LLNL-CONF-436076

LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

On the Verification and Validation of Geospatial Image Analysis Algorithms

R. S. Roberts, T. G. Trucano, P. A. Pope, C. R. Aragon, M. Jiang, T. Wei, L. K. Chilton, A. Bakel

June 16, 2010

International Symposium Geoscience and Remote Sensing
Honolulu, HI, United States
July 25, 2010 through July 30, 2010

ON THE VERIFICATION AND VALIDATION OF GEOSPATIAL IMAGE ANALYSIS ALGORITHMS

Randy S. Roberts[1], *Timothy G. Trucano*[2], *Paul A. Pope*[3], *Cecilia R. Aragon*[4],
Ming Jiang[1], *Thomas Wei*[5], *Lawrence K. Chilton*[6] *and Alan Bakel*[5]

{[1]Lawrence Livermore, [2]Sandia, [3]Los Alamos, [4]Lawrence Berkeley,
[5]Argonne, [6]Pacific Northwest} National Laboratory, USA
Email: roberts38@llnl.gov

ABSTRACT

Verification and validation (V&V) of geospatial image analysis algorithms is a difficult task and is becoming increasingly important. While there are many types of image analysis algorithms, we focus on developing V&V methodologies for algorithms designed to provide textual descriptions of geospatial imagery. In this paper, we present a novel methodological basis for V&V that employs a domain-specific ontology, which provides a naming convention for a domain-bounded set of objects and a set of named relationships between these objects. We describe a validation process that proceeds through objectively comparing benchmark imagery, produced using the ontology, with algorithm results. As an example, we describe how the proposed V&V methodology would be applied to algorithms designed to provide textual descriptions of facilities.

Index Terms— Verification and Validation, ontology, geospatial image analysis, image benchmarks, facility detection, text-to-image, image-to-text

1. INTRODUCTION

Verification and validation (V&V) of geospatial image analysis algorithms is a difficult task and is becoming increasingly important. The amounts and types of imagery produced by existing geospatial sensors readily overwhelm the abilities of human analysts, and future sensing capabilities will add to the torrent of data. Moreover, geospatial image analysis is increasingly called upon to answer very complex questions. For example, consider problems such as detecting nuclear proliferation activities, or performing time-dependent environmental characterizations. Analysis of complex spatio-temporal problems such as these typically requires large quantities of multi-modal imagery collected over long periods of time. As the sophistication, automation, and scope of geospatial image analysis increases, so does the need to verify and validate the performance of the underlying algorithms.

While many types of image analysis algorithms (along with a comprehensive V&V methodology) are needed to unravel the complex scenarios mentioned above, we focus on developing V&V methodologies for algorithms designed to provide textual descriptions of geospatial imagery. Textual descriptions of imagery are the backbone of high-level image exploitation tasks such as image indexing and retrieval, data mining and image understanding. Moreover, textual descriptions of imagery may become a key provision to realizing the potential of the Semantic Web. Creating a textual description of the context and meaning in imagery is in itself a very difficult problem [1]. The description of the imagery should not only include the salient objects and their attributes, but also the geospatial, temporal and functional relationships between the objects.

In this paper, we summarize a novel methodological basis for V&V of algorithms designed to process complex geospatial imagery. Note that in this paper we will generally to refer to "V&V," but the reader should understand that our emphasis is on validation. We begin by surveying the state-of-the-art in methodologies for algorithm V&V and argue that these approaches are not well suited for V&V of algorithms that process geospatial imagery. We then describe an approach employing a domain-specific ontology to enable the proposed V&V methodology. The ontology, as an interpretive conceptual basis for geospatial analysis, provides descriptions of objects and relationships between the objects. Using the ontology, benchmark imagery is produced for three purposes: algorithm verification, calibration and validation. We describe a process by which validation proceeds through objectively comparing benchmark imagery with algorithm outputs. We conclude the paper by describing how the proposed V&V methodology would be applied to algorithms designed to provide textual descriptions of facilities, and point out gaps in technology that need to be addressed before the V&V methodolgy can be fully implemented.

2. CURRENT APPROACHES TO V&V OF GEOSPATIAL IMAGE ANALYSIS ALGORITHMS

There currently exist a number of V&V principles, conceptual frameworks, and guidance. Verification is defined as the process of evaluating an algorithm to determine if it has been correctly implemented in software. Validation is defined as the process of evaluating an algorithm to determine if it satisfies specific requirements, or, more generally, to determine if it is the "correct" algorithm for the intended applications. The broadest scope V&V frameworks are probably those of the IEEE [2], which is heavily software centric, and the Department of Defense Modeling and Simulation Coordination Office (MSCO) [3], which has a huge modeling and simulation scope including individual, organizational and social models, war games, and so on. AIAA and ASME have developed guidance and frameworks specific to the needs and requirements of computational physics and engineering [4, 5]. The formal V&V program associated with the DOE NNSA Advanced Simulation and Computing (ASC) program directly targets large-scale computational physics and engineering [6, 7]. There have also been publications related to the V&V of image processing algorithms [8].

Also of interest are on-line algorithm competitions and de facto standard sets of test imagery, although these are not necessarily as-

To appear in *Proc. IGARSS 2010, July 25-30, Honolulu, HI, USA*

sembled specifically for the purpose of assessing geospatial algorithms (see, for example, the Caltech 101 image benchmark suite [9], and the Overhead Imagery Research Data Set (OIRDS) [10]). These compendiums and competitions have not been incorporated in a formal V&V context, and whether the test problems involved are appropriate for use in a rigorous V&V methodology is therefore an open question. There is published debate about the appropriateness of, for example, the Caltech 101 benchmarks [11, 12], which suggests that the question of overall suitability of these kinds of test collections for V&V may be a good one.

While the aforementioned V&V frameworks have important commonalities, which we address in the paper, we emphasize that they do not encompass important geospatial algorithm V&V issues. The very broad V&V frameworks of IEEE and MSCO, having their centers-of-gravity on software implementation assessment, do not address the range of complexities that arise when validation benchmarks are defined by physical observational data. V&V guidance from ASME, AIAA and ASC commonly acknowledges specific difficulties associated with observational data, and significant complexities are introduced in their frameworks to respond to the intricacies of observation-based validation procedures. However, ultimately observation-based validation is highly subject matter (domain) specific. Therefore, while we generally apply principles identified by ASME, AIAA and ASC for our consideration of validation, the devil remains in the details of the differences between geospatial analysis and computational fluid or solid mechanics. Dealing with these differences is the essential novelty of our endeavor.

3. PROPOSED APPROACH TO V&V OF GEOSPATIAL IMAGE ANALYSIS ALGORITHMS

A comprehensive methodology for V&V of geospatial image algorithms should contain the following attributes: 1) Quantitative measures of usability and end-user needs; 2) Precisely defined types and quantities of geospatial benchmarks required for V&V; 3) Specification of the kinds and degree of geospatial benchmark variability; 4) Quantitative methods for comparing and summarizing geospatial benchmarks with algorithm outputs; 5) Methods for determining if the accuracy achieved meets the application requirements for the algorithm; and 6) Procedures for quantifying and tracking uncertainties throughout the entire V&V process. Note that we use the word "benchmark" to refer to "reference data." The former term is common in V&V-related literature, while the latter term is used in geographic information science.

3.1. Ontology for Geospatial V&V

Our approach to the V&V of geospatial image analysis algorithms begins with an application-specific ontology. The role of an ontology as an enabler in the V&V of geospatial image analysis algorithms is novel. An ontology provides an agreed upon conceptualization of reality for a particular knowledge domain. It is used as a guide to define objects, and their spatial and temporal interrelationships, which comprise the scenes and scenarios captured by the benchmark imagery. The ontology has a fundamental role in defining observational validation benchmarks. In fact, the validity of the underlying ontology becomes an additional factor in the overall V&V assessment, which further complicates the goals, conduct, and outcomes of geospatial algorithm validation. We observe that "validity" of the underlying ontology, or "evaluation" as it is commonly called in the ontology literature [13], is an epistemic uncertainty for the overall V&V process.

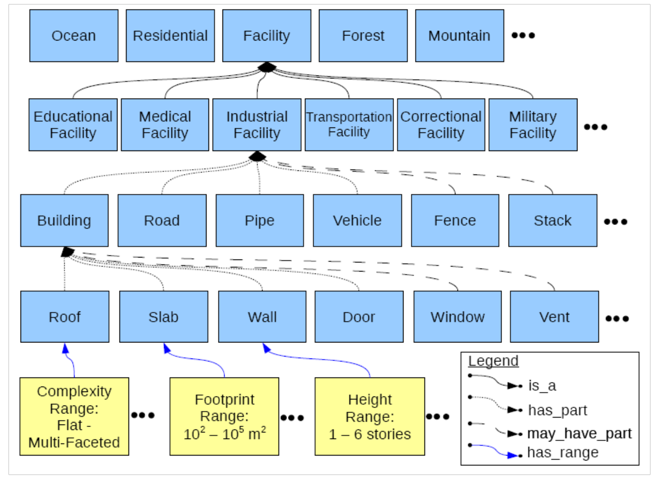

Fig. 1. Schematic of a facility ontology for V&V of geospatial algorithms

An ontology is a "specification of a conceptualization" for a particular knowledge domain [14]. Ontologies are typically constructed through knowledge elicitation of Subject Matter Experts. Note that choice of the knowledge domain represented by the ontology is implicitly set by the end-user needs. It can be interpreted as a taxonomy of concepts, with terms shown in the boxes and the relations between them portrayed as connecting lines. A schematic diagram of a facility ontology is depicted in Figure 1. More abstract terms are at the top, with more specific terms toward the bottom, and several relations exist. Only a portion of the larger ontology is shown; the ellipses imply the fact that numerous other terms exist at each level. Although an ontology is often depicted in the form of simple tree graphs, like we have presented here, they are typically too large and complex for a human to grasp directly; these constructs are more easily leveraged via computer. Also shown are one characteristic each for the concepts "roof", "slab", and "wall", all of which are parts of the concept "building." Note that the concepts included in an ontology generally strive for words that are generic and do not imply a particular use (e.g. "slab" versus "foundation") such that they are universal among numerous domains. It is through this type of design that an ontology serves its greatest purpose, that is, connecting heterogeneous databases of information about disparate domains of knowledge [15]. For example, this ontology could serve both as a guide to the generation of synthetic imagery, as well as act at the core of a Geospatial Object-based Image Analysis algorithm (GEOBIA) which searches an image to detect and identify these concepts [16].

3.2. Proposed Geospatial V&V Methodology

Using the aforementioned list of desirable attributes of a V&V methodology, we have defined a conceptual model, or process, for V&V of geospatial algorithms, with explicit emphasis on validation. The major elements of this process, in the sequential order of their application, are: 1) Specification of the application requirements for the algorithm slated for V&V; 2) Identify or create the minimal ontology that spans the application space of the algorithm; 3) Identify or create benchmark imagery using the ontology identified in step 2, thereby defining specific validation tests. Derive reference data from the benchmark imagery as required by the specifications;

4) Process the benchmark imagery with the algorithm undergoing V&V; 5) Quantitatively assess the performance of the algorithm on the reference data; and 6) Assess the adequacy of the algorithm's performance relative to the intended application. These process components are intensely domain specific. Their specific form and application for geospatial processing algorithms is another key difference in our work from existing guidance. We elucidate some aspects of the proposed methodology below through example and discussion.

Consider the validation of algorithms designed to provide textual descriptions of facilities. As noted by Yao, et al. [1], this is a very complex problem, and many different types of algorithms are required for a functional system. For specificity, we focus on the V&V of segmentation algorithms, a key component of an image-to-text system. Applying the first step of the V&V process, the validation team and the end-user agree that the algorithm must accept a geospatial image of a facility, and return an image that segments the following objects: buildings, roads, fences, pipes and effluent stacks. Ranges on the dimensions of these objects are provided. Furthermore, the imagery will be collected at noon on clear to partially-cloudy days in the summertime. The terrain and landcover surrounding the facilty are unspecified, and left to the validation team as free parameters.

Given these algorithm specifications, the validation team identifies (or constructs) an ontology like the one illustrated in Figure 1. This ontology indicates the components of an industrial facility, along with ranges for their dimensions. Note that the onotology should also indicate the geospatial relationships between these objects (e.g. "next to", "on top of", etc.), but this is a topic of current research in mereotopology [17]. Additionally, the ontology also provides geospatial models for other objects that provide scene clutter. The application specific ontology required by the methodology imposes rigor on the specification of the test data. It is used to drive selection or generation of benchmark imagery, which are used directly, or to derive, "ground truth." In this sense, the ontology is the foundation of the text-to-image process used to create benchmark imagery.

Creation of benchmark imagery can be accomplished in a number of ways, including observational (real-world) collection, image composition, and image synthesis as potential methods. There are two aspects of benchmark imagery: the physical model that underlies the imagery (composed of objects and topography), and realizations of the physical model modulated by environmental factors such as illumination and atmospheric conditions, sensor characteristics and collection geometries. The ontology guides the selection of models from which physical scenes are composed, and includes both objects of interest and objects that provide degrees of scene clutter. Extrinsic parameters such as environmental factors, sensor attributes and collection geometries are separate from the ontology, yet equally important. These parameters are inherently part of observational collections but can be manipulated in synthetic imagery.

In our example, the validation team would assemble a large number of benchmark images that contain the required objects and span the ranges of object dimensions, geospatial arrangements, and extrinsic parameters as specified by the application requirements. The ensemble of imagery would contain a range of segmentation problems, from images that a simple segmentation algorithm would correctly segment, to images that state-of-the-art algorithms would not be able to segment. Given the potential range of scene objects, clutter, and extrinsic parameters, the number of potential benchmark images is enormous. Methods to select or design imagery with appropriate scene content and variability for algorithm V&V is a gap in current research. Dealing with the complexity of this test data is a distinguishing feature of the V&V of geospatial algorithms compared to the V&V of other types of algorithms.

After an appropriate collection of benchmark imagery has been assembled, it is used either directly, or indirectly, as reference data to validate the algorithm. In our example, segments (represented by imagery, vectors, or other means) are derived from the benchmark imagery and this reference data is compared to the output of the algorithm in the validation process. If the benchmark imagery is observational, a combination of existing segmentation algorithms and human subject-matter-experts can be used to create the reference data. If the imagery is composite or synthetic, the reference data can be created as a by-product of the composition or synthesis process (e.g., auto-annotation).

The comparison of reference data derived from benchmarks with output from the algorithm in the validation process concentrates on one or more quantities, characteristics or features which can be rigorously compared. The design of metrics that assess the performance of the algorithm on benchmarks is at the heart of the technical validation challenges in this process and influences the design of effective validation tests. In our example, the reference data is compared to the segmentations produced by the algorithm undergoing validation. There are many potential metrics that can be used for such comparisons, and finding optimal measures is an active area of research [18]. The key is to produce quantifiable measures that can be used to assess the statistical performance of the algorithm on the ensemble of reference data.

The last step in the vaidation process is to evaluate the performance of the algorithm relative to the application requirements. This step essentially determines if the performance of the algorithm is "good enough" for operational use. All of the statistical evidence of the algorithm's performance, along with the accumulated uncertainties of the validation process, are used in the evaluation. If the algorithm achieves an acceptable level of accuracy, and the uncertainties associated with the validation process are sufficiently understood and accepted, the algorithm is considered valid within the uncertainty of the validation process. On the other hand, if the accuracy of algorithm performance is insufficient, or the uncertainities associated with the validation process are unacceptable, then either the algorithm, and/or elements of the validation process require modification. Finally, in order to be successful at validating geospatial image analysis algorithms, we will need to measure their effectiveness in an operational environment via usability metrics. The ultimate success of the software will be dependent upon how effectively humans are able to use the tools to achieve the goal of detecting nuclear proliferation from geospatial imagery.

4. SUMMARY AND TECHNICAL CHALLENGES

This paper describes a novel approach to the Verification and Validation (V&V) of geospatial imagery analysis algorithms. Although we are concerned with all aspects of V&V, our emphasis is algorithm validation. Several approaches to algorithm V&V are available in the literature, but these approaches do not address intricacies unique to geospatial imagery. Desirable attributes of a V&V methodology specifically tailored to geospatial analysis algorithms are presented, and a V&V process is developed from these attributes. Fundamental to the proposed V&V methodology is an ontology that is used to guide the composition of benchmark imagery. Some aspects of the methodology are described using image segmentation algorithms as an example.

A full implementation of the proposed methodology would be difficult given current gaps in theory and technology. Rather than present an exhaustive list of technical challenges, two outstanding issues are mentioned. As previously noted, while the ontology imposes structure on scene composition, the number of potential benchmark images is huge. Developing techniques to select a resonable number of images while spaning the space of scene content prescribed by the ontology, as well as the space spanned by the extrinsic parameters, is critically important. A second outstanding issue is the use of composite and synthetic imagery as surrogates for real-world imagery. Both composite and synthetic imagery offer the potential for cost effective, comprehensive, and rigourous algorithm validation. But, under what conditions can composite and synthetic imagery be credibly used for algorithm validation? Further research is needed to close these (and other) technology gaps, thereby enabling robust validation of geospatial image analysis algorithms.

5. ACKNOWLEDGEMENTS

The authors would like to acknowledge the support of Dr. Alexander Slepoy, Program Manager, Simulations, Algorithms, and Modeling; Office of Nonproliferation Research & Development, National Nuclear Security Administration.

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. LLNL-CONF-XXXXXX.

6. REFERENCES

[1] B. Yao et al., "I2T: Image parsing to text description," *IEEE Proceedings*, 2010, (to appear).

[2] Software Engineering Standards Committee of the IEEE Computer Society, *IEEE Guide for Software Verification and Validation Plans*, IEEE Std 1059-1993. Institute of Electrical and Electronics Engineers, 1994.

[3] Navy Modeling and Simulation Management Office, *Modeling and Simulation Verification, Validation, and Accreditation Implementation Handbook*, Volume 1 VV&A Framework. United States Department of the Navy, 30 Mar 2004.

[4] American Institute of Aeronautics and Astronautics Staff, *AIAA Guide for the Verification and Validation of Computational Fluid Dynamics Simulations*, American Institute of Aeronautics & Astronautics, 1998.

[5] American Society of Mechanical Engineers, *V&V 10 - 2006 Guide for Verification and Validation in Computational Solid Mechanics*, American Society of Mechanical Engineers, 2006.

[6] United States Department of Energy, "ASCI Software Quality Engineering: Goals, Principles and Guidelines," Tech. Rep. DOE/DP/ASC - SQE-2000-FDRFT-VERS2, United States Department of Energy, Feb 2001.

[7] R. Klein et al., "ASC predictive science academic alliance program verification and validation whitepaper," Tech. Rep. UCRL-TR-220711, Lawrence Livermore National Laboratory, 2006.

[8] D. Desovski et al., "Verification and validation of a fingerprint image registration software," *EURASIP Journal on Applied Signal Processing*, vol. 2006, Article ID 15940, 9 pages, 2006.

[9] L. Fei-Fei, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories," *Computer Vision and Image Understanding*, vol. 106, no. 1, pp. 59–70, 2004.

[10] F. Tanner et al., "Overhead imagery research data set - an annotated data library & tools to aid in the development of computer vision algorithms," in *Proceedings of the 2009 IEEE Applied Imagery Pattern Recognition Workshop (AIPRW)*, Oct 14-16 2009, pp. 1–8.

[11] N. Pinto, D. Cox, and J. DiCarlo, "Why is real-world visual object recognition hard?," *PLoS Computational Biology*, vol. 4, no. 1, pp. 0151–0156, Jan 2008.

[12] J. Ponce et al., "Dataset issues in object recognition," in *Toward Category-Level Object Recognition*, vol. 4170/2006 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg, 2006.

[13] L. Obrst et al., "The evaluation of ontologies: Toward improved semantic interoperability," in *Semantic Web: Revolutionizing Knowledge Discovery in the Life Sciences*, C. Baker and K. Cheng, Eds. Springer, New York, NY, 2007.

[14] T. Gruber, "Ontology," in *Encyclopedia of Database System*, Ling Liu and Tamer M. Özsu, Eds. Springer-Verlag, 2008.

[15] W. Ceusters, B. Smith, and J. Flanagan, "Ontology and medical terminology: Why description logics are not enough," in *Towards an Electronic Patient Record (TEPR 2003)*, May 10-14 2003, Unpagenated CD.

[16] G.J. Hay and T. Blaschke, "Special issue: Geographic object-based image analysis (GEOBIA)," *Photogrammetric Engineering and Remote Sensing*, vol. 76, no. 2, pp. 121–122, 2009.

[17] M. Kavouras and M. Kokla, *Theories of Geographic Concepts: Ontological Approaches to Semantic Integration*, CRC Press, Boca Raton, Florida, 2008.

[18] X. Jiang et al., "Distance measures for image segmentation evaluation," *EURASIP Journal on Applied Signal Processing*, vol. 2006, Article ID 35909, 10 pages, 2006.