all metaheuristic techniques (see Table VI). Notice the experiments were conducted on personal computer equipped with an Intel i7 5500U processor with 8 GB of RAM using Linux as the operational system. Usually, when one deal with metaheuristic-driven optimization, swarm-based approaches take longer to execute, since they update all agents at each iteration, while HS-based approaches do that only once per iteration. Although they are much faster, they usually pay the price of a slow convergence process, but it can be worth in the end. If we consider the times displayed in Table VI, the slowest technique was BA, which may slow down GP as well. This means we could try to remove BA from the experimental setting, and try to make GP faster. Clearly, as we are adding one more optimization step, we can also expect a higher computational load. Another possibility would be to employ HS-based approaches only, which can turn the process of learning trees much more feasible.

2) *Experiment II*: In this second round of experiments, we aimed at showing the robustness of the proposed approach when removing a very good individual technique from the previous experiment. For such purpose, we considered removing SGHS. Table VII and VIII present the mean values of \mathbb{M}_1 and \mathbb{M}_2 performance evaluators considering GP only, but using the very same experimental methodology of the former experiment. GP still obtained the best results for five out six images with respect to both \mathbb{M}_1 and \mathbb{M}_2 , except for Indian Pines for the two metrics. This experiment is interesting to show we can also obtain better combined results even if we have individual optimizers with similar results.

Fig. 13 depicts a best tree from the sixth execution of the cross-validation procedure considering Salinas-A image and performance evaluator \mathbb{M}_2 . Since BA obtained very good results concerning experiment I (see Table VIII), it is expected it should be part of some best tree as a terminal node. In this case, one can realize we obtained a best tree composed of BA only. However, this was a specific situation, which means other best trees can also be obtained with or without BA.

V. CONCLUSION

Unsupervised land-cover classification has been one of the foremost applications in the last years, mainly due to the lack of high-quality labeled data. Still one of the most used pattern recognition techniques out there, the well-known k -means can achieve interesting results, but its original formulation to find the nearest centroid to each dataset sample may get trapped from local optima.

In this study, we attempt at modeling the k -means algorithm as a metaheuristic optimization problem, in which the starting centroids are then computed by distinct optimization techniques, that are further combined in a hyperheuristic methodology using GP. Experiments over six satellite and two performance evaluators (fitness function) assessed the robustness of the proposed approach in two distinct rounds: in the former, we employed eight different metaheuristic optimization techniques, being the best results obtained by GP in almost all datasets. Some results concerning the proposed approach were quite close, but even better, to the best individual ones. These results seem to confirm the hypothesis that GP tends to degenerate to the more accurate approaches when we try to combine distinct techniques with very different behaviors. In a second round of experiments, we removed SGHS from the collection of techniques to be combined by GP in order to show it can still obtain very good results, thus improving the results of individual techniques.

In regard to future works, we aim at combining more metaheuristic optimization techniques, as well as to evaluate different GP operators to the very same context of this study.

REFERENCES

- [1] R. J. Pisani, R. Y. M. Nakamura, P. S. Riedel, C. R. L. Zimback, A. X. Falcão, and J. P. Papa, "Toward satellite-based land cover classification through optimum-path forest," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6075–6085, Oct. 2014. [Online]. Available: <http://dx.doi.org/10.1109/TGRS.2013.2294762>
- [2] H. Bagan, W. Takeuchi, T. Kinoshita, Y. Bao, and Y. Yamagata, "Land cover classification and change analysis in the Horqin sandy land from 1975 to 2007," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 3, no. 2, pp. 168–177, Jun. 2010. [Online]. Available: <http://dx.doi.org/10.1109/JSTARS.2010.2046627>
- [3] M. N. Capucim, V. S. Brand, C. B. Machado, L. D. Martins, D. G. Al-lasia, C. T. Homann, E. D. Freitas, M. A. F. S. Dias, M. F. Andrade, and J. Martins, "South America land use and land cover assessment and preliminary analysis of their impacts on regional atmospheric modeling studies," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 3, pp. 1185–1198, Mar. 2015. [Online]. Available: <http://dx.doi.org/10.1109/JSTARS.2014.2363368>
- [4] B. Banerjee, F. Bovolo, A. Bhattacharya, L. Bruzzone, S. Chaudhuri, and B. K. Mohan, "A new Self-Training-Based unsupervised satellite image classification technique using cluster ensemble strategy," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 4, pp. 741–745, Apr. 2015. [Online]. Available: <http://dx.doi.org/10.1109/LGRS.2014.2360833>
- [5] G. Dong, G. Kuang, N. Wang, L. Zhao, and J. Lu, "SAR target recognition via joint sparse representation of monogenic signal," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 7, pp. 3316–3328, Jul. 2015. [Online]. Available: <http://dx.doi.org/10.1109/JSTARS.2015.2436694>
- [6] M. Martorella, E. Giusti, L. Demi, Z. Zhou, A. Cacciamano, F. Berizzi, and B. Bates, "Target recognition by means of polarimetric ISAR images," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 47, no. 1, pp. 225–239, Jan. 2011. [Online]. Available: <http://dx.doi.org/10.1109/TAES.2011.5705672>
- [7] L. Du, P. Wang, H. Liu, M. Pan, F. Chen, and Z. Bao, "Bayesian spatiotemporal multitask learning for radar HRRP target recognition," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3182–3196, Jul. 2011. [Online]. Available: <http://dx.doi.org/10.1109/TSP.2011.2141664>
- [8] J. Li, X. Huang, P. Gamba, J. M. B. Bioucas-Dias, L. Zhang, J. A. Benediktsson, and A. Plaza, "Multiple feature learning for hyperspectral

- image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1592–1606, Mar. 2015. [Online]. Available: <http://dx.doi.org/10.1109/TGRS.2014.2345739>
- [9] A. Voisin, V. A. Krylov, G. Moser, S. B. Serpico, and J. Zerubia, "Supervised classification of multisensor and multiresolution remote sensing images with a hierarchical Copula-Based approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3346–3358, Jun. 2014. [Online]. Available: <http://dx.doi.org/10.1109/TGRS.2013.2272581>
 - [10] H. Yang, Q. Du, and G. Chen, "Unsupervised hyperspectral band selection using graphics processing units," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 4, no. 3, pp. 660–668, Sep. 2011. [Online]. Available: <http://dx.doi.org/10.1109/JSTARS.2011.2120598>
 - [11] Y. Yuan, G. Zhu, and Q. Wang, "Hyperspectral band selection by multitask sparsity pursuit," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 631–644, Feb. 2015. [Online]. Available: <http://dx.doi.org/10.1109/TGRS.2014.2326655>
 - [12] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, L. M. L. Cam and J. Neyman, Eds. Berkeley, CA, USA: Univ. of California Press, 1967, pp. 281–297.
 - [13] Y.-R. Chen, J.-W. Chen, S.-C. Hsieh, and P.-N. Ni, "The application of remote sensing technology to the interpretation of land use for Rainfall-Induced landslides based on genetic algorithms and artificial neural networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 2, no. 2, pp. 87–95, Jun. 2009. [Online]. Available: <http://dx.doi.org/10.1109/JSTARS.2009.2023802>
 - [14] R. Y. M. Nakamura, L. M. G. Fonseca, J. A. Santos, R. S. Torres, X.-S. Yang, and J. P. Papa, "Nature-Inspired framework for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 4, pp. 2126–2137, Apr. 2014. [Online]. Available: <http://dx.doi.org/10.1109/TGRS.2013.2258351>
 - [15] S. Goel, M. Gaur, and E. Jain, "Nature inspired algorithms in remote sensing image classification," *Procedia Comput. Sci.*, vol. 57, pp. 377–384, 2015. [Online]. Available: <http://dx.doi.org/10.1016/j.procs.2015.07.352>
 - [16] J. Senthilnatha, X.-S. Yang, and J. A. Benediktsson, "Automatic registration of multi-temporal remote sensing images based on nature-inspired techniques," *Int. J. Remote Sens.*, vol. 5, pp. 263–284, 2014.
 - [17] R. Kheddami and A. Belhadj-Aissa, "Classification of remotely sensed images using clonal selection theory of Artificial Immune System, presented at IEEE 4th Int. Conf. Image Process. Theory Tools Appl., Oct. 2014. [Online]. Available: <http://dx.doi.org/10.1109/IPTA.2014.7001971>
 - [18] L. P. Chandran and K. A. A. Nazeer, "An improved clustering algorithm based on K-means and harmony search optimization, presented at the IEEE Recent Adv. Intell. Comput. Syst. Sep. 2011. [Online]. Available: <http://dx.doi.org/10.1109/RAICS.2011.6069352>
 - [19] R. Forsati, M. R. Meybodi, M. Mahdavi, and A. G. Neiat, "Hybridization of K-Means and harmony search methods for web page clustering, presented at the IEEE/WIC/ACM Web Intell. Intell. Agent Technol., pp. 329–335, Dec. 2008. [Online]. Available: <http://dx.doi.org/10.1109/WIAT.2008.370>
 - [20] Y. Lin, N. Tong, M. Shi, K. Fan, D. Yuan, L. Qu, and Q. Fu, "K-means optimization clustering algorithm based on particle swarm optimization and multiclass merging," in *Advances in Intelligent and Soft Computing*. Berlin, Germany: Springer, pp. 569–578, 2012. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-30126-1_90
 - [21] R. J. Kuo, S. S. Chen, W. C. Cheng, and C. Y. Tsai, "Integration of artificial immune network and K-means for cluster analysis," *Knowl. Inform. Syst.*, vol. 40, no. 3, pp. 541–557, May. 2013. [Online]. Available: <http://dx.doi.org/10.1007/s10115-013-0649-3>
 - [22] I. B. Saida, N. Kamel, and B. Omar, "A new hybrid algorithm for document clustering based on cuckoo search and K-means," in *Advances in Intelligent Systems and Computing*. Switzerland: Springer, pp. 59–68, 2014. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-07692-8_6
 - [23] S. Fong, S. Deb, X.-S. Yang, and Y. Zhuang, "Towards enhancement of performance of K-Means clustering using nature-inspired optimization algorithms," *Scientific World J.*, pp. 1–16, 2014. [Online]. Available: <http://dx.doi.org/10.1155/2014/564829>
 - [24] P. Ross, "Hyper-heuristics," in *Search Methodologies*. New York, NY, USA: Springer, 2005, pp. 529–556. [Online]. Available: http://dx.doi.org/10.1007/0-387-28356-0_17
 - [25] J. P. Papa, L. P. Papa, D. R. Pereira, and R. J. Pisani, "Unsupervised land-cover classification through hyper-heuristic-based harmony search," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2015, pp. 69–72.
 - [26] E. R. Hruschka, R. J. G. B. Campello, A. A. Freitas, and A. C. P. L. F. Carvalho, "A survey of evolutionary algorithms for clustering," *IEEE Trans. Syst. Man, Cybern. C, Appl. Rev.*, vol. 39, no. 2, pp. 133–155, Mar. 2009.
 - [27] M. C. Naldi, R. J. G. B. Campello, E. R. Hruschka, and A. C. P. L. F. Carvalho, "Efficiency issues of evolutionary k-means," *Appl. Soft Comput.*, vol. 11, no. 2, pp. 1938–1952, 2011.
 - [28] M. C. Naldi and R. J. G. B. Campello, "Evolutionary k-means for distributed data sets," *Neurocomputing*, vol. 127, no. 15, pp. 30–42, 2014.
 - [29] D. Luchi, W. Santos, A. A. Rodrigues, and F. M. Varejão, "Genetic sampling k-means for clustering large data sets," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications (Lecture Notes in Computer Science)*, vol. 9423, A. Pardo and J. Kittler, Eds. Switzerland: Springer, 2015, pp. 691–698.
 - [30] A. Bousbaci and N. Kamel, "A parallel sampling-PSO-multi-core-k-means algorithm using mapreduce," in *Proc. 14th Int. Conf. Hybrid Intell. Syst.*, 2014, pp. 129–134.
 - [31] A. B. S. Serapião, G. S. Corrêa, F. B. Gonçalves, and V. O. Carvalho, "Combining k-means and k-harmonic with fish school search algorithm for data clustering task on graphics processing units," *Appl. Soft Comput.*, vol. 41, pp. 290–304, 2016.
 - [32] I. B. Saida, K. Nadjet, and B. Omar, "A new algorithm for data clustering based on cuckoo search optimization," in *Genetic and Evolutionary Computing (Advances in Intelligent Systems and Computing)*, vol. 238, J.-S. Pan, P. Krömer, and V. Snášel, Eds. Switzerland: Springer, 2014, pp. 55–64.
 - [33] D. Binu, M. Selvi, and A. George, "MKF-cuckoo: Hybridization of cuckoo search and multiple kernel-based fuzzy c-means algorithm," *AASRI Procedia*, vol. 4, pp. 243–249, 2013.
 - [34] I. B. Saida, K. Nadjet, and B. Omar, "A new hybrid algorithm for document clustering based on cuckoo search and k-means," in *Recent Advances on Soft Computing and Data Mining (Advances in Intelligent Systems and Computing)*, vol. 287, T. Herawan, R. Ghazali, and M. M. Deris, Eds. Switzerland: Springer, 2014.
 - [35] T. Niknam and B. Amiri, "An efficient hybrid approach based on PSO, ACO and k-means for cluster analysis," *Appl. Soft Comput.*, vol. 10, no. 1, pp. 183–197, 2010.
 - [36] T. Özzyer, R. Alhajj, and K. Barker, "Clustering by integrating multi-objective optimization with weighted k-means and validity analysis," in *Intelligent Data Engineering and Automated Learning (Lecture Notes in Computer Science)*, vol. 4224, E. Corchado, H. Yin, V. Bott, and C. Fyfe, Eds. Berlin, Germany: Springer, 2006, pp. 454–463.
 - [37] M. Anusha and J. G. R. Sathiaselan, "Feature selection using k-means genetic algorithm for multi-objective optimization," *Procedia Comput. Sci.*, vol. 57, pp. 1074–1080, 2015.
 - [38] J. R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. Cambridge, MA, USA: MIT Press, 1992.
 - [39] M. Mahdavi, M. Fesanghary, and E. Damangir, "An improved harmony search algorithm for solving optimization problems," *Appl. Math. Comput.*, vol. 188, no. 2, pp. 1567–1579, May. 2007. [Online]. Available: <http://dx.doi.org/10.1016/j.amc.2006.11.033>
 - [40] M. G. H. Omran and M. Mahdavi, "Global-best harmony search," *Appl. Math. Comput.*, vol. 198, no. 2, pp. 643–656, May. 2008. [Online]. Available: <http://dx.doi.org/10.1016/j.amc.2007.09.004>
 - [41] D. Zou, L. Gao, J. Wu, S. Li, and Y. Li, "A novel global harmony search algorithm for reliability problems," *Comput. Ind. Eng.*, vol. 58, no. 2, pp. 307–316, Mar. 2010. [Online]. Available: <http://dx.doi.org/10.1016/j.cie.2009.11.003>
 - [42] Q.-K. Pan, P. N. Suganthan, M. F. Tasgetiren, and J. J. Liang, "A self-adaptive global best harmony search algorithm for continuous optimization problems," *Appl. Math. Comput.*, vol. 216, no. 3, pp. 830–848, Apr. 2010. [Online]. Available: <http://dx.doi.org/10.1016/j.amc.2010.01.088>
 - [43] J. Kennedy and R. C. Eberhart, *Swarm Intelligence*. San Francisco, CA, USA: Morgan Kaufmann, 2001.
 - [44] X.-S. Yang and A. H. Gandomi, "Bat algorithm: A novel approach for global engineering optimization," *Eng. Comput.*, vol. 29, no. 5, pp. 464–483, Jul. 2012. [Online]. Available: <http://dx.doi.org/10.1108/02644401211235834>
 - [45] X.-S. Yang, "Firefly algorithm, stochastic test functions and design optimisation," *Int. J. Bio-Inspired Comput.*, vol. 2, no. 2, pp. 78–84, 2010.
 - [46] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics Bull.*, vol. 1, no. 6, pp. 80–83, 1945.

Authors' photographs and biographies not available at the time of publication.