A Framework for Ontology-Supported Intelligent Geospatial Feature Discovery Services

Liping Di, Peng Yue, Peisheng Zhao, Wenli Yang, Weiguo Han
Center for Spatial Information Science and Systems
George Mason University
Fairfax, Virginia 22030
(ldi, pyue, pzhao, wyang1, whan)@gmu.edu

***Abstract*—Geospatial feature discovery from remote sensing imageries is widely used in national defense and security communities. Existing methods in the geospatial image mining and feature extraction focus on the manual or automated processing of images to detect individual elementary features, such as building and highway. Such elementary features don't tell much semantic information about the features. Compound geospatial features such as Weapons of Mass Destruction (WMD) proliferation facilities are spatially composed of elementary features (e.g., buildings for hosting fuel concentration machines, cooling ponds, and transportation railways). The identity and much semantic information of a compound geospatial feature can be derived from the spatial relationship among the elementary elements. In this paper, we propose a flexible service framework for discovering compound geospatial features using an ontology-supported approach. The ontology for facilities helps find compound features that contain the specified spatial relationships among constituent features. The framework uses Web services for elementary feature extraction or access of existing elementary features, identifies facilities based on semantic descriptions of elementary feature constituents and their spatial relationships, and composes workflow-based service chains for automatic feature discovery.**

Keywords-image mining, semantic web, ontology, geospatial services, workflow, feature discovery

I. INTRODUCTION

Rapid increasing in the remote sensing capability in recent years, especially in the high spatial resolution imaging, has allowed identification of geospatial features and their changes over time. Consequently, analysis of geospatial imagery has become a promising approach for characterizing Weapons of Mass Destruction (WMD) (including nuclear) proliferation. However, the overwhelming volume of routine image acquisition has greatly outpaced the increase in the capacity of manual interpretation by intelligent analysts and has prompted automated approaches for image processing for information or knowledge generation that can be used in the decision making. An automated system, which can automatically identify geospatial features such as suspicious WMD proliferation sites in images for intelligence analysts to further investigate, can significantly reduce the workload of human intelligence analysts and increase the possibility of prompt detection of WMD proliferation.

Current methods in the geospatial image mining and feature extraction focus on the manual or automated processing of the incoming images to detect elementary visual features, such as building, highway [1-3]. Classification is often performed on per-pixel basis, although region-based characterization has received increasing attention in the recent years [4, 5]. On the other hand, geospatial images contain complex (compound) features and patterns. These features and patterns contain spatial relationships (metric, topological, etc). Traditional image analysis approaches mainly exploit image features, such as, color and texture, and, to some extent, size and shape. These image features ignore important spatial (topological) relationships [6], without which we cannot accurately discover complex (compound) features that relate to facilities used for manufactory, storage, and transportation of WMD.

In this paper, we present the concept and implementation architecture of an ontology-supported approach for discovering complex features in a service oriented environment. Web service technologies can be used to extract, access, and discover features in a distributed environment. Ontologies for complex features include semantic descriptions of elementary features and their spatial relationships. Using ontological reasoning and planning, we can decompose a detection task into a series of spatial/temporal relationship queries. The automatic execution of the series of queries to detect complex features such as possible WMD proliferation site can be performed through the workflow execution under a Service-Oriented Architecture (SOA)-based system. In the rest of the paper, we use the semiconductor manufacturing facility as an example of complex features.

The remainder of the paper is organized as follows. Section II describes the discovery of elementary geospatial features. These features can be either extracted from imagery using feature extraction services or accessed from existing feature repositories. In Section III, we discuss the ontologies for complex features, whose constituent features and spatial relations among them support the decomposition of detection tasks into workflows in Section IV. Section V describes the implementation architecture in SOA. Discussion and conclusion are given in Section VI.

II. DISCOVERY OF ELEMENTARY GEOSPATIAL FEATURES

A manufacturing facility consists of a group of elementary ground features (e.g., buildings for hosting fuel concentration machines, cooling ponds, transportation railways, fence, etc). Discovery of elementary features can be conducted by either 1) performing new feature extraction from high resolution remote sensing images or 2) accessing existing feature repositories. In both cases, the Catalogue Service for the Web (CSW) from the Open Geospatial Consortium (OGC) can be used by front-end users to find the features of interest.

There are already a significant number of algorithms for extracting elementary features (e.g., building, railways, highway, and ponds) from high spatial resolution images and other sources [1-3]. It is promising to adopt new technologies and flexible systems to make these existing algorithms capable of plug-in-and-play for on-demand feature extraction from high-resolution images.

Web services provide a promising prospect to have feature extraction done automatically over the Web. A geospatial Web service is a modular Web application that operates on geospatial data, information, or knowledge. Geospatial Web services can perform any function from simple geospatial data request to complex geospatial analysis [7]. Individual geospatial Web services can be chained together as workflows to accomplish complex tasks. Figure 1 shows an example of service components connected in the pipeline as a workflow for providing new geospatial features.

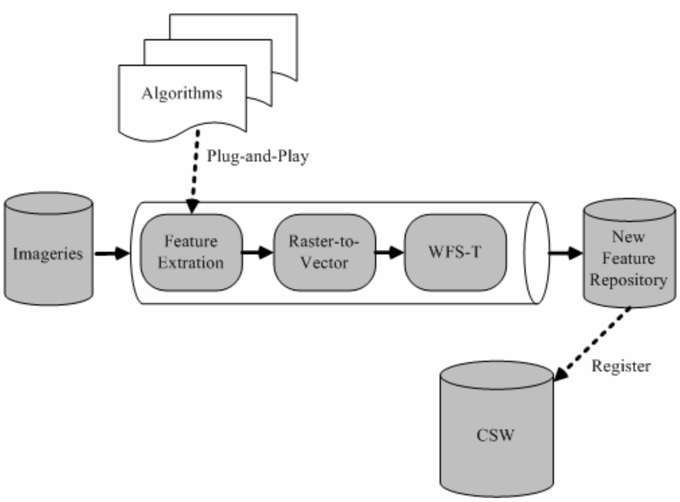

Figure 1. Feature Extraction and Registration

Most of remote sensors collect data as images in raster form. The feature extraction algorithms can recognize objects on images. Each algorithm can be provided as a feature extraction service and plugged into the workflow. The geospatial data at the feature level is in a vector form. The extraction results need to be converted into vectors using a raster-to-vector conversion service. The vector is inserted in a transactional Web Feature Service (WFS-T) and stored in a new feature repository. WFS-T is an OGC standard specification. The new features can be registered in the CSW. The feature extraction services can also follow the OGC Web Processing Service (WPS) specification. These OGC service specifications define the standard interfaces and protocols for geospatial services to ensure the interoperability of these services.

III. ONTOLOGIES FOR COMPOUND FEATURES

The development of ontologies for compound features is to encode the knowledge of the subject matter experts in a form that can be used for automated inference and retrieval of the knowledge items. In turn, automated inference and knowledge item retrieval will be used to facilitate matching between the subject area concepts and classes of the features extracted from the imagery.

Two types of ontologies are used in describing the semantics of compound features:

1) Spatial and spatiotemporal ontologies for compound geospatial features associated with manufacturing facilities. For example, spatial predicate such as surrounded by, near, and cross can be created in the ontologies.

2) Ontologies for describing semantic concepts of and assigning semantic labels to compound geospatial features and their constituents for manufacturing domain scenarios based on current and future functional and operational requirements. The elementary features, such as building, fences, bridges, railway, railway stations, and airports, can be defined here.

By combining the two ontologies together, the application ontology, such as the manufacturing facility ontology, can describe the concepts of manufacturing facility, related compound ground features and their elementary features, and spatial/temporal relationships among the member features and with the surrounding environment (surrounding ground features).

To achieve maximum flexibility of the ontology design, a layered approach can be utilized: for most general concepts we can use existing upper-level ontologies such as DOLCE or BFO [8, 9]. The next level, describing concepts specific to geographic domain and spatial relations, can be based on the existing ontologies, the SWEET ontology [10] and Ordnance Survey Ontologies [11]. SWEET Ontology can be used in parts relating to geosciences and remote sensing, and Ordnance Survey Ontologies can be used in parts relating to buildings and facilities. Ontologies pertaining to manufacturing facility can be developed by extending existing ontologies.

IV. TASK DECOMPOSITION FOR COMPLEX FEATURE DISCOVERY

When elementary features are available from either image extractions or existing feature repositories, the task of detecting a manufacturing facility in an intelligent system can be decomposed into a series of queries of geospatial features based on the spatial relationship among the features. Examples of the queries are:

- Find groups of buildings surrounded by fences
- Find a school near metro station in a specific city
- Find roads cross parks

The above queries are constructed on spatial relationships between two groups of geographic features (binary query). A more complex query may involve multiple geographic features. For example: *Find a group of buildings surrounded by fence and near a railway station*.

The binary query can be generalized as:

Find [<quantifier1>] <feature1> <operator> [<quantifier2>] <feature2> [in <spatial area>] [at <time>]

where, feature1 and feature2 are the geographic features, such as building, schools, railways, highways, airports, bridges, fences, factories, etc, which can be topologically expressed as point, line, or polygon. Only features available in the CSW are allowed here. The quantifier1 and quantifier2 put the quantification on the features, e.g., single, a group of. The operator defines the relationship between the two geospatial features. Example of operators includes "surrounded by", "near", "containing", "cross", etc.

Complex queries can be accommodated by a chain of the binary queries with the output of the previous query is the input of the next query. For example, the complex query above can be viewed as *Find (Find groups of building surrounded by fence) near a railway station.*

The decomposition process, which decomposes a detection task into a series of spatial relationship queries, can be automated through ontological reasoning and planning. A simple scenario of manufacturing facility detection is as following:

1) Users submit a request to detect an instance of facility whose concept has been described in an ontology;

2) The request is converted to a complex spatial relationship query based on the constituent features and their spatial relationship defined in the ontology for manufacturing facilities;

3) The complex query is transformed into a chain of the binary queries;

4) The chain of the binary queries is used to construct a workflow;

5) The workflow is transformed into a service chain, consisting of services and data in the CSW.

6) The workflow engine executes the service chain to generate the answer, which is displayed in the graphical user interface overlaid with the source images to show locations of matched geographic features.

Thus, the execution of series of queries to detect the possible compound features can be performed through the workflow execution under a service system. Spatial operators in the binary queries are implemented as geospatial web services. This dynamic and automatic reasoning process takes into consideration the feature semantics and service semantics when generating workflows. The workflows can be combined with feature extraction workflows for on-demand detection of facilities from high-resolution images.

V. IMPLEMENTATION ARCHITECTURE

The ontology-supported intelligent geospatial feature discovery system is a SOA-based system. The system is designed to be application neutral (i.e., it can be used to solve the similar issues in different applications, for example, either WMD proliferation sites or other facilities detection, by using different domain ontologies). Figure 2 shows the architecture of the system.

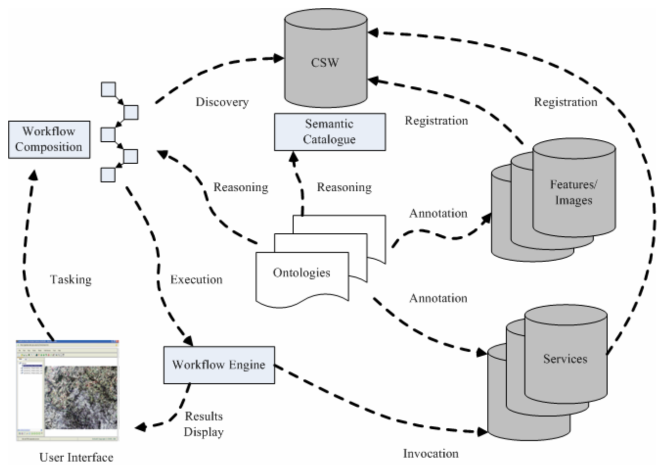

Figure 2. Implementation Architecture

A. Ontoloiges

The system can be used in different applications as long as ontologies for such applications are available. The ontologies are described using Web Ontology Language (OWL). The inference engine is used to find all relevant entailments.

B. CSW

CSW is an OGC standard specification for metadata catalogue services. Metadata for data and services are registered in the CSW for search. The semantic annotations are registered in an OGC compliant semantic catalogue to allow the semantics-enhanced discovery of available data and services [12].

C. Workflow

The facility detection task can be decomposed into a series of binary queries (spatial operations) on elementary features using the ontological reasoning and planning in the workflow composition service. The series of operations are encoded as a service chain, which will be executed by a workflow engine.

D. *Data and Services*

The data includes the features that have been already extracted from images. When only raw images are available, feature extraction services or workflows will be used. Feature extraction, spatial relations, and other utility services such as coordinate transformation and file format conversion can be provided. Some of them are already available in the GeoBrain project [13]. The semantics for data and services are annotated using entities from ontologies.

E. *User Interface*

The user interface includes both the graphical user interface (GUI) and API interface. The GeoBrain project has developed a powerful portal called GeOnAS for user interaction [14]. It can be modified in the prototype implementation here.

VI. DISCUSSION AND CONCLUSION

Currently extraction of semantic information and semantic labeling of the features in high-resolution remotely sensed imagery is a very actively developing area of remote sensing [15-17, 6]. Typically such algorithms use training data in the form of image segments with known objects and then use various statistics to match training data with the imagery. Even though effective for many purposes, such one-step approaches are likely to fail when there are subtle differences between the complex features on an image. For example, presence of an industrial chimney is a salient feature distinguishing nuclear power plant from a coal-firing plant. However, chimney is a relatively small feature in the planar view that is unlikely to produce distinguishable effect in matching statistics. The use of formally encoded semantic information in the form of ontologies can solve these problems by explicitly identifying salient features of the compound objects and their constituents.

The approach described in this paper is a two-step approach, with step 1 to identify the location and type of elementary ground features (such as building, road, etc) from high-resolution images, which has mature technology already, and step 2 to extract high-level semantic information (such as nuclear fuel concentration facilities, nuclear test sites, missile test sites, etc) through discovering compound ground features based on spatial (topological) relationships among the elementary features. The concepts and spatial relationships are described in ontologies. Therefore, ontologies are essential for system to work. However, the system is ontology independent. It can work in different domain with different domain ontologies.

The development of the system follows the SOA paradigm and using Web services and Semantic Web technologies. It allows discovery and composition of features and services in a distributed environment. The automatic discovery of compound features using workflow composition and execution can reduce the workload of human intelligent analysts and provide valuable information in decision making. The next step will be the proof-of-concept implementation, and evaluation the ontology approach.

ACKNOWLEDGMENT

This research is supported by a grant from U.S. Department of Energy (grant #DE-NA0001123, PI: Prof. Liping Di)

REFERENCES

[1] J. B. Mena and J. A. Malpica, "An automatic method for road extraction in rural and semi-urban areas starting from high resolution satellite imagery," Pattern Recognition Letters, vol 26, issue 9, pp. 1201-1220, July 2005.

[2] J. Ahmad and CK Murali, "Automatic Feature Extraction: A Solution for Extracting the Features from High Resolution Satellite Images," in Map Asia 2006. http://www.gisdevelopment.net/technology/rs/ma06_147abs.htm.

[3] G. Sohn and I. Dowmana, "Data fusion of high-resolution satellite imagery and LiDAR data for automatic building extraction," ISPRS Journal of Photogrammetry and Remote Sensing, vol 62, issue 1, pp. 43-63, May 2007.

[4] A. Frome, Y. Singer, F. Sha, and J. Malik, "Learning globally-consistent local distance functions for shape-based image retrieval and classification," in IEEE 11th International Conference on Computer Vision (ICCV 2007), pp. 1-8.

[5] X. Wang, B. Waske, and J. A. Benediktsson, "Ensemble methods for spectral-spatial classification of urban hyperspectral data," in Geoscience and Remote Sensing Symposium (IGARSS), 2009 IEEE International, 2009, pp. 944-947.

[6] R. R. Vatsavai, B. Bhaduri, A. Cheriyadat, L. Arrowood, E. Bright, S. Gleason, C. Diegert, A. Katsaggelos, T. Pappas, R. Porter, J. Bollinger, B. Chen, and R. Hohimer, "Geospatial image mining for nuclear proliferation detection: Challenges and new opportunities," in Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International, 2010, pp. 48−51.

[7] L. Di, P. Zhao, W. Yang, G. Yu, and P. Yue, "Intelligent geospatial web services," in Geoscience and Remote Sensing Symposium (IGARSS), 2005 IEEE International, 2005, pp. 1229 - 1232.

[8] DOLCE, "DOLCE: A Descriptive Ontology for Linguistic and Cognitive Engineering," Laboratory for Applied Ontology - DOLCE. Retrieved November 6, 2010, from http://www.loa-cnr.it/DOLCE.html

[9] B. Smith, "Basic Formal Ontology," Retrieved November 5, 2010, from http://www.ifomis.org/bfo

[10] R. Raskin, "SWEET Ontology. Semantic Web for Earth and Environmental Terminology (SWEET)," Retrieved November 6, 2010, from http://sweet.jpl.nasa.gov/

[11] Ordnance Survey, "Ordnance Survey Ontologies," Retrieved November 6, 2009, from http://www.ordnancesurvey.co.uk/oswebsite/ontology/

[12] P. Yue, J. Gong, L. Di, L. He, and Y. Wei, "Integrating semantic web technologies and geospatial catalog services for geospatial information discovery and processing in cyberinfrastructure," GeoInformatica, vol 15, issue 2, 2011, pp. 273–303.

[13] L. Di, "GeoBrain-a web services based geospatial knowledge building system," in Proceedings of NASA Earth Science Technology Conference 2004. June 22-24, 2004. Palo Alto, CA, USA. (8 pages, CD-ROM).

[14] W. Han, L. Di, P. Zhao, Y. Wei, an dX. Li, "Design and implementation of GeoBrain online analysis system (GeOnAS)," In Proceedings of 8th International Symposium on Web and Wireless Geographical Information System, Michela Bertolotto, Cyril Ray, Xiang Li (Eds):Lecture Notes in Computer Science (LNCS), 5373, 2008, pp. 27-36.

[15] K. W. Tobin, B. L. Bhaduri, E. A. Bright, A. Cheriyadat, T. P. Karnowski, P. J. Palathingal, T. E. Potok, et al., "Automated feature generation in large-scale geospatial libraries for content-based indexing," Photogrammetric engineering and remote sensing, vol 72, issue 5, May 2006, pp. 531-540.

[16] S. Gleason, R. Ferrell, A. Cheriyadat, R. R. Vatsavai, and S. De, “Semantic information extraction from multispectral geospatial imagery via a flexible framework,” in Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International, 2010, pp. 166-169.

[17] R. R. Vatsavai, A. Cheriyadat, and S. Gleason, “Unsupervised semantic labeling framework for identification of complex facilities in high-resolution remote sensing images,” in 2010 IEEE International Conference on Data Mining Workshops (ICDMW), Sydney, Australia, pp. 273 – 280.