SAND2016-9243C

A Generalizable Radiography Algorithm Test Environment for NDE Applications

Edward S. Jimenez, Ismael Perez and Andrew C. Wantuch

Sandia National Laboratories
Software Systems R&D
PO Box 5800; Mail Stop 0932
Albuquerque, NM 87185-0932
(505) 284-9690; Fax: (505) 844-2018; E-mail: esjimen@sandia.gov

1 ABSTRACT

This work demonstrates a generalizable test environment platform that can load multiple algorithms and automatically run a test-set of X-ray Data from a Radiography Image Database to determine their efficacy based on various metrics of interest. The output will be text and graphical data communicating performance such as hit rates, ROC Curves, Proportion correct, and other methods of estimating detection. The platform can easily integrate algorithms written in various programming languages, providing great flexibility and allows for the measurement of similar algorithms deployed in different software environments. This platform has many applications within the Non-Destructive Evaluation Community, including anomaly detection, feature extraction, and quality assurance.

2 INTRODUCTION

The objective of this work is to develop a generalizable algorithm testbed for Radiography NDE applications. This effort is a subset of a larger effort called the Open Threat Assessment Platform (OTAP) that is investigating the development of an open architecture radiography screening prototype in partnership with several screening technology manufacturers to allow third-party vendors to implement and develop detection, recognition, and diagnostic algorithms as well as the ability to integrate and deploy specialized or non-traditional hardware into the screening system. Modularization of the screening system components, both hardware and software is key to developing an open architecture-based screening system that is dynamic and fluid in adapting and integrating new technology.

The standard methodology for a algorithm testbed is quite simple. An algorithm, along with metadata such as a template for execution, is inserted into a specific interface that may also receive input from a database or other repository, it then executes the algorithm according to the execution template, compares performance against ground truth data or against another algorithm with similar output, and in turn generates metrics that describe the performance of the algorithm with respect to that which it is being compared. Metrics are application-specific and are not necessarily trivial in some fields. In some cases, some testbeds are closed loop and automatic [11]. In other cases there is a need to coupling of two or more testbeds in order to modularize more components in the interface [8]. Future versions of the testbed could potentially take advantage of more complex features, but as a proof of concept, this work will use the standard methodology of a testbed which is described in work done by LaViola [7].

3 IMPLEMENTATION

The implementation of the testbed was executing using Matlab GUI development [9] as well as allowing for the rapid implementation of various numerically stable and robust metrics. In addition the ability to perform function calls to executables written in other languages such as C, C++, Fortran Java, Python, and .NET; thereby allowing third-party developers to choose the preferred language to implement their algorithms. Therefore, provide the ability to let them natively without any need to convert the algorithm implementation into another language. The ability to test an algorithm in its original form also allows for the comparison of the same algorithm, implemented in other programming languages, to investigate numerical stability, execution time, accuracy, etc.

The input to the testbed may vary depending on the desired metric or ouput. Generally, the testbed will require an image repository, at least one candidate algorithm, and gold standard or ground truth metadata for metrics where the objective is to measure performance against an absolute truth (i.e. signal present/not present). For metrics where multiple algorithms are measured against each other but not necessarily against a known truth or quantity (i.e. raw pixels labeled), the testbed required input would only need the data repository and the candidate algorithms. As shown in figure 1, the data repository and candidate algorithms feed into a task execution environment; generally speaking, there may be instances where the image repository is not in the standard format for the algorithms to operate. One example would be a image repository consisting of TIFF images [2], where the test algorithms may expect DICOM [10], DICONDE [3], or DICOS [12] formats which are popular image formats within the medical, non-destructive evaluation, and security communities, respectively. In these cases, the task execution would pre-process the raw image data (the repository's native format), into a standard image format such as DICOS, DICOM, DICONDE, or any other image format whether common or not; Matlab allows for rapid development of non-traditional image formats that can be specified by the user as well as conversion to traditional formats such as JPEG [13] or PNG [1]. Once the images have been converted to the standard format, the candidate algorithms are then applied to the images. A comparator is then used to measure each algorithm's output to either the ground-truth metadata or to the output of the other algorithms. The Detected Signal Information from the comparator is then compiled into the desired metrics such as false-positive, false-negative, ROC, or any other desired metrics. Once all the processing is completed, it is visualized to a Graphical User Interface as well as documented to a text file.

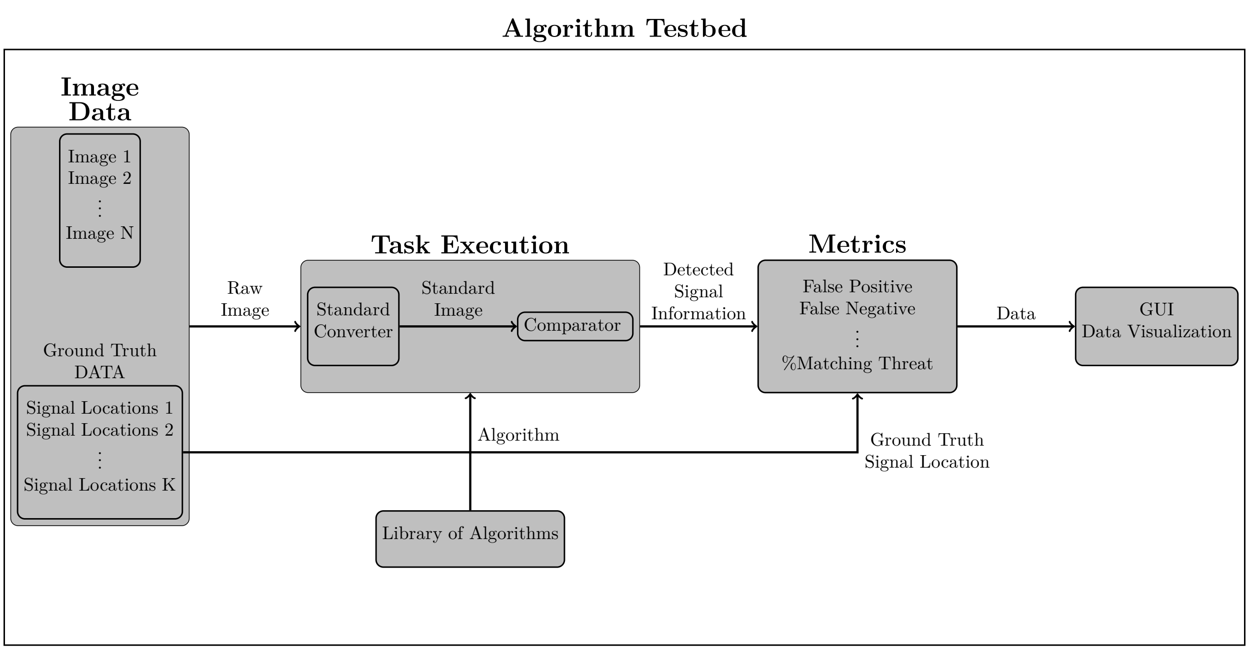

Figure 1: Algorithm Testbed Flowchart.

4 RESULTS

Figure 4 presents a screen shot of the prototype algorithm testbed. In this manifestation, the user will be prompted to select a directory locations where the image repository is stored and a directory where the compiled candidate algorithms reside. Once selected, the user will specify the metrics to perform, in the current implementation, only 2 metrics can be visualized at any given moment, but future versions will allow for arbitrary metrics to be selected and visualized. Once all selections have been made, the testbed will begin the analysis process. The example in figure 4 demonstrates 4 algorithms being compared to each other on the left side of the GUI and against a ground-truth on the right side of the GUI.

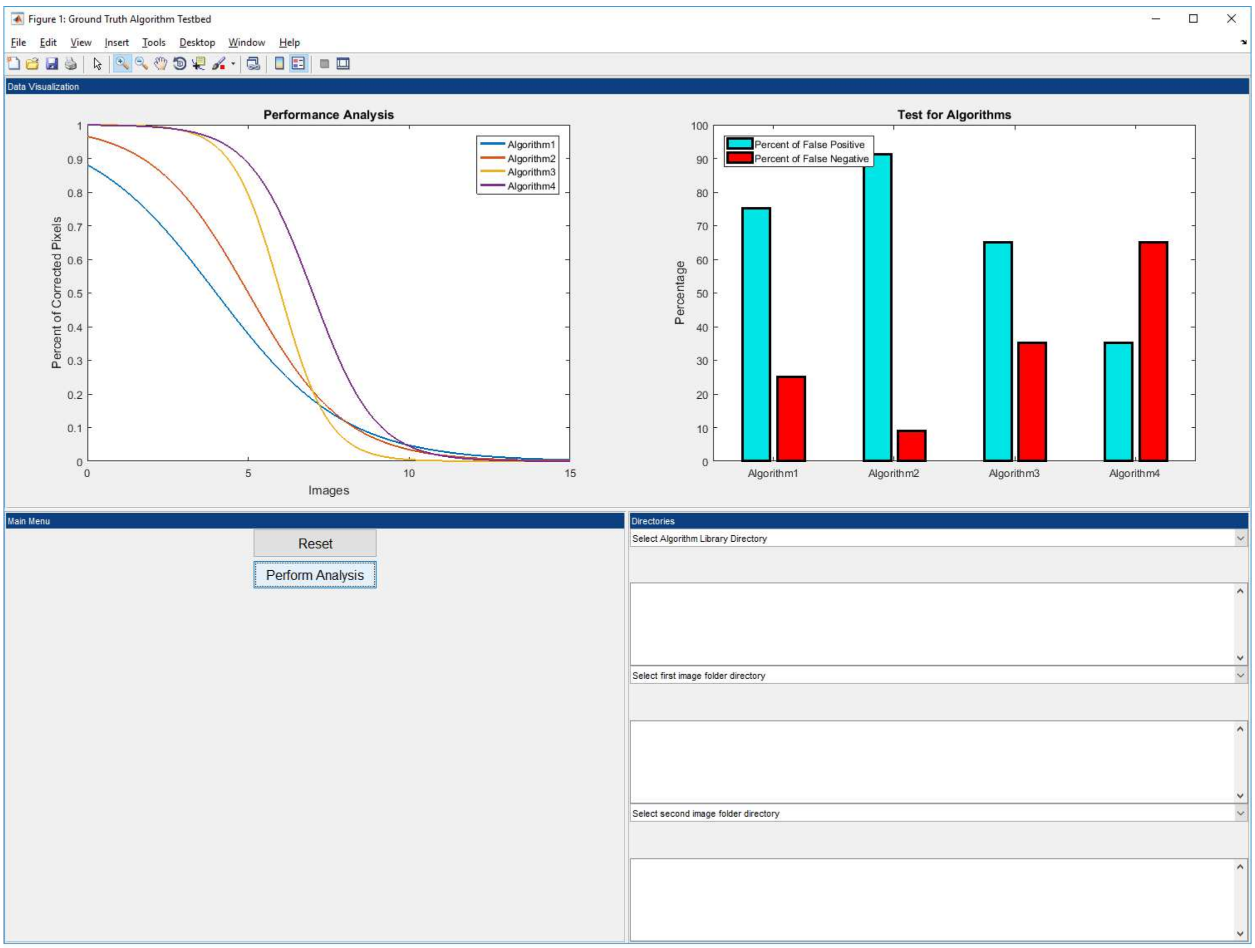

Figure 2: Screenshot of the Algorithm Testbed

5 CONCLUSIONS

This work has demonstrated a very early prototype of a generalizable algorithm testbed that allows for algorithms written in different programming languages to be compared in a common environment without the need to port the algorithm in a different language. The ability to test algorithms of various species will be crucial to the success of the OTAP effort as the ability to verify and validate performance across different environments will allow for the open architecture to flourish. The algorithm testbed will provide the ability to modularly insert algorithms and be able to visualize the performance of each one. Although the main motivation for this work is for security applications, the testbed could easily extend itself to other medical and industrial non-destructive evaluation applications such as quality assurance, verification and validation, anomaly detection, feature extractions, and root-cause analysis.

6 FUTURE WORK

This work has shown there is much potential in a general NDE algorithm testbed albeit much work still needs to be done. Future work includes implementing a variety of metrics, testing against other image repositories, and testing on real-world applications. Other work includes the ability to incorporate spectral images from nascent NDE detector technologies such as the Multix detector that is discussed in Gorecki et. Al. [5] as there exist many preliminary studies that are beginning to investigate these new technologies to improve performance metrics where many traditional algorithms may directly apply such as those by Wurtz et. al. [14] and by Blake and Fogelman [4]. Similarly, new approaches in materials identification using non-destructive evaluation approaches with spectral images such as those proposed by Jimenez et. Al. [6] are performing estimation tasks that are no longer pixel-based and thus traditional metrics specific to image processing may not necessarily apply. Although the testbed could potentially benefit Wurtz et. Al., Blake and Fogelman, and Jimenez et. Al., investigations must be performed to determine the suitable implementation.

7 ACKNOWLEDGEMENTS

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000 476395 SAND2016-6808 C. This work is funded by the United States Transportation Security Agency.

References

[1] Mark Adler, T Boutell, John Bowler, C Brunschen, AM Costello, LD Crocker, A Dilger, O Fromme, J Gailly, C Herborth, et al. Portable network graphics (png) specification. *Specification V1*, 2:W3C, 2003.

[2] Adobe Developers Association et al. Tiff revision 6.0. *Internet publication: http://www. adobe. com/Support/TechNotes. html*, 1992.

[3] ASTM E ASTM. 2339, standard practice for digital imaging and communication in nondestructive evaluation (diconde), 2011.

[4] Glen M Blake and Ignac Fogelman. Technical principles of dual energy x-ray absorptiometry. In *Seminars in nuclear medicine*, volume 27, pages 210–228. Elsevier, 1997.

[5] A Gorecki, A Brambilla, V Moulin, E Gaborieau, P Radisson, and L Verger. Comparing performances of a cdte x-ray spectroscopic detector and an x-ray dual-energy sandwich detector. *Journal of Instrumentation*, 8(11):P11011, 2013.

[6] Edward S Jimenez, Laurel J Orr, and Kyle R Thompson. Object composition identification via mediated-reality supplemented radiographs. In *2014 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, pages 1–6. IEEE, 2014.

[7] Joseph J. LaViola Jr. A testbed for studying and choosing predictive tracking algorithms in virtual enviroments. *EGVE '03 Proceedings of the workshop on Virtual environments 2003*, April 2003.

[8] Matthias Keller, Christoph Robbert, and Manuel Peuster. An evaluation testbed for adaptive, topology-aware deployment of elastic applications. *Proceedings of ACM SIGCOMM 2013*, August 2013.

[9] The Mathworks, Inc., Natick, Massachusetts. *MATLAB version 9.0.0.341360 (R2016a)*, 2016.

[10] Peter Mildenberger, Marco Eichelberg, and Eric Martin. Introduction to the dicom standard. *European radiology*, 12(4):920–927, 2002.

[11] Parashar Barve M. Lutful Hai Kittisak Sakkapongse Durgesh K. Sharma Tina Smilkstein, Kiran Kumar Tati. An evolutionary algorithm testbed for quick implementation of algorithms in hardware. *Evolving and Self-Developing Intelligent Systems, 2009*, April 2009.

[12] NEMA IIC 1 v01. Digital imaging and communications in security (dicos) information object definitions (iods), 2010.

[13] Gregory K Wallace. The jpeg still picture compression standard. *IEEE transactions on consumer electronics*, 38(1):xviii–xxxiv, 1992.

[14] Ron Wurtz, Sean Walston, Dan Dietrich, and Harry Martz. Metrics for developing an endorsed set of radiographic threat surrogates for jinii/caars. Technical report, Lawrence Livermore National Laboratory (LLNL), Livermore, CA, 2009.