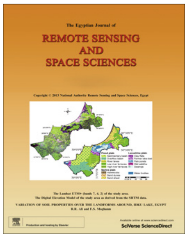
National Authority for Remote Sensing and Space Sciences

The Egyptian Journal of Remote Sensing and Space Sciences

www.elsevier.com/locate/ejrs
www.sciencedirect.com



RESEARCH PAPER

A web based semi automatic frame work for astrobiological researches

P.V. Arun *

National Institute of Technology, Bhopal, India

Received 28 June 2013; revised 27 September 2013; accepted 3 November 2013

KEYWORDS

Astrobiology;
Extra-terrestrial;
Cellular Automata;
Signature

Abstract Astrobiology addresses the possibility of extraterrestrial life and explores measures towards its recognition. Researches in this context are founded upon the premise that indicators of life encountered in space will be recognizable. However, effective recognition can be accomplished through a universal adaptation of life signatures without restricting solely to those attributes that represent local solutions to the challenges of survival. The life indicators should be modelled with reference to temporal and environmental variations specific to each planet and time. In this paper, we investigate a semi-automatic open source frame work for the accurate detection and interpretation of life signatures by facilitating public participation, in a similar way as adopted by SETI@home project. The involvement of public in identifying patterns can bring a thrust to the mission and is implemented using semi-automatic framework. Different advanced intelligent methodologies may augment the integration of this human machine analysis. Automatic and manual evaluations along with dynamic learning strategy have been adopted to provide accurate results. The system also helps to provide a deep public understanding about space agency's works and facilitate a mass involvement in the astrobiological studies. It will surely help to motivate young eager minds to pursue a career in this field.

1. Introduction

Exploration of extra-terrestrial intelligence (ETI) requires universal construal towards life signatures in a way that will help us to recognize biospheres quite different from our own. Astrobiological explorations that address various possibilities in this context have been based on the potential recognition of life indicators (Chyba and Phillips, 2001). Literature reveals a great deal about various related missions and their scopes (Hoover, 2011). The Prospecting Autonomous Nano Technology Swarm Mission (PAM) is an advanced mission concept for the 2020s that seeks to map surface characteristics and other properties of the asteroid Main Belt (Curtis et al., 2003). Recent discoveries of around 10,000 Near Earth Objects (NEOs) suggest the possible investigations over these bodies for exploring their constituents. The limited availability of direct samples

---

* Tel.: +91 4828249999.
E-mail address: arunpv2601@gmail.com

Peer review under responsibility of National Authority for Remote Sensing and Space Sciences.

Production and hosting by Elsevier

1110-9823 © 2013 Production and hosting by Elsevier B.V. on behalf of National Authority for Remote Sensing and Space Sciences.
http://dx.doi.org/10.1016/j.ejrs.2013.11.002

Please cite this article in press as: Arun, P.V., A web based semi automatic frame work for astrobiological researches, Egypt. J. Remote Sensing Space Sci. (2013), http://dx.doi.org/10.1016/j.ejrs.2013.11.002

in these contexts has made us to rely primarily on remote detection methods, however processing of large bulks of data produced in these missions requires effective automation.

Usually referred distinguishing characteristics of life such as metabolism, growth, reproduction and adaptation (Cavicchioli, 2002), cannot be used as remote sensing parameters due to their limited spatial and temporal scale (Schulze-Makuch et al., 2002). Life may exist in a form unknown to us as it may occur beneath an opaque surface or may be too small to cause environmental transformations or may be insufficiently intricate to generate complex phenomena, such as roads or radio waves (Schulze-Makuch and Irwin, 2008). Under these circumstances, more sophisticated and abstract definitions of life alluded to the above must be used for alternative set of parameters that could point to conditions favourable for generic forms of life (Victor and Schulze-Makuch, 2004). Direct consequences of biological activity (bio-signatures) and alterations of the geological environment due to biological processes (geo-signatures) are usually employed as life signatures.

Bio-geo signatures should reflect fundamental and universal characteristics of life, and thus are not restricted solely to those attributes that represent local solutions to the challenges of survival. Automatic modelling, similarity deduction and extra-terrestrial considerations are required for interpretation of the imageries in this regard (Chyba and Phillips, 2001). Advanced learning and random modelling approaches can be used to dynamically model signatures with reference to the specific conditions.

The data from different spectral sources have to be intelligently interpreted for effective decisions which make manual approach literally inadequate. However complete automatic systems are deficient of rational thinking capability and creativity of human expertise. Involvement of public in identifying the patterns can bring a thrust to the mission however lack of skill may be a matter of concern. This can be made good using semi-automatic approaches where man and machine will work side by side, integrating the rational nature of former with the acquired skill of latter. Semi-automatic approaches also resolve the human tendency to assume facial resemblance (Hoover, 2011). Image data made available through web portal facilitate public involvement as in the case of SETI@home project.

In this paper, we investigate the feasibility of a semi-automatic open source frame work for accurate bio-geo signature detection using effective public participation. We have focussed over semi-automatic enhancement of methodologies to distinguish artificial and natural structures, and have evaluated the efficiency over controversial traditional datasets (say, 35A72 and 70A13). Proposed methodology will explore reliable semi-automatic recognition of basic geological elements under varying conditions, with a view to produce informative, concise summaries of science observations and to guide spot analyses at sites of detailed study. Different advanced intelligent methodologies enhance the integration of these human machine analyses.

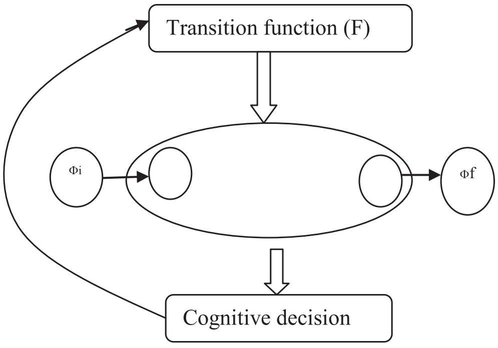


Figure 1 Schematic representation of CTSLA.

2. Bio-geo signatures

Bio-geo signatures of life which can be readily detected and validated on Earth from space are usually adopted as life indicators (Marais et al., 2003). Commonly used life indicators whose consequences can be detected remotely as described in literatures (Schulze-Makuch et al., 2002; Sudhir et al., 2010; Dawyndt et al., 2005; Russell et al., 1999) are analysed and a possible way of their remote detection is summarized in Tables 1 and 2.

3. Theoretical background

The frame work has been implemented using various intelligent methodologies and advanced web mining techniques. Different methodologies namely cognitive networks (Kandasamy and Smardandache, 2003; Ziaei and Hajizade, 2011), classifiers (Huang et al., 2002; Lee et al., 2005), and evolutionary computing techniques have been employed for the purpose. Cognitive variation of Learning Automata (LA) (Arun and Katiyar, 2013b), namely Cognitive Time Specific LA (CTSLA) has been proposed for effective modelling of life indicators.

Evolutionary computing approaches such as Cellular Automata and their variants such as Cellular Neural Network (CNN) and Multiple Attractor Cellular Automata (MACA) have been found to be useful for modelling random signatures. CNN (Arun and katiyar, 2013a) is effectively used for modelling object shape to facilitate feature interpretation. MACA is a special type of CA that converges to certain attractor states on execution (Arun and katiyar, 2013b) and is employed to identify classes of patterns for object interpretation. N-Dimensional classifiers such as Support Vector Machines have been used along with the kernel functions to implement initial clustering for accurate detection. Mixture density kernels have been used to integrate an adaptive kernel strategy to the Support Vector Random Field (SVRF) based clustering as it facilitates the learning of kernels directly from image data (Srivastava, 2004).

NLP parsers along with WordNet have been used for query interpretation and dynamic updation of probabilistic rules. CTSLA is an enhanced modification of LA to facilitate the incorporation of a dynamic learning strategy. It also enables to distinguish timely variation of inputs and implement automatic modelling better than its traditional counterparts. Schematic representation of the model is presented in Fig. 1.

As shown in the figure F is dynamically updated based on expected and acquired states, thus implementing a cognitive dynamism. The CTSLA model has been proposed to incorporate effective modelling of life indicators based on temporal and planetary environmental variations. Neutrosophic Cognition Techniques (Kandasamy and Smardandache, 2003)

Please cite this article in press as: Arun, P.V., A web based semi automatic frame work for astrobiological researches, Egypt. J. Remote Sensing Space Sci. (2013), http://dx.doi.org/10.1016/j.ejrs.2013.11.002

Table 1 Geo signatures.

S. No	Geo signature	Remote detection	Remarks
1	Optimal composition of atmosphere	Hyper spectral remote sensing in combination with the microwave approach	Medium for dynamic energy gradients, affords a stabilizing and protective shield, and may help the presence of liquid
2	Flow of energy	Thermal remote sensing (heat), optical remote sensing (light), surface mapping (tectonic and internal differentiation)	To organize its material substance and maintain its low entropic state
3	Liquid medium	Detected by radar and the absorption spectrum of water but not when it is present in the deep subsurface or shielded by a thick layer of ice	Concentrating without immobilizing interacting constituents within a bounded environment Water–ammonia–organic compound mixtures can also provide a medium as they can exist as liquid at very low temperature
4	Chemical complexity-chemical cycling	Polymeric organic compounds and chemical cycling can be inferred from absorption spectra and gradients in surface colouration Detection of alteration minerals indicates geochemical cycling	Chemical cycling on Earth is known to occur through oxidation–reduction cycles that are actively maintained by organisms or that occur inorganically
5	Tectonic activities	Tectonics can be identified based on measured magnetic properties of the rock, visible symmetry along a spreading axis, and specific patterns in fracture orientation and propagation	The recycling of nutrients caused by tectonic movements is required for the sustainment of life Plate tectonics on Earth also constantly produce greenhouse gases that acted as a global thermostat providing stability for the evolution of life

Table 2 Bio signature.

S. No	Bio signature	Remote detection	process
1	Optimal atmospheric gas composition	Detected remote sensing of its absorption spectrum	The disequilibrium associated with high amounts of oxygen can be used as an indicator for oxygen producing photoautotrophs
2	Chemosynthesis	Radar is used for mapping of topographic and geomorphologic characteristics, including even surface roughness Stereoscopic methods can be used for enhancing the detection limit of surface features	Chemoautotrops produced exhibit large-scale geomorphological characteristics such as stromatolite colonies and coral reefs (e.g. Great Barrier reef of moon)
3	High rates of erosion	Visible and microwave wavelengths of the electromagnetic spectrum	Erosion observed on Earth due to biological and chemical weathering induced by living organisms (Fungus-lichen rock)
4	Biogenic macromolecules	Radiance spectra in the visible region and by advanced very high resolution radiometer measurements	Property of terrestrial biogenic molecules
6	Structural complexity	High Resolution imagery	Represents presence of life
7	Biogenic heat	Thermal Remote sensing	Biogenic heat liberated by continually drawing and release of energy by the living beings

Please cite this article in press as: Arun, P.V., A web based semi automatic frame work for astrobiological researches, Egypt. J. Remote Sensing Space Sci. (2013), http://dx.doi.org/10.1016/j.ejrs.2013.11.002

enable the modelling of indeterminate conditions and are effectively used for modelling relation between signature components and their interpretations. Fuzzy AHP is a synthetic extension of the classical AHP method where the fuzziness of decision makers is considered for evaluating complex multi attribute alternatives (Ziaei and Hajizade, 2011). Cognitive maps along with the Fuzzy AHP approach have been employed for proper integration of signature components. Client–server architecture using Lamp server, PHP, Phython and oracle has been adopted for implementing a user friendly interactive framework.

4. Methodology

Advanced web mining and artificial intelligence techniques have been used to provide a semi-automatic framework for effective signature detection. Schematic representation of adopted methodology is presented in Fig. 2. Multilayer CTSLA has been adopted in which each of the layers is being dedicated to a signature. Let the CTSLA be given as $[\varphi, \alpha, \beta, A, G]$ where φ is the state of automaton, α is the output, β is the input, A is the learning algorithm, and G is the output function. Let $p_j(n)$ be the probability that automaton is in state j at iteration n, then the reinforcement scheme, if $\alpha(n) = \alpha i$ and for $j \neq i$; ($j = 1$ to r), is given by $pj(n + 1) = pj(n) - g\ [pj(n)]$ when $\beta(n) = 0$ && $pj(n + 1) = pj(n) + h\ [pj(n)]$ when $\beta(n) = 1$. In order to preserve probability measure, $\sum pj(n) = 1$, for $j = 1$ to r, if $\alpha(n) = \alpha i$, reinforcement is modified as $pi(n+1) = pi(n) + \sum_{j=1}^{r} g(P_j(n))$ when $\beta(n) = 0$ and $pi(n+1) = pi(n) - \sum_{j=1}^{r} g(P_j(n))$ when $b(n) = 1$, where $g(.)$ is the reward function and $h(.)$ is the penalty function.

Image data as well as user interactions are made available through 3-tier client server architecture. Public participations have been facilitated as in the case of SETI@home and different intelligent methodologies augment the integration of these human–machine analyses. Users classify the data based on signature content and semi-automatic environment is provided to increase the effectiveness. A prior level accuracy checking has also been automated to enable the system to improve itself. The users have been ranked based on their accuracy and actions of high skilled users are used for improving the machine performances. The data marked as high signature content by lower skilled users are automatically cross validated and are moved to higher category users for thorough high skilled analysis. The semi-automatic operations facilitated by the system are detailed below.

4.1. Extraction of components

Different signature components are effectively segmented in a semi supervised frame work where users are facilitated to interactively provide various segmentation parameters such as number of classes, seed locations, stopping conditions etc. Initial clustering is accomplished using Support Vector Random Field approach that uses composite mixed density kernel function. Composite kernel concept incorporates spectral and spatial information to represent contextual information. The adaptive mixture density kernels are exploited to dynamically adjust the kernel parameters in accordance with the data distribution. Detected objects along with boundary information are optimized using the coreset approach (Agarwal et al., 2001) to reduce the complexity of shape modelling.

4.2. Component interpretation

Training phase of feature interpretation facilitates users to interactively model random signature components with reference to terrain parameters (tone, texture, type, and shape). Clustered objects along with edge information are exploited using CNN and MACA to model various components. CNN rules corresponding to a particular feature are used to distinguish it and these rules along with possible variation thresholds are stored in Prolog DB for effective interpretation. Deviation of feature geometry from fractal model is also calculated to provide accurate description of these features.

Detection phase involves interpretation of features using definitions acquired during training. Probabilistic rules corresponding to different signature components are stored in the form of grammar productions and are updated dynamically. MACA is initialized with an unknown pattern and operated for a maximum (depth) number of cycles until it converges to an attractor. PEF bits after convergence are extracted to identify the class of the pattern and are compared with stored rules to interpret the object. Probabilistic rules are used to produce most likely predictions based on the previous experiences.

4.3. Component interpolation

Interpolations of signature components are accomplished using CA rules integrated with stored predicate rules. For example, given a feature such as a river, training experience or stored rules are used to set the threshold value for particular

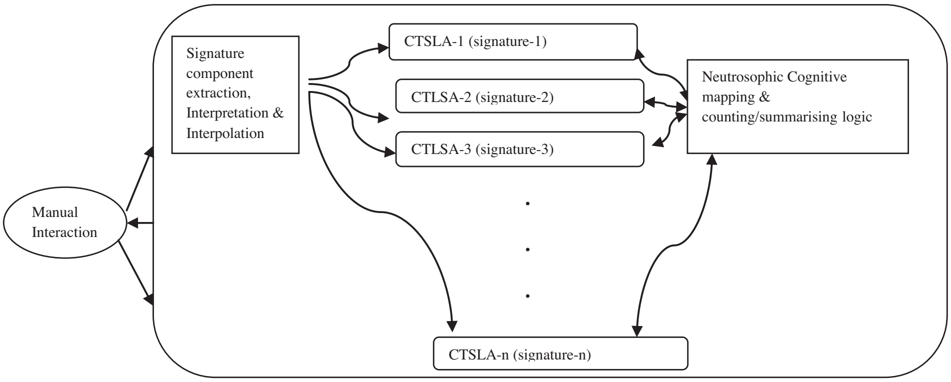


Figure 2 Schematic representation of the system.

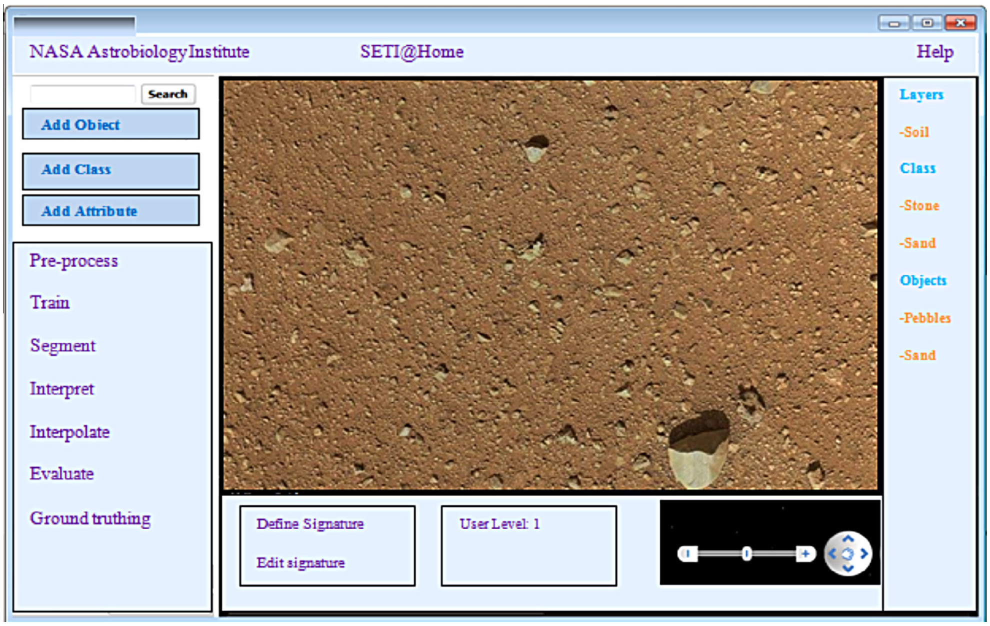
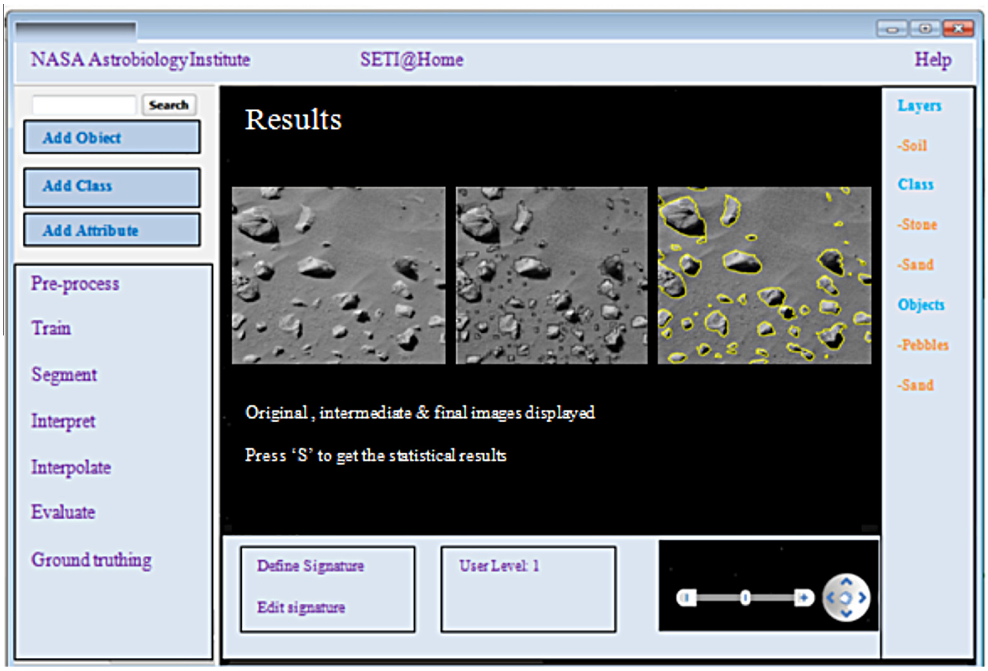


Figure 3 Basic user interface.



Original, intermediate & final images displayed

Press 'S' to get the statistical results

Figure 4 Results.

evolution rule to be adopted. PCFG rule sets are used to govern the topology extraction and the relative positions are determined based on the coordinate information associated with each feature. Comparisons of boundary pixel positions are adopted for determining relative positions of random features. Topology information, along with simple spatial buffering, is explored to assess the proximity queries.

4.4. Signature formulation

Different components extracted from images are intelligently integrated to form required life indicators. A predicate rule database is also maintained for guiding proper combination of signature components. Effective dynamic modelling of various components with reference to time as well as planetary conditions is accomplished using CTSLA. The transition functions are changed based on cognitive decisions which in turn depend on the achieved states with reference to a particular time domain. Hence a dynamically updating cognitive network is created which can dynamically model the signatures with reference to the environmental variations.

N-bit codes based on PEF bit pattern of MACA facilitate the representation of signature components. Binary representations of individual components are combined (say XOR) to get the signature code and each signature is modelled using a CTSLA. Thus binary representation of a life indicator reveals information about its components and can incorporate temporal variations. These signature codes along with their details are stored in prolog DB and are organized using a fuzzy AHP approach.

4.5. Natural language interpretation

NLP interface enables to interactively update various rule formulations and component definitions. PCFG grammar formulations have been adopted for different rule sets and are dynamically updated based on training. We have employed Stanford parser along with Wordnet for syntactic as well as lexical analyses to facilitate effective interactions.

5. Results and discussions

Client server model has been adopted for implementation, college level students from engineering as well as science back grounds, having no prior astronomical knowledge were allowed to use the system, and have been found to be successful. System has been evaluated with students of Royal University of Bhutan (Bhutan), Cochin University (India), and Pokhara University (Nepal). Students accepted the model with wide enthusiasm and have participated effectively in the attempt. Investigations over different traditional image sets have revealed that considerable success has been achieved with the procedure.

Fig. 3 shows basic interface of the system. Pre-processing facilitates the removal of noises and sensor effects from the images. Training enables users to provide samples to increase detection accuracy whereas a semi-supervised clustering is facilitated through segmentation. Ground truthing enables to provide metadata that can be used for accuracy evaluation. Provisions are also provided for editing and defining signature formulation manually. Summary tab is also provided to list the different classes, objects and layers so far defined with reference to the displayed image. Availability of multiple images has been explored to provide effective classification along with enhanced zooming (as in the case of Google Earth).

Framework also facilitates the incorporation of ground truthing information to measure accuracy of user tasks. Lack of the data has limited the accuracy of evaluation; however it is compensated using relative analysis based on highly skilled user inputs. Provisions have also been provided to enable effective evaluation through manual interpretation. The performance accuracy is statistically evaluated using parameters such as kappa statistics, over-all efficiency as well as producer and consumer accuracy (Arun and Katiyar, 2013). Report based on confusion matrix is also provided to improve classification accuracy. Fig. 4 shows the results of stone detection where the third image shows a filtered output based on size.

Different traditional datasets have been considered to validate the effectiveness of the method. The controversial Viking image samples of 1976 that resembled humanoid faces (35A72 and 70A13) have been accurately resolved using the approach. The framework has been successful in resolving many natural structures that otherwise ordinary human interpretation may similarize to artificial ones. Applicability of the framework in other areas has been investigated with reference to public health; Spatial analysis framework has been implemented in NIMHANS, India where terrestrial images have been used to model epidemic related issues.

The main disadvantage of the method is its computational complexity which has been improved through coreset optimization and similar approximation techniques. Complexity can be further reduced by storing the detected rule variations; optimization methods such as GA can be exploited to optimize the strategy. This research provides a basic framework and further investigations are needed to optimize it. Integration of a fuzzy approach to the inverse mapping also seems to be promising, since fuzzy/neutrosophic cognitive maps can be exploited for effectively organizing and selecting CA rules. The PCFG grammar update approach also needs further improvement especially in the context of topological attributes. Efficiency of the system needs to be improved through incorporation of more signature definitions and has to be trained with more samples. Distributed implementation of the system may result in considerable reduction of complexity.

6. Conclusion

In this paper we investigated the feasibility of a semi-automatic framework for augmenting analysis of extra-terrestrial remote sensing images. The proposed approach adopts dynamic modelling techniques along with manual interpretation to interactively analyse large bulk of ET data. We have also proposed a cognitive variation of LA (called CTSLA) to effectively model life indicators. The proposed methodology seems to be capable of handling various issues available in the automation of astrobiological techniques. The accuracy assessment of these methodologies has been difficult due to the unavailability of ground truthing or reference data. Generalization of this framework has been investigated in public health domain and has been found to be an effective tool for epidemic modelling. The dynamic learning network adopted in this approach can be further improved by incorporating the parallelism approach. Further research is required over the optimization as well as parallelization of this framework.

References

Agarwal, P.K., Aronov, B., Sharir, M., 2001. Exact and approximation algorithms for minimum-width cylindrical shells. Discrete Comput. Geom. 26 (3), 307–320.

Arun, P.V., Katiyar, S.K., Cellular Automata based approach towards automatic object extraction and shape modelling. J. GI Sci. Remote Sens., Accepted and published online, July 2013.

Arun, P.V., Katiyar, S.K., An evolutionary computing based approach towards automatic object extraction. Egypt. J. Remote Sens., Accepted and published online, September 2013.

Please cite this article in press as: Arun, P.V., A web based semi automatic frame work for astrobiological researches, Egypt. J. Remote Sensing Space Sci. (2013), http://dx.doi.org/10.1016/j.ejrs.2013.11.002

Cavicchioli, R., 2002. Extremophiles and the search for extraterrestrial life. Astrobiology 2, 281–292.

Chyba, F.F., Phillips, C.B., 2001. Possible ecosystems and the search for life on Europa. Proc. Natl. Acad. Sci. USA 98, 801–804.

Curtis, S.A., Rilee, M.L., Clark, P.E., Marr, G.C., 2003, Use of swarm intelligence in spacecraft constellations for the resource exploration of the asteroid belt. Proceedings 3rd International Workshop on Satellite Constellations and Formation Flying, Pisa, Italy.

Dawyndt, P., Vancanneyt, M., De Meyer, H., Swings, J., 2005. Knowledge accumulation and resolution of data inconsistencies during the integration of microbial information sources. IEEE Trans. Knowl. Data Eng. 17 (8), 1111–1126.

Hoover, R.B., 2011. Fossils of cyanobacteria in CI1 carbonaceous meteorites. J. Cosmol. 13 (2), 120–125.

Huang, C., Davis, L.S., Townshend, J.R.G., 2002. An assessment of support vector machines for land cover classification. Int. J. Remote Sens. 23 (4), 725–749.

Kandasamy, W.B.V., Smardandache, F., 2003. Fuzzy cognitive maps and neutrosophic cognitive maps. Phoenix: Xiquan 23, 211.

Lee, C.H., Schmidt, M., Greiner, R., 2005. Support vector random fields for spatial classification. 9th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD), Portugal, pp.170–175.

Marais, D.J., Louis, J.A., Louis, J.A., Steven, A.B., Alan, P.B., David, D., Paul, G.F., 2003. The NASA astrobiology roadmap. Astrobiology 3 (2), 219–235.

Russell, M.J., Ingham, J.K., Zedef, V., Maktav, D., Sunar, F., Hall, A.J., Fallick, A.E., 1999. Search for signs of ancient life on Mars: expectations from hydromagnesite microbialites, Salda Lake, Turkey. J. Geolog. Soc. London 156, 869–888.

Schulze-Makuch, D., Irwin, L., 2008. Life in the Universe: Signatures of Life, vol. 3. Springer, Berlin/Heidelberg (pp.165–182).

Schulze-Makuch, D., Irwin, L.N., Guan, H., 2002. Search parameters for the remote detection of extraterrestrial life. Planet. Space Sci. 50, 675–683.

Srivastava, A.N., 2004. Mixture Density Mercer Kernels: A Method to Learn Kernels Directly from Data. Accepted in SLAM Data Mining Conference.

Sudhir, K.J., Abdul, A.K., Mahendra, K.R., 2010. Geomicrobiology. Sci. Publ. 2010, 175–208.

Victor, A.G., Schulze-Makuch, D., 2004. Genetic code: lucky chance or fundamental law of nature? Phys. Life Rev. 1 (3), 202–229.

Ziaei, M., Hajizade, F., 2011. Fuzzy analytical hierarchy process (FAHP): a GIS-based multicriteria evaluation/selection analysis. 19th International Conference on Geoinformatics, 19 (11), 1–6, 24–26.

Further reading

Marion, G.M., Fritsen, C.H., Eicken, H., Payne, M.C., 2003. The search for life on Europa: limiting environmental factors, potential habitats, and Earth analogues. Astrobiology 3, 785–811.

Please cite this article in press as: Arun, P.V., A web based semi automatic frame work for astrobiological researches, Egypt. J. Remote Sensing Space Sci. (2013), http://dx.doi.org/10.1016/j.ejrs.2013.11.002