Nuclear Instruments and Methods in Physics Research A 784 (2015) 274–280

Contents lists available at ScienceDirect

Nuclear Instruments and Methods in Physics Research A

journal homepage: www.elsevier.com/locate/nima

Data and software tools for gamma radiation spectral threat detection and nuclide identification algorithm development and evaluation

David Portnoy*, Brian Fisher, Daniel Phifer

Johns Hopkins University Applied Physics Laboratory 11100 Johns Hopkins Road Laurel, MD 20723, USA

ARTICLE INFO

Article history:
Received 21 July 2014
Received in revised form
24 October 2014
Accepted 4 November 2014
Available online 14 November 2014

Keywords:
Gamma-ray spectroscopy
Nuclear threat detection
Radiation detection
Nuclide identification algorithms
Radiation detection algorithms

ABSTRACT

The detection of radiological and nuclear threats is extremely important to national security. The federal government is spending significant resources developing new detection systems and attempting to increase the performance of existing ones. The detection of illicit radionuclides that may pose a radiological or nuclear threat is a challenging problem complicated by benign radiation sources (e.g., cat litter and medical treatments), shielding, and large variations in background radiation. Although there is a growing acceptance within the community that concentrating efforts on algorithm development (independent of the specifics of fully assembled systems) has the potential for significant overall system performance gains, there are two major hindrances to advancements in gamma spectral analysis algorithms under the current paradigm: access to data and common performance metrics along with baseline performance measures. Because many of the signatures collected during performance measurement campaigns are classified, dissemination to algorithm developers is extremely limited. This leaves developers no choice but to collect their own data if they are lucky enough to have access to material and sensors. This is often combined with their own definition of metrics for measuring performance. These two conditions make it all but impossible for developers and external reviewers to make meaningful comparisons between algorithms. Without meaningful comparisons, performance advancements become very hard to achieve and (more importantly) recognize. The objective of this work is to overcome these obstacles by developing and freely distributing real and synthetically generated gamma-spectra data sets as well as software tools for performance evaluation with associated performance baselines to national labs, academic institutions, government agencies, and industry. At present, datasets for two tracks, or application domains, have been developed: one that includes temporal spectral data at 1 s time intervals, which represents data collected by a mobile system operating in a dynamic radiation background environment; and one that represents static measurements with a foreground spectrum (background plus source) and a background spectrum. These data include controlled variations in both Source Related Factors (nuclide, nuclide combinations, activities, distances, collection times, shielding configurations, and background spectra) and Detector Related Factors (currently only gain shifts, but resolution changes and non-linear energy calibration errors will be added soon). The software tools will allow the developer to evaluate the performance impact of each of these factors. Although this first implementation is somewhat limited in scope, considering only NaI-based detection systems and two application domains, it is hoped that (with community feedback) a wider range of detector types and applications will be included in the future. This article describes the methods used for dataset creation, the software validation/performance measurement tools, the performance metrics used, and examples of baseline performance.

1. Introduction

The detection of radiological and nuclear threats is extremely important to national security. The federal government is spending significant resources developing new detection systems and attempting to increase the performance of existing ones. The detection of illicit radionuclides that may pose a radiological or nuclear threat is a challenging problem complicated by benign radiation sources (e.g., cat litter and medical treatments), shielding, and large variations in background radiation [1–3].

Radiation detection systems range greatly in size and application. They include large Radiation Portal Monitors (RPM), used for the screening of cargo, vehicles, individuals, etc; portable units mounted on water-borne vehicles [4], air-borne assets, SUVs, and in trailers [5]; man-portable equipment in the form of backpacks [6] and handheld devices (so-called Radiation Isotope Identifiers). Although diverse in design and application, the majority of these

---

* Corresponding author.

http://dx.doi.org/10.1016/j.nima.2014.11.010
0168-9002/

systems rely on gamma radiation energy spectra to identify materials and threats. The evaluation of their performance often requires expensive and time-consuming measurement campaigns using fully assembled systems and real radioactive materials, and (with somewhat less frequency) high-fidelity computer models of the complete systems. These types of evaluations are necessary before final decisions can be made for procurement, specific application domains, and concept of operations (CONOPS). However, relying solely on fully assembled systems, real materials, and high-fidelity computer models can stifle algorithm development and community engagement.

Although there is a growing acceptance within the community that concentrating efforts on algorithm development (independent of the specifics of fully assembled systems) has the potential for significant overall system performance gain [9], there are two major hindrances to advancements in gamma spectral analysis algorithms under the current paradigm: access to data and common performance metrics with baseline performance measures. Because many of the signatures collected during performance measurement campaigns are classified, dissemination to algorithm developers is extremely limited. This leaves developers no choice but to collect their own data if they are lucky enough to have access to material and sensors. This is often combined with their own definition of metrics for measuring performance. (The Department of Homeland Security's (DNDO) GRaDER program [http://www.dhs.gov/guidance-grader-program] has the stated mission of "… [evaluating] commercial off-the-shelf (COTS) Rad/Nuc detection equipment against national consensus standards adopted by the Department of Homeland Security and TCSs, but this service is not free and requires fully assembled detection systems.) These two conditions make it all but impossible for developers and external reviewers to make meaningful comparisons between algorithms. Without meaningful comparisons, performance advancements become very hard to achieve and (more importantly) recognize. The objective of this work is to overcome these obstacles by developing and freely distributing real and synthetically generated gamma-spectra datasets as well as software tools for performance evaluation with associated performance baselines to national labs, academic institutions, government agencies, and industry.

In general, the sensors used in radiation detection systems to measure gamma radiation produce spectra fall into one of three resolution classes: high-resolution, such as high-purity germanium (HPGe); medium-resolution, such as sodium iodide (NaI); and low-resolution, such as polyvinyltoluene (PVT). Although in the future high-resolution and low-resolution data sets will be generated and distributed, currently our datasets are based on the widely used NaI-based gamma radiation sensors, which produce medium resolution spectra.

Although this first implementation is somewhat limited in scope, considering only NaI-based detection systems and two application domains (temporal and static), it is hoped that (with community feedback) a wider range of detector types and applications will be included in the future.

This article describes the methods used for dataset generation and collection, the software validation/performance measurement tools, the performance metrics used, and examples of baseline performance.

2. Data sets

The data are split into three general classes: training, open, and closed. The training data consist of computer generated "pure" spectra for nuclides under randomly sampled shielding conditions and distances, and the nuclides' energy lines with relative intensities. The pure spectra are single nuclide spectra with no background and are meant to provide a developer with ideal examples of the nuclide features likely to be encountered for a particular detector geometry and source configuration (material, activity, shielding, distance). Currently each nuclide has about 2,000 random samples in the training set. The pure spectra provide developers the data required for template-based algorithms, while the energy line data provide developers the data required for peak-finding algorithms. A limited amount of real spectra is provided along with the synthetic computer-generated data for training and/or testing. The open dataset consists of a set of labeled spectra covering various source and shielding configurations. The open dataset is primarily to be used to generate performance baselines for existing algorithms, and to give developers performance targets. The third type of data are the closed data. These data are used by evaluators to perform a second level of validation of algorithms. It consists of data similar to the open dataset, but will also include a more significant amount of real data. The closed datasets will be more tightly controlled than the open datasets and not distributed to everyone.

All data for this work were generated for or collected using NaI-based detectors. These data include both natural gain variations and non-linear energy calibration errors, deliberately added gain variations up to 2%, and perfect gain and energy calibration (for the synthetic source data). The sources and shielding materials and offset distances included in the training data and open data sets are as follows (areal densities of the shielding materials and distances are randomly sampled):

Sources:	^{18}F, ^{40}K, ^{54}Mn, ^{56}Co, ^{57}Co, ^{60}Co, ^{67}Ga, ^{75}Se, ^{82}Rb, ^{88}Y, ^{90}Sr, ^{99}Mo, ^{99}Tc$_{m}$, ^{109}Cd, ^{111}In, ^{113}Sn, ^{123}I, ^{131}I, ^{131}Xe$_{m}$ ^{135}Xe, ^{133}Ba, ^{140}La, ^{137}Cs, ^{152}Eu, ^{156}Eu, ^{166}Ho$_{m}$ ^{169}Yb, ^{182}Ta, ^{192}Ir, ^{194}Ir, ^{200}Tl, ^{201}Tl, ^{202}Tl, ^{207}Bi, ^{210}Po, ^{226}Ra, ^{228}Th, ^{22}Na, ^{232}Th, ^{235}U, ^{237}Np, ^{238}U, ^{239}Pu, ^{241}Am
Shielding Materials:	Water, Carbon, Iron, Tin, Tungsten, Lead, Depleted Uranium (DU)
Distances:	0.5 to 60 m
Max Shielding Thickness:	95% reduction in effective does at 2 m

At present, datasets for two tracks, or application domains, have been developed: one that includes temporal spectral data at 1 s time intervals (data at 0.1 time intervals are also available), which represents data collected by mobile systems operating in a dynamic radiation background environment; and one that represents static measurements with a foreground spectrum (background plus source) and a background spectrum. These data include controlled variations in both Source Related Factors (nuclide, nuclide combinations, activities, distances, collection times, shielding configurations, and background spectra) and Detector Related Factors (currently only gain shifts, but resolution changes and non-linear energy calibration errors will be added soon). The software tools allow the developer to evaluate the performance impact of each of these factors.

The temporal spectral data consist of real background data with and without real sources, simulated temporal background data, utilizing static background measurements taken in 30 different locations in an urban environment, simulated temporal background data with elevated radon levels caused by real rain events collected at the same 30 locations, and simulated temporal background data with synthetic sources. The simulated temporal data was generated first by creating a speed model with varying speeds nominally from 15 to 35 mph and random stops with durations up to 60 s; adjustable maximum acceleration and deceleration rates as well as stochastic speed variations are also included in the model (Fig. 1). Then long-dwell static background measurements