GEOSPATIAL IMAGE MINING FOR NUCLEAR NONPROLIFERATION DETECTION: CHALLENGES AND NEW OPPORTUNITIES

Ranga Raju Vatsavai[1], *Budhendra Bhaduri*[1,*], *Anil Cheriyadat*[1], *Lloyd Arrowood*[2], *Eddie Bright*[1], Shaun Gleason[1], Carl Diegert[5], Aggelos Katsaggelos[7], Thrasos Pappas[7], Reid Porter[8], Jim Bollinger[3], Barry Chen[4], Ryan Hohimer[6]

[1]Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA
[2] Y-12 National Security Complex, Oak Ridge, TN 37831, USA
[3] Savannah River National Laboratory, Aiken, SC 29808, USA
[4] Lawrence Livermore National Laboratory, Livermore, CA 94550, USA
[5] Sandia National Laboratories, Albuquerque, NM 87123, USA
[6] Pacific Northwest National Laboratory, Richland, WA
[7] Argonne National Laboratory, Argonne, IL 60439, USA
[8] Los Alamos National Laboratory, New Mexico, USA

1. INTRODUCTION

With increasing understanding and availability of nuclear technologies, and increasing persuasion of nuclear technologies by several new countries, it is increasingly becoming important to monitor the nuclear nonproliferation activities. There is a great need for developing technologies to automatically or semi-automatically detect nuclear proliferation activities using remote sensing technologies. Images acquired from earth observation satellites is an important source of information in detecting nonproliferation activities. High-resolution remote sensing images are highly useful in verifying the correctness and as well as completeness of any countries nuclear program. DOE national laboratories were interested in detecting nuclear proliferation by developing advanced geospatial image mining algorithms. In this paper we describe the current understanding of geospatial image mining techniques and enumerate key gaps and identify future research needs in the context of nuclear nonproliferation.

2. GEOSPATIAL IMAGE MINING

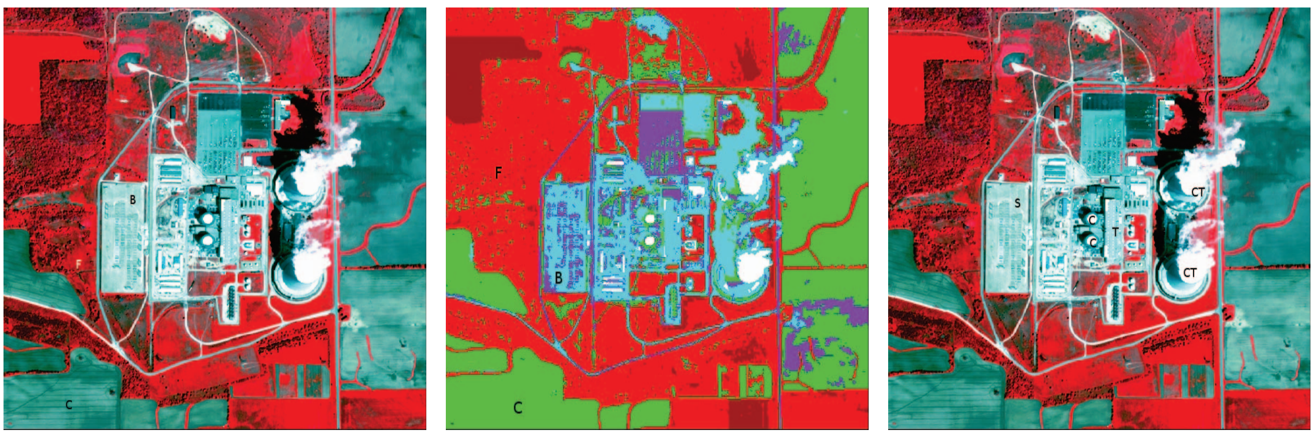
Increasing resolution, volume, and availability of remote sensing imagery made it possible to accurately identify key geospatial features and their changes over time. Recent studies have shown the usefulness of remote sensing imagery for monitoring nuclear safeguards and proliferation activities [1]. Classification is one of the widely used technique for extracting thematic information. Classification is often performed on per-pixel basis, however proliferation detection requires identification of complex objects, patterns and their spatial relationships. One key distinguishing feature as compared to traditional thematic classification is that the objects and patterns that constitute a nuclear facility have interesting spatial relationships (metric, topological, etc) among themselves. These limitations are clearly evident from Figure 1. Classification technology is matured for extracting thematic classes such as buildings, forest, crops, etc. However, such thematic labels are not enough to capture

*Contact: bhaduribl@ornl.gov

<table>
<tr><th colspan="6">Key thematic areas in geospatial and spatiotemporal data mining</th></tr>
<tr><th rowspan="2">Geospatial Data Characteristics</th><th rowspan="2">Geospatial Ontology</th><th colspan="2">Geospatial Image Mining</th><th rowspan="2">Geospatial Video Analysis</th></tr>
<tr><th>Feature Extraction</th><th>Semantics Extraction</th></tr>
<tr>
<td>• Available Sensors: Optical, thermal, IR, Radar (Note: the scope of this project is restricted to optical imagery)
• Spatial Resolution: Coarse (30m) to Fine (sub-meter)
• Temporal Resolution: Daily to 22 days
• Acquisition: Offline vs. One-demand (with targeting capabilities)
• Modes: Ground, Air, Satellites</td>
<td>• Number of well-established geospatial ontologies - SWEET, EnvO, Ordnance Survey ontologies, . . .
• A number of standards for geospatial data: GML, SDTS, ISO/OGC specifications
• Gazetteers and feature catalogs by national mapping agencies</td>
<td>• A number of low-level feature extraction techniques, including scale invariance (SIFT) image features exist</td>
<td>• Mostly pixel-based supervised classification approaches
• Recently Object-based classification has gained popularity
• Most approaches are bottom-up
• (Ontology driven) top-down and generative models like LDA, probabilistic, and graphical models hold promise</td>
<td>• Graphical models are predominantly used to represent event scenarios
• Low-level descriptors are for detection and tracking</td>
</tr>
<tr>
<td>• Inadequate understanding of the effectiveness of various sensor data for monitoring nuclear proliferation
• Lack of efforts generating multi-sensor, multi-resolution fused images
• Inadequate and incomplete training and test image database</td>
<td>• Lack of consistency in the existing ontologies
• No unification and compatibility among geospatial ontologies</td>
<td>• Need for medium-level features
• Need for shape features</td>
<td>• Missing geaospatial semantics, spatial and spatiotemporal relationships in complex object recognition
• Lacks top-down and ontology driven classification framework
• Definition of (geospatial) saliency, and distinction between saliency and outliers and anomalies requires maturity</td>
<td>• Little effort for abstract event detection over wide area persistence surveillance sensors</td>
</tr>
</table>

Table 1. Table Caption

the fact that the given image contains a nuclear power plant. What is missing is the fact that the objects (e.g., switch yard, containment building, turbine generator, cooling towers) and their spatial relationships (arrangements or configurations) are not captured in traditional thematic classification. In addition, traditional image analysis approaches mainly exploit low-level image features (such as, color and texture and, to some extent, size and shape) and are oblivious to higher level descriptors and important spatial (topological) relationships without which we can not accurately discover these complex objects or higher level semantic concepts. One stumbling block in exploiting such relationships is in the description of compound objects and the spatial relationships among the object constituents. Therefore, for effective utilization of remote sensing imagery, first it is important to identify key concepts that describe a nuclear facility and describe (or encode) them using formal ontologies. Formal geospatial image ontology model then should include descriptions for image features and spatial relationships. Proliferation activity can then be monitored through a top-down, ontology guided semantic classification and search framework. We now describe several key geospatial image mining techniques that can be leveraged into this framework. Results of our current survey work are summarized in Table 1.



(a) FCC Image with Thematic class labels (B-Buildings, C-Crop, F-Forest)

(b) Thematic Classified Image (B-Buildings, C-Crop, F-Forest)

(c) FCC Image with Semantic Labels (S-Switch Yard, C-Containment Building, T-Turbine Generator, CT-Cooling Tower

Fig. 1. Thematic Classes vs. Semantic Classes

Geospatial Ontologies As discussed above, one of key requirement in proliferation monitoring is first to encode the key signatures or concepts that describe a nuclear facility or a proliferation scenario using formal ontologies. An image is essentially composed of a set of objects where the object view can be modeled more appropriately by geospatial image semantics. The conjunction of objects and their spatial relationships gives raise to semantic objects or higher-level concepts such as facility, airport, etc. Such high-level concepts play an important role in the way a human analyst perceives geospatial images (visual interpretation) or by image mining systems to extract such patterns (e.g., classification) or measure similarity among images containing such concepts (e.g., image retrieval).

Geospatial Image Features Feature extraction from spatial or hyperspectral imagery is a low-level operation, but it is crucial in the detection and characterization of complex geospatial objects for the assignment of semantic labels in nuclear proliferation scenarios. Feature extraction from spatial imagery and the subsequent translation of pixel-based features into subjects and predicates for query languages have been studied for many years. Objects are segmented from image backgrounds based on pixel operations that identify sharp discontinuities between pixel values within images. These gradient-based techniques attempt to identify regions or edge boundaries that represent distinct objects or parts of objects within a scene. In this manner, imagery can be transformed into a collection of local features that are invariant to affine transformations. Feature detection is the identification of regions of interest within the imagery and feature extraction is the analysis of a local patch associated with the feature in order to characterize that feature's properties. In nuclear proliferation detection, features may represent new building structures, changes in land use or temperature gradients within building structures or bodies of water. In recent study [2], several statistical properties of high-resolution satellite images were studied. Such statistics are highly useful in identifying and distinguishing structures within a image, and can be exploited in semantic classification of images.

Object-based Classification Increasing spatial resolution render per-pixel based classification schemes obsolete. To fully exploit the richness of high-resolutions geospatial image data, it is necessary to extend the per-pixel classification schemes to incorporate object relationships or hierarchies in learning process. Recent literature shows the importance and growth of the object-based classification approaches. Typical object-based classification involves two steps: in the first step, low-level image primitives such as image segmentation and clustering were applied to partition the image into a set of jointly exhaustive and mutually disjoint regions, and in the second step, the segments or clusters were subjected to a object fusion procedure. Fusion is typically done though a set of rules that were aimed at unifying the adjacent segments into an object. Recent advances in object classification can be found in [3]. Though object-based classification is an important step towards identifying objects, further research is needed to fully exploit spatial and semantic relationships to accurately identify complex spatial objects (or compound objects).

Segment-based Classification Since the detection of objects is quite difficult, an intermediate approach is to rely on the classification of image segments or fixed size blocks. Fixed size blocks are easy to obtain but may lead to misclassification, as they may not follow object boundaries. On the other hand, image segmentation techniques attempt to decompose an image into perceptually/semantically uniform regions, which can lead to better classification. A third alternative that has been proposed recently relies on Lempel-Ziv incremental parsing to decompose the image into variable size rectangular patches, which provide an asymptotically optimal representation for image compression and retrieval applications [4]. This parsed representation, in combination with latent semantic analysis (LSA), has been shown to provide semantic classification that is superior to that provided by fixed-block techniques.

Saliency Detection Several models of visual saliency have been developed to model the human vision system in detecting regions of interest within scenes [5]. Many techniques are based on the bottom-up model of visual saliency proposed in [6] which identifies locations within a scene that are sufficiently different from the remainder of the scene to warrant increased attention. This approach attempts to develop saliency maps based on multiscale features that have been undergone filtering. Recent work utilizing the Bayesian model of surprise may be of interest here [6]. Recent studies have shown that bottom-map saliency approaches do not perform well at change detection unless some semantic information is included.

Pattern Mining Patterns can be defined as recurrent and predictable occurrences of one or more objects that satisfy user defined constraints. Though pattern recognition is a well-established discipline, when it comes to massive datasets with complex objects and numerous interrelationships between the objects, it would be very hard to define a good pattern. Therefore one should able to search for patterns in the data even in situations where definition of pattern is not known. This automated and systematic search for patterns in large volumes of data is called pattern mining. Recent approaches in spatial data mining, like colocation mining, are a first step towards automatic discovery of patterns in large volumes of spatial and spatiotemporal data.

Colocation Mining Boolean spatial features are geographic object types that are either present or absent at different location in a two dimensional or three dimensional spaces, e.g., the surface of the earth. Examples of Boolean spatial features include plant and animal species, cancers, crime; neighborhoods (house and swimming pool; building and parking lot; etc.). Colocation patterns represent the subset of the Boolean spatial features whose instances are often located in close geographic proximity. Though there are several efficient algorithms for colocation mining exist, further research is needed to incorporate spatial relationships other than geographic proximity.

3. CONCLUSIONS

In this paper we described an important problem of developing geospatial image mining framework for monitoring nuclear nonproliferation using remote sensing imagery. We described key components of this framework. This is an ongoing project and we present recent results and discuss important open research problems in the final paper.

4. ACKNOWLEDGMENTS

Research is supported by the Simulation, Algorithms and Modeling (SAM) program office, NA-22, National Nuclear Safety Administration, U. S. Department of Energy. Authors of this manuscript are employed by the national laboratories management contractors for the U. S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

5. REFERENCES

[1] Bhupendra Jasani, Sven Nussbaum, and Irmgard Niemeyer, *International Safeguards and Satellite Imagery: Key Features of the Nuclear Fuel Cycle and Computer-Based Analysis*, Springer, Berlin, 2009.

[2] V. Vijayaraj, A.M. Cheriyadat, Phil Sallee, Brian Colder, R.R. Vatsavai, E.A. Bright, and B.L. Bhaduri, "Overhead image statistics," *Applied Image Pattern Recognition Workshop,*, vol. 0, pp. 1–8, 2008.

[3] T. Blaschke, S. Lang, and G.J. Hay, *Object-Based Image Analysis spatial concepts for knowledge-driven remote sensing applications*, Springer, Berlin, 2008.

[4] Soo Hyun Bae and Biing-Hwang Juang, "Incremental parsing for latent semantic indexing of images," in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, Oct. 2008, pp. 925–928.

[5] Timor Kadir and Michael Brady, "Saliency, scale and image description," *Int. J. Comput. Vision*, vol. 45, no. 2, pp. 83–105, 2001.

[6] Laurent Itti, Christof Koch, and Ernst Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, 1998.