Algorithmic Information Theory-Based Analysis of Earth Observation Images: An Assessment

Daniele Cerra, *Student Member, IEEE*, Alexandre Mallet, Lionel Gueguen, and Mihai Datcu, *Senior Member, IEEE*

Abstract—**Earth observation image-understanding methodologies may be hindered by the assumed data models and the estimated parameters on which they are often heavily dependent. First, the definition of the parameters may negatively affect the quality of the analysis. The parameters could not be captured in all aspects, and those resulting superfluous or not accurately tuned may introduce nuisance in the data. Furthermore, the diversity of the data, as regards sensor type, spatial, spectral, and radiometric resolution, and the variety and regularity of the observed scenes make it difficult to establish enough valid and robust statistical models to describe them. This letter proposes algorithmic information theory-based analysis as a valid solution to overcome these limitations. We will present different applications on satellite images, i.e., clustering, classification, artifact detection, and image time series mining, showing the generalization power of these parameter-free data-driven methods based on the computational complexity analysis.**

Index Terms—**Clustering, data compression, image classification, Kolmogorov complexity.**

I. Introduction

CLASSIC image analysis methodologies often require strong *a priori* knowledge of the data, which may be, in some cases, available, but often put undesired limitations in applications to the Earth observation image databases. In fact, the large and steadily growing volume of data provided by satellites, along with the great variety, diversity, and irregularity of the observed scenes, makes it hard to establish enough general statistical description models for these data sets. Furthermore, both the definition and the setting of the parameters have a strong subjective component and may either underfit the data, failing to capture relevant information, or overfit them, introducing nuisance. This drawback in several typical image-understanding problems like segmentation, classification, or clustering is particularly affecting the image information mining applications, which usually process large volumes of data, often not restricted to homogeneous data sets.

These limitations are overcome with the methods described in this letter, which aim at extracting information and computing similarities within the data on the sole basis of their computational complexity, estimated with solutions emerging from

algorithmic information theory. The concept to the base of these methodologies, not yet familiar within the geoscience community, is the Kolmogorov complexity [1], which is incomputable *in se* but can be approximated by compression algorithms, as we will see more in detail. This approach is totally data-driven, independent of any statistical model, and requires no parameter to be set.

This letter is structured as follows. Section II presents a short overview on the theory on which these methods are based, together with the state of the art. The subsequent sections report practical applications in the field of satellite image analysis. Section III illustrates an unsupervised hierarchical clustering and classification experiments on optical data, Section IV presents artifact detection, and Section V introduces an application of image time series mining. Finally, Section VI reports our conclusions.

II. Preliminaries

A. Algorithmic Information Theory Frame

Complexity and information content are two tightly related concepts that can be addressed by Shannon's classic probabilistic approach [2] or by the more recent algorithmic point of view, defined by Kolmogorov [1] and later on by Chaitin [3]. These concepts are strongly linked [4] but different since Shannon's approach does not provide the informational content of individual objects or their regularity. The Kolmogorov complexity, instead, evaluates the intrinsic information content of a single object independently from any description formalism. The Kolmogorov complexity $K(x)$ of a string x is the length $|q|$ of the program q that outputs x and halts on an appropriate universal machine, such as the universal Turing machine, and it is defined as $K(x) = \min_{q \in Qx} |q|$, with Qx being the set of programs generating x. Therefore, the strings presenting the recurring patterns have low complexity, whereas the complexity of random strings is high and almost equals their own length. It is important to remark that $K(x)$ is not a computable function of x. This assimilation of information content to computational complexity is generating an interest which is not confined within the information theory community, as shown in a recent book by Chaitin [5] which also aimed at an audience outside of the field, and it opens the way to many interesting concepts and applications. For our purposes, the perhaps most interesting one is the normalized information distance (NID) [6], which has its roots in the Shannon's concept of mutual information. The NID between two objects x and y is defined as the length of the shortest program that computes x knowing y, as well as computing y knowing x, i.e., it is proportional to the quantity

$\max\{K(x|y), K(y|x)\}$, where $K(x|y)$ is the conditional Kolmogorov complexity of x when y is given as an auxiliary input for the computation; after normalization, the distance is

$$NID(x,y) = \frac{K(x,y) - \min\{K(x), K(y)\}}{\max\{K(x), K(y)\}}. \qquad (1)$$

$NID(x,y)$ is a metric minimizing all normalized distances, with $NID(x,y) = 0$ iff $x = y$ and $NID(x,y) = 1$ representing the maximum distance between x and y.

B. Computable Distances and State of the Art

The NID is based on Kolmogorov complexity, and is thus incomputable, but it *comes to reality* with an approximation proposed by Li *et al.* [7]. To find a suitable approximation for $K(x)$, the latter is stated to represent the lower bound for what a real compressor can achieve. This allows approximating $K(x)$ with $C(x)$, i.e., the length of the compressed version of x obtained with a standard lossless compressor C such as Gzip. Equation (1), in its explicitly computable form, is the normalized compression distance $NCD(x,y)$, which is defined for the two objects x and y as

$$NCD(x,y) = \frac{C(x,y) - \min\{C(x), C(y)\}}{\max\{C(x), C(y)\}} \qquad (2)$$

with $C(x,y)$ representing the size of the file obtained by compressing the concatenation of x and y. The NCD can be adopted to compute similarities between any two objects, and the conditions for it to be a metric hold under certain assumptions. Before the concept of information distance, the idea of exploiting the intrinsic power of data compression to match recurring patterns within the data can be found *in nuce* in [8]. Subsequently, the pattern recognition based on data compression (PRDC), another methodology to classify general kinds of data (text, drawings, proteins [10], images, music), which are encoded into strings in a preliminary step, was published by Watanabe *et al.* [9]; to the best of our knowledge, the PRDC is the only classification methodology based on data compression which has been tested on an airborne or satellite imagery, apart from other works by Cerra and Datcu [11]–[13]. Definition (2), which relies on a more solid theoretical background, provides a more robust index since it is normalized, and can be applied to any kind of file by using any off-the-shelf compressor, allowing to discard any limitation due to the physical format of the objects, which previously had to be strings or preencoded into strings. This information distance is very general, and experiments have been carried out with the NCD-like measures and other indexes to compute similarities within very diverse data, such as simple text files [14], music samples [15], dictionaries from different languages [14], and tables [16]. There are many applications in the field of bioinformatics. DeoxyriboNucleic Acid sequences are classified in [14], [17], and [18], and an information content estimation of RiboNucleic Acid is obtained in [19]. An extensive test of the power and the adaptability of these techniques is presented by Keogh *et al.* [18], [20], with experiments of clustering, anomaly detection, and classification, carried out on different data types, and backed by

comparisons with 51 other measures. Cohen *et al.* [21] use a similar approach, also entering in the field of Kolmogorov's algorithmic structure functions [22], to summarize changes in biological image sequences. Among the most unusual applications, we recall a program to detect plagiarism [23] and a study on the evolution of chain letters [24]. Semantic relations between words may be discovered on the Web using Google as a compressor [25]. Many distance measures similar to the NCD have been used in these works (see, for example, the compression-based dissimilarity measure [18]), but, recently in [26], it has been shown that all of them differ very little since they can be brought in a canonical form with the normalization term as the only difference between them. Also in this work, these distance measures are assimilated to distances of implicit vectors projected on implicit invisible feature spaces, hinting potential cross relations with the area of machine learning. Research in this area is still very active. Recently, a link is being established between the NCD and the PRDC [11], and the NCD has been shown to be noise-resistant [27].

III. EARTH-OBSERVATION IMAGES CLUSTERING AND CLASSIFICATION

A. Hierarchical Clustering

The compression-based similarity measures introduced in the previous section are a powerful tool to discover similarities within satellite data with a total data-driven model-free approach. In the first experiment, we have tested a modified version of the NCD, the NCD using smallest grammars (NCDG) [11], which performs compression by extracting the approximated smallest context-free grammars, regarded as a compact representation of the relevant information in an object. The experiment has been carried out on 60 image subsets equally divided in six classes from a labeled data set containing 600 SPOT5 single-band subsets. The NCDG has been computed between each pair of objects, generating a distance matrix. The open-source utility CompLearn [28] is then used to perform an unsupervised clustering, generating a dendrogram which fits (suboptimally) the distance matrix. The results in Fig. 1 show that all the classes are well separated with only one "false alarm." The classes fields, city, and desert are considered closer to each other, whereas clouds and sea behave in a similar way and yield the only false alarm since both of them have a simple structure and relevant portions with the same brightness. A similar approach has been tested also on a synthetic aperture radar (SAR) TerraSAR-X data set, with no false alarms in [11], confirming the universality of compression-based distance measures.

B. Classification

Compression-based similarity measures yield a good performance also in the classification tasks. In the following experiment, we have split our data set of 600 images in 200 randomly chosen training images, which are picked in equal amount among all the classes, and used the remaining 400 as test set. After building a distance vector using the NCD, we have performed classification on a simple nearest neighbor
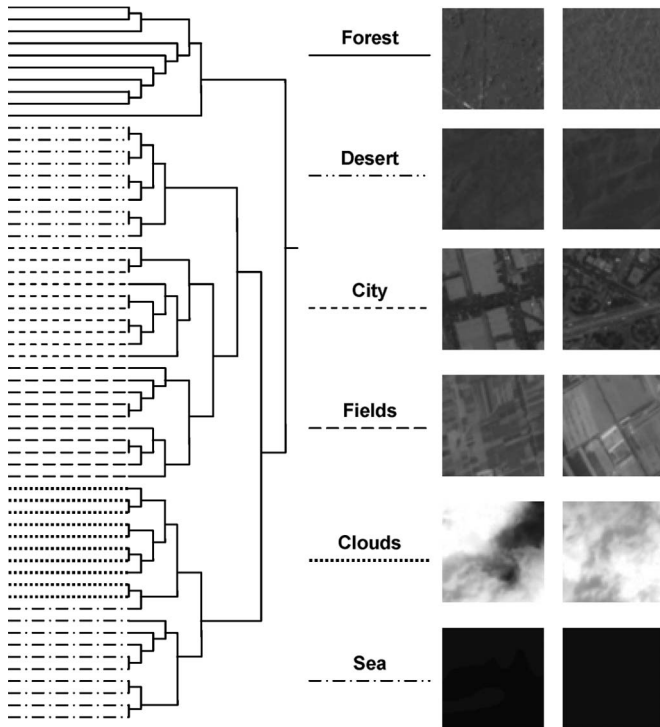
Fig. 1. (Left) Hierarchical clustering on a distance matrix containing the NCDG values applied to 60 images from the data set of which (right) a sample is reported. The classes result well separated. The only false alarm is a sea subset confused with clouds.

TABLE I
AVERAGE CLASSIFICATION ACCURACY USING DIFFERENT COMPRESSORS

LZW	LZW + Peano Scanning	JPEG2000
90.3	90.5	93.3

TABLE II
JPEG2000+NCD. CONFUSION MATRIX, NEAREST NEIGHBOR METHOD

	Clouds	Sea	Desert	City	Forest	Fields
Clouds	90.9	0	1.5	0	0	7.6
Sea	0	92.6	0	0	7.4	0
Desert	0	1.5	88	0	9	1.5
City	0	0	0	100	0	0
Forest	0	1.5	1.5	0	97	0
Fields	1.5	0	6	0	1.5	91

Average accuracy 93.3

TABLE III
JPEG2000+NCD. CLASSIFICATION ACCURACY, TOP-RANKING METHOD

Clouds	Sea	Desert	City	Forest	Fields	Average
90.9	100	98.5	100	98.5	86.5	95.7

TABLE IV
AVERAGE DISTANCES USING DIFFERENT COMPRESSORS

	NCD (LZW)	NCDG	NCD (Jpg2000)
Intra-class	1.03	0.86	0.70
Inter-class	1.1	0.98	0.90
Discrimination	0.07	0.12	0.2

basis in a 6-D space, where each dimension represents the average distance from a class. We applied the NCD with the Lempel–Ziv–Welch (LZW) algorithm [29], with and without a first step of data encoding with a space-filling Hilbert–Peano curve [30], and with lossless JPEG2000 compression [31]. The average accuracies reported in Table I show that the latter yields the best results. This is justified by the fact that the JPEG2000 compression allows keeping the vertical spatial information contained within the images, exploiting it intrinsically within the computation of the information distance, whereas a general compressor such as the LZW is limited since it linearly scans the data. We also notice that the Hilbert–Peano encoding applied before compression does not help considerably in keeping the structural information within the images. The dependence on the choice of the compressor is not a free parameter in itself, and, for each data set, a compression algorithm that is able to fully exploit the redundancies in that kind of data should be adopted [18]. Better compression, in fact, means better approximation of the Kolmogorov complexity. Nevertheless, performance comparisons for general compression algorithms have shown that this dependence is generally loose [32].

Table II shows the confusion matrix for the NCD+JPEG2000 method, whereas Table III shows what happens when, instead of the average distance from a class, we just consider the class of the top-ranked retrieved object (i.e., the closest to the query image) as in [8]. The accuracy reaches 95.7%, and an object of the correct class is retrieved within the two top ranked for 98.5% of the test set. Anyway, such a decision rule would make the classification method sensitive to potential outliers, as in the case of the class fields, which may present saturated areas or brightness similar to a desert zone, so an image representing a cloudy or desertic area could be retrieved as the best match. As a comparison, we have tried a totally different approach with support vector machine [33], using as input parameters the mean value and the variance of each image subset and performing a multiclass classification [34]. The resulting average accuracy was just 35.2%, and only the classes clouds and sea were recognized in a satisfactory way. This dramatic difference is justified by the fact that a totally feature-free approach has two advantages: It avoids underfitting the data (although, better results may be anyway obtained with the same parameters by using latent Dirichlet allocation [35]), and it skips the clustering step, which may introduce false alarms.

Compression with grammars has not been tested in this case since it is a computationally intensive procedure to be carried out offline, requiring approximately 5 s on a laptop computer to output a distance between two 64×64 tiles, so less suitable for applications on large data sets. Nevertheless, an empirical test carried out on a restricted test set of 100 images from the same data set suggests that the NCDG has better discrimination power with respect to the NCD (when a standard LZW-based compressor is used), and injection of the JPEG2000 compression in the NCD once again outperforms both (Table IV).
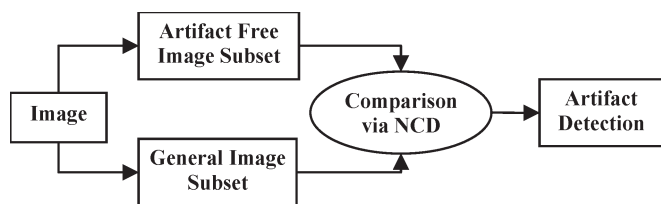
Fig. 2. Workflow for artifact detection. All image elements (size 4×4) are compared with the artifact-free elements (size 32×32) manually selected. A complexity comparison is carried out by applying the NCD to build a significant feature space. A classification or a simple thresholding is then applied to that feature space, providing a detection map.
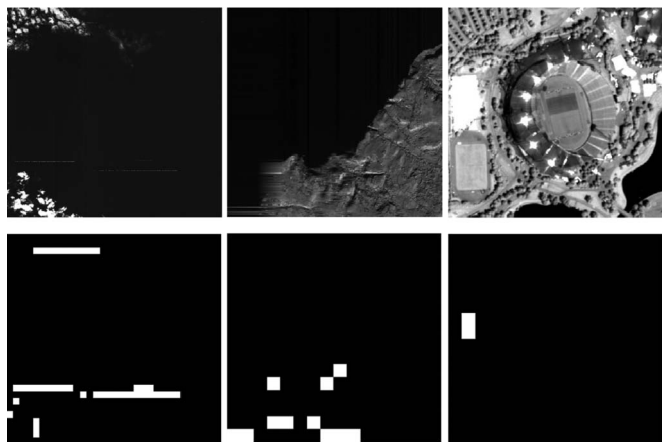

Fig. 3. Artifact detection. (Top, from left to right) SPOT image with the typical stripes of dropout artifacts (uppermost stripe not fully visible due to the reduced picture size), infrared SPOT image presenting degradation over the bottom left corner due to electronics failures in the sensor, and IKONOS saturated image. (Bottom, from left to right) Manually thresholded detections. The inconsistent areas are highlighted. (Bottom right) In the last case, only the biggest saturated area is detected because of the size of the windows employed in the algorithm.

IV. IMAGE ARTIFACT DETECTION

Satellite images may contain artifacts that decrease the quality of the images and can lead to analysis and interpretation problems. These anomalies are introduced in the processing step or coming directly from the sensors (refer to the upper row in Fig. 3) and are sometimes hard to spot visually. It would then be desirable to detect them automatically. An anomaly is not defined *in se* by the objective parameters; hence, the idea to adopt a parameter-free approach fits well the problem. Such an algorithm using the compression-based distance measures is successfully employed in [18], following a simple *divide et impera* algorithm, whereas a rate-distortion-based approach for artifact detection is described in [36]. The general idea for a methodology to automatically detect these artifacts based on the algorithmic information theory is then proposed in [37], and its workflow and the empirical results are presented here.

Generally, artifacts alter the local complexity within the images, resulting in areas with a complexity that is either too low or too high. Therefore, a similarity measure is applied to detect these defects, under the assumption that it is possible to identify some artifact-free subsets within each image. In the workflow sketched in Fig. 2, the image elements are compared via the NCD to the artifact-free image subsets, building a feature vector for each element. The classification can then be
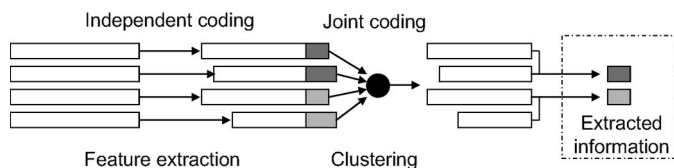

Fig. 4. Workflow to build a compressed indexed database. Dictionaries (represented in gray), containing the objects' relevant information, are extracted from the data and used to optimally compress the database.
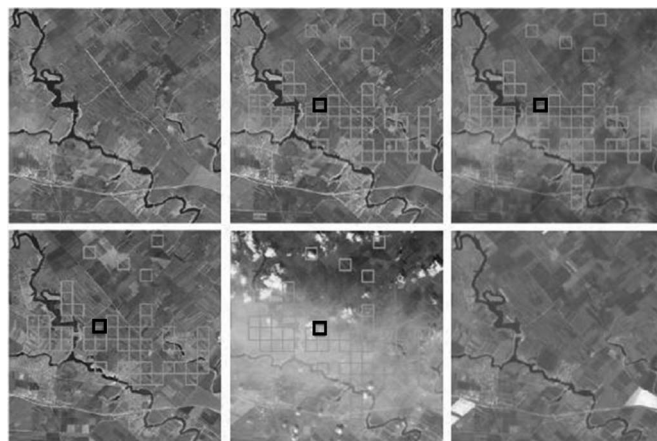

Fig. 5. Detection of fog-covered areas. This sequence of size $600 \times 600 \times 6$ is a subset of an encoded series of size $2000 \times 1500 \times 35$. The whole series contains about $25\,000$ spatiotemporal patterns. The dark square represents the query, and the light gray squares represent the query answer, i.e., the areas that evolve in a similar way and are subject to similar changes.

applied to the feature space to detect the artifacts. Lacking a ground truth database, only preliminary results are reported. Fig. 3 shows the first results obtained with this technique: The complexity comparison output is clustered with the k-means algorithm and manually thresholded, yielding the artifact detection. The results show the adaptability of this methodology to detect different types of artifact without setting any parameter, suggesting that a fully automated detection could be possible in the future.

V. MINING SITS

In [38], a method is proposed to build an index of the content of a compressed satellite image time series (SITS) database. The approach is presented in Fig. 4. First, a set of dictionaries is independently extracted from the database; then, the best ones are selected using a compression-based similarity measure to take into account the interobject correlations; finally, the selected dictionaries are used to code efficiently each object, which is thus defined by a two-part representation [39]. The dictionary is a lossy representation of the database, containing the minimal sufficient information to discriminate the objects, and it is the only information analyzed when the database is queried.

The experiments have been carried out on the SITS data set.[1] Fig. 5 shows new results for a typical query, where patches with similar evolution (in this case, covered by fog for a short period)

are retrieved. Both precision and recall [40] averages for these queries have been estimated to be 0.67, despite the high subjectivity of the results. The compression of SITS databases with the proposed method achieves then two goals: It compresses in a lossless way the images with a ratio of approximately 1 : 3, and at the same time, it enables query on the compressed database content with an acceptable precision–recall score.

VI. CONCLUSION

This letter has presented a recent approach to image analysis based on algorithmic information theory, along with different applications based on these concepts. Clustering may be carried out to estimate the number of classes in an unsupervised way, without any *a priori* knowledge of the data, and the compression-based similarity measures give a very good performance also in classification tasks and can successfully be applied to perform artifact detection and joint compression and mining of SITS. These methodologies do not need any parameter definition or *a priori* knowledge of the data and skip the feature extraction step; thus, satellite images varying greatly in content, resolution, and also sensor type may be analyzed using the same tools. This would be particularly valuable for new applications in the field of image information mining. Our perspectives for the future include exploiting these techniques to perform a joint semantic coding of data and extracted features [41] in a single step.

ACKNOWLEDGMENT

REFERENCES

[1] M. Li and P. M. B. Vitányi, *An Introduction to Kolmogorov Complexity and Its Applications*. Berlin, Germany: Springer-Verlag, 1997.

[2] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 1948. 623–656.

[3] G. Chaitin, *Information, Randomness and Incompleteness: Papers on Algorithmic Information Theory*. Singapore: World Scientific, 1990.

[4] P. Grünwald and P. M. B. Vitányi, "Shannon information and Kolmogorov complexity," *IEEE Trans. Inf. Theory*, to be published. [Online]. Available: http://www.cwi.n1/~paulv/papers/info.pdf

[5] G. Chaitin, *Meta Math! The Quest for Omega*. New York: Pantheon, 2005.

[6] C. H. Bennett, P. Gács, M. Li, P. Vitányi, and W. Zurek, "Information distance," *IEEE Trans. Inf. Theory*, vol. 44, no. 4, pp. 1407–1423, Jul. 1998.

[7] M. Li, X. Chen, X. Li, B. Ma, and P. M. B. Vitányi, "The similarity metric," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3250–3264, Dec. 2004.

[8] D. Benedetto, E. Caglioti, and V. Loreto, "Language trees and zipping," *Phys. Rev. Lett.*, vol. 88, no. 4, p. 048 702, Jan. 2002.

[9] T. Watanabe, K. Sugawara, and H. Sugihata, "A new pattern representation scheme using data compression," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 579–590, May 2002.

[10] K. Sugawara and T. Watanabe, "Classification and function prediction of proteins using data compression," *Artif. Life Robot.*, vol. 6, no. 4, pp. 185–190, Dec. 2002.

[11] D. Cerra and M. Datcu, "Image classification and indexing using data compression based techniques," in *Proc. IGARSS*, Boston, MA, 2008, pp. I-237–I-240.

[12] D. Cerra and M. Datcu, "Algorithmic cross-complexity and conditional complexity," in *Proc. DCC*, 2009, pp. 342–351.

[13] D. Cerra and M. Datcu, "A model conditioned data compression based similarity measure," in *Proc. DCC*, 2008, p. 509.

[14] R. Cilibrasi and P. M. B. Vitányi, "Clustering by compression," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1523–1545, Apr. 2005.

[15] R. Cilibrasi, P. Vitányi, and R. de Wolf, "Algorithmic clustering of music based on string compression," *Comput. Music J.*, vol. 28, no. 4, pp. 49–67, Dec. 2004.

[16] A. Apostolico, F. Cunial, and V. Kaul, "Table compression by record intersections," in *Proc. DCC*, 2008, pp. 13–22.

[17] M. Li, J. H. Badger, X. Chen, S. Kwong, P. Kearney, and H. Zhang, "An information-based sequence distance and its application to whole mitochondrial genome phylogeny," *Bioinformatics*, vol. 17, no. 2, pp. 149–154, 2001.

[18] E. J. Keogh, S. Lonardi, and C. Ratanamahatana, "Towards parameter-free data mining," in *Proc. SIGKDD*, 2004, pp. 206–215.

[19] Q. Liu, Y. Yang, C. Chen, J. Bu, Y. Zhang, and X. Ye, "RNACompress: Grammar-based compression and informational complexity measurement of RNA secondary structure," *BCM Bioin.*, vol. 9, p. 176, Mar. 2008.

[20] E. J. Keogh, S. Lonardi, C. Ratanamahatana, L. Wei, S. Lee, and J. Handley, "Compression-based data mining of sequential data," *Data Min. Knowl. Discov.*, vol. 14, no. 1, pp. 99–129, Feb. 2007.

[21] A. Cohen, C. Bjornsson, S. Temple, G. Banker, and B. Roysam, "Automatic summarization of changes in biological image sequences using algorithmic information theory," *IEEE Trans. Pattern Anal. Mach. Intell*. to be published. [Online]. Available: http://ieeexplore.ieee.org/xpl/preabsprintf.jsp?arnumber=4544515. Abstract.

[22] N. K. Vereshchagin and P. Vitányi"Kolmogorov's structure functions and model selection," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3265–3290, Dec. 2004.

[23] X. Chen, B. Francia, M. Li, B. McKinnon, and A. Seker, "Shared information and program plagiarism detection," *IEEE Trans. Inf. Theory*, vol. 50, no. 7, pp. 1545–1551, Jul. 2004.

[24] C. Bennett, M. Li, and B. Ma, "Chain letters and evolutionary histories," *Sci. Amer.*, vol. 288, no. 6, pp. 76–81, Jun. 2003.

[25] R. L. Cilibrasi and P. Vitányi, "The Google similarity distance," *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 3, pp. 370–383, Mar. 2007.

[26] D. Sculley and C. E. Brodley, "Compression and machine learning: A new perspective on feature space vectors," in *Proc. DCC*, 2006, pp. 332–341.

[27] M. Cebrian, M. Alfonseca, and A. Ortega, "The normalized compression distance is resistant to noise," *IEEE Trans. Inf. Theory*, vol. 53, no. 5, pp. 1895–1900, May 2007.

[28] R. Cilibrasi, A. Cruz, S. de Rooij, and M. Keijzer, *CompLearn Tool*. [Online]. Available: http://www.complearn.org/index.html

[29] J. Ziv and A. Lempel, "Compression of individual sequences via variable-rate coding," *IEEE Trans. Inf. Theory*, vol. 24, no. 5, pp. 530–536, Sep. 1978.

[30] G. Peano, "Sur une courbe, qui remplit toute une aire plane," *Math. Ann.*, vol. 36, no. 1, pp. 157–160, Mar. 1890.

[31] *Jpg2k for Java*. [Online]. Available: http://jj2000.epfl.ch/

[32] M. Cebrian, M. Alfonseca, and A. Ortega, "Common pitfalls using the normalized compression distance: What to watch out for in a compressor," *Commun. Inf. Syst.*, vol. 5, no. 4, pp. 367–384, 2005.

[33] T. Joachims, B. Schölkopf, C. Burges, and A. Smola, *Making Large-Scale SVM Learning Practical. Advances in Kernel Methods—Support Vector Learning*. Cambridge, MA: MIT Press, 1999.

[34] T. Joachims, *Multi-Class Support Vector Machine*. [Online]. Available: http://www.cs.cornell.edu/people/tj/svm_light/svm_multiclass.html

[35] M. Liénou, H. Maître, and M. Datcu, "Semantic annotation of satellite images using latent Dirichlet allocation," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 4, Oct. 2009, to be published.

[36] A. Mallet and M. Datcu, "Rate distortion based detection of artifacts in Earth observation images," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 3, pp. 354–358, Jul. 2008.

[37] A. Mallet and M. Datcu, "Complexity based image artefact detection," in *Proc. DCC*, 2008, p. 534.

[38] L. Gueguen and M. Datcu, "A similarity metric for retrieval of compressed objects: Application for mining satellite image time series," *IEEE Trans. Knowl. Data Eng.*, vol. 20, no. 4, pp. 562–575, Apr. 2008.

[39] J. J. Rissanen, "A universal data compression system," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 5, pp. 656–664, Sep. 1983.

[40] R. Baeza-Yates and B. Ribeiro-Neto,in *Modern Information Retrieval*. New York: ACM Press, 1999.

[41] E. Kokiopoulou and P. Frossard, "Semantic coding by supervised dimensionality reduction," *IEEE Trans. Multimedia*, vol. 10, no. 5, pp. 806–818, Aug. 2008.