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]

{Lawrence Livermore[1], Sandia[2], Los Alamos[3], Lawrence Berkeley[4], Argonne[5], and Pacific Northwest[6]} National Laboratory

Verification and validation (V&V) of geospatial image processing and 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.

In this paper, we present a rigorous methodological basis for V&V of algorithms designed to process complex geospatial imagery. We begin by surveying the state-of-the-art in methodologies for algorithm verification and validation 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 frame the proposed V&V methodology. The ontology, as an interpretive conceptual basis for geospatial analysis, provides definitions of objects, relationships between objects, and similar attributes. 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 results. We conclude the paper by applying the proposed V&V technique in two different, complex, spatio-temporal analysis problems, and discussing the specifics of how our novel method may be applied more generally to the V&V of other geospatial image algorithms.

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 IEEE [1], which is heavily software centric, and the Department of Defense Modeling and Simulation Coordination Office (MSCO) [2], 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 [3, 4]. The formal V&V program associated with the DOE NNSA Advanced Simulation and Computing (ASC) program directly targets large-scale computational physics and engineering [5]. There have also been publications related to the V&V of image processing algorithms [6]. Also of interest are on-line algorithm competitions and de facto standard sets of test imagery, although these are not necessarily assembled specifically for the purpose of assessing geospatial algorithms (see, for example, the Caltech 101 image benchmark suite [7], and the Overhead Imagery Research Data Set (OIRDS) [8]). Note that datasets such as these have been found to lack necessary qualities of V&V data [9].

While the aforementioned 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 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 all 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.

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) precisely defined types and quantities of geospatial benchmarks required for V&V; 2) specification of the kinds and degree of geospatial benchmark variability; 3) quantitative methods for comparing and summarizing geospatial benchmarks with algorithm outputs; 4) methods for determining if the accuracy achieved meets the application requirements for the algorithm; 5) quantitative measures of usability and end-user needs; 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 systems. Also note that in this paper we will generally to refer to “V&V,” but the reader should understand that our emphasis is validation.

Our approach to V&V of geospatial image processing and analysis algorithms begins with an application-specific ontology. An ontology provides an agreed upon conceptualization of reality for a particular 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. Additionally, it describes the variability of scene content (e.g., different types of objects, their characteristics and positions, atmospheric conditions, sensor attributes, etc.) required for comprehensive V&V. The ontology has a fundamental role in defining observational validation benchmarks as well as in assessment of algorithm-to-benchmark comparisons. 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. The role of ontology in the V&V of geospatial image analysis algorithms is novel. We will discuss this issue in greater detail in the paper, but we observe that "validity" of the underlying ontology is an epistemic uncertainty for the overall V&V process.

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 geospatial analysis algorithm that is the target of validation, with its precisely defined application requirements; (2) the reference imagery (benchmark) defining specific validation tests; (3) quantitative methods for determining the difference between algorithm and benchmarks; and (4) validation adequacy assessment of the algorithm-benchmark differences. There are feedbacks in this process that we will discuss in the paper. The aforementioned ontology is used to drive selection or generation of benchmark imagery. Benchmarks are assigned the role of "ground truth" in this process and must be very carefully selected. These process components are intensely domain specific. Their specific form and application for geospatial processing algorithms is the key difference in our work from existing guidance. Creation of benchmark imagery can be accomplished in a number of ways, including observational collection, image composition, and image synthesis as potential methods. The comparison of algorithms and benchmarks concentrates on one or more quantities, characteristics or features which can be rigorously compared, and are referred to as *System Response Quantities* (SRQs). We call the rigorous quantitative techniques for comparing algorithm and benchmark SRQs *Validation Metrics.* The design of these metrics is at the heart of the technical validation challenges in this process and influences the design of effective validation tests. Finally, adequacy assessment eventually leads to an exit from the V&V process, and on to operational application of the method or tool.

In the paper we illustrate the V&V process on algorithms designed to process imagery related to two specific scenarios. In the first example, we consider a simple nuclear proliferation scenario where a pattern of shifting of barrels in a facility indicates a proliferation activity. In the second example, we consider a scenario of illegal manufacturing of a particular type of semiconductor chip. These scenarios involve several types of algorithms, ranging from image registration, to feature extraction and pattern analysis, and finally to extraction of semantic content. The V&V process is described for individual algorithms as well as the entire processing chains. By comparing and contrasting the two scenarios, we demonstrate the generality of our V&V process. We conclude the paper with a discussion of how the V&V process can be applied to geospatial image processing and analysis algorithms, specifically imagery change and anomaly detection algorithms, and directions for further research.

BIBLIOGRAPHY

[1] IEEE, *IEEE Guide for Software Verification and Validation Plans*, IEEE Computer Society Software Engineering Standards Committee standard, IEEE Std 1059-1993, 1994.

[2] United States Department of the Navy, "Modeling and Simulation Verification, Validation, and Accreditation Implementation Handbook: Volume 1 – VV&A Framework," Navy Modeling and Simulation Management Office report, 2004. Also, see the Department of Defense Modeling and Simulation Coordination Office Web site http://www.msco.mil/ .

[3] AIAA, *Guide for the verification and validation of computational fluid dynamics simulations*, AIAA-G-077-1998, American Institute of Aeronautics and Astronautics, Reston, VA, 1998.

[4] ASME , *V&V 10 - 2006 Guide for Verification and Validation in Computational Solid Mechanics*, ASME, 2006.

[5] United States Department of Energy (2001), "ASCI Software Quality Engineering: Goals, Principles and Guidelines," DOE/DP/ASC – SQE-2000-FDRFT-VERS2.

[6] D. Desovski et al., "Verification and Validation of a Fingerprint Image Registration Software," *EURASIP Journal on Applied Signal Processing*, Volume 15940, pp. 1-9, 2006

[7] 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*, Volume 106, Number 1, pp. 59-70, 2004.

[8] F. Tanner, B. Colder, C. Pullen, D. Heagy, M. Eppolito, V. Carlan, C. Oertel, & P. Sallee, "Overhead Imagery Research Data Set – an annotated data library and tools to aid in the development of computer vision algorithms", *Proceedings of IEEE Applied Imagery Pattern Recognition Workshop 2009*.

[9] N. Pinto et al., "Why is Real-World Visual Object Recognition Hard?" *PLoS Computational Biology*, Volume 4, Number 1, pp. 0151-0156, 2008.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the support of 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-ABS-421217