Please cite this paper as: N. Singh, J. T. Horwood, J. M. Aristoff, and J. Murray-Krezan, "Athena: a data-driven anomaly detection and space object classification tool for SSA," in Proceedings of the 26th AAS/AIAA Space Flight Mechanics Meeting, Napa, CA, February 2016 (Paper AAS-16-447)

AAS 16-447

ATHENA: A DATA-DRIVEN ANOMALY DETECTION AND SPACE OBJECT CLASSIFICATION TOOL FOR SSA

Navraj Singh[*], Joshua T. Horwood[†], Jeffrey M. Aristoff[‡], and Jeremy Murray-Krezan[§]

We present *Athena*, a data-driven system we have developed for space object anomaly detection and classification. Although our algorithmic framework is designed for processing multiple data types, *Athena v1.0* focuses primarily on exploiting non-resolved photometric data (light-curves) obtained from optical sensors. The main techniques developed can be viewed as components of a machine learning pipeline, and include (i) feature extraction using ideas inspired by compressed sensing, (ii) unsupervised learning (via robust principal component analysis) for anomaly detection, (iii) supervised learning for object classification, and (iv) a unifying database that enables all of the above. This paper describes the *Athena* system components and demonstrates some of its use cases on both real and simulated photometric data.

INTRODUCTION

The characterization, classification, and identification of near-Earth objects, in multiple regimes of space, is a critical requirement for maintaining space situational awareness (SSA). A related but equally important goal is the development of anomaly detection and identification, and change detection capabilities. Such tasks are critical for supporting the current and future needs of the SSA mission, including the maintenance of a growing catalog of resident space object (RSO) characterizations and classifications, and for providing indications and warnings (I&W) and facilitating threat identification and notification (TIN). These are important goals for maintaining space superiority and control, both from a military and a commercial viewpoint. However, with the ever-growing number of observable objects in space, the problem of detecting, tracking, and cataloging the type and capabilities for each RSO, as well as detecting and identifying anomalies, is becoming increasingly challenging.

Purely kinematic (astrometric) data, while sufficient for tasks such as orbit maintenance, is generally insufficient for higher-level SSA functions such as object classification and anomaly detection/identification. As a result, the analysis of raw images obtained from electro-optical sensors has received much attention in recent years within the SSA community. However, space objects are often difficult to resolve in such images, even with the newest sensors, due to their small relative sizes (especially in deep-space regimes). Fortunately, photometric signatures (i.e., an object's brightness or apparent visual magnitude variation over time, also called a "light curve") have emerged as a

[*]Research Scientist, Numerica Corporation, 5042 Technology Parkway, Suite 100, Fort Collins, CO, 80528.
[†]Senior Research Scientist, Numerica Corporation.
[‡]Program Manager, Numerica Corporation.
[§]Program Manager, Air Force Research Laboratory/RVS, 3550 Aberdeen Avenue SE, Kirtland AFB, NM 87117.

very promising data type extracted from such non-resolved images.[1,2] With new sensors such as the space surveillance telescope (SST) and the space-based space surveillance (SBSS) system already online within the space surveillance network (SSN), and other commercial data sources as well as amateur data sources coming online (e.g., via DARPA's OrbitOutlook program), photometric data collections on near-Earth objects are expected to increase.[3] As a result, advanced algorithms are needed for processing large amounts of light-curve data in support of the aforementioned SSA functions, but this is a difficult task. For example, an Earth-orbiting object's photometric signature is a function of several factors, including its kinematic state, attitude, material composition, size and shape, stability, illumination geometry, atmospheric effects, biases, and sensor characteristics.[4] While the inverse problem of solving for one or more of these factors, given a real photometric signature, can be mathematically formulated, it has proven far more difficult to treat using purely theoretical physics-based models, especially for noisy and sparsely sampled photometric data, due to the following: (i) the challenges in developing physically-realistic models without incorporating hundreds or even thousands of parameters to be estimated, (ii) the limited observability of said parameters in realistic models given sparse data, and (iii) the computational effort required to fit highly-parameterized and complex physical models.[5,6]

With the goal of overcoming these challenges, Numerica has developed *Athena*, a *data-driven* system using modern *machine learning* methods on *sparse feature representations* of photometric data, in order to facilitate high-level I&W by providing anomaly detection, change detection, and object classification reports. The goal of *Athena v1.0* is to provide an end-to-end machine learning-based I&W capability using a variety of historical and real-time photometric data sources.[7] Importantly, *Athena* does not rely on a particular set of theoretical generative models of photometric data, but instead focuses on learning the observable physics for different classes of space objects (as well as for specific space objects) from the data itself, by exploiting historical data collections intelligently and rigorously. The knowledge gained by applying the learning algorithms on historical data ultimately enables real-time I&W capabilities on previously unseen data on both known and unknown space objects. For example, for anomaly detection and identification, *Athena* learns the "normal" behavior (in its feature space) of specific objects of interest or broader object classes, and then detects departure from normality, given new data. For object classification, *Athena* learns the feature behavior patterns of different object classes from historical data. The learned features and classifiers ultimately lead to classification of unknown objects based on new data.

To accomplish these objectives, *Athena* provides (i) feature extraction algorithms based on *compressed sensing*[8] for extracting sparse representations of photometric data, (ii) *unsupervised learning* algorithms[9,10] based on *robust principal component analysis* for anomaly/change detection, (iii) *supervised learning* algorithms for object classification based on the sparse feature representations, and (iv) a *unifying database* that enables the above algorithms. This paper describes the *Athena* data flow pipeline, software architecture, the *Athena* database, and supporting algorithmic components, as well as demonstrates the system's distinguishing features and benefits with specific use case results on both simulated and real data. Results are presented on feature extraction using real photometric data, and on anomaly detection as well as object classification using high-fidelity simulated photometric data obtained via AFRL's *Ananke* simulation software.

PROBLEM STATEMENT AND GOALS

Before describing the *Athena* system components, we first provide the reader a better understanding of the types of problems we aim to tackle with the system, by way of example use cases. Toward

Use Case 1: Detect Anomalies in Data on Known Object

Goal: Detect anomalies in photometric signatures of a ***known*** space object, using real-time photometric collections from one or more EO sensors, for threat identification and notification (TIN)

Primary Actor: Operator/Analyst

Pre-Condition: *Athena* has an established and tested connection with the data source(s) and operator console

Trigger: *Athena* receives a newly collected photometric signature on a ***known*** space object

Main Success Scenario (Story):

1. *Athena* receives a new photometric signature on a ***known*** space object
2. *Athena* retrieves historically collected photometric signatures of the ***same object, or similar objects, in*** **Athena** ***database***
3. *Athena* determines if the new photometric signature is anomalous in some way with respect to the retrieved historical photometric signatures
4. *Athena* compiles and returns an analysis report to the operator consisting of the following: (i) description on the detected anomalies, (ii) a measure of confidence in any detected anomalies, (iii) auxiliary information to aid operator/analyst in determining source(s) of anomaly
5. Operator/analyst receives the report and takes necessary course of action to resolve the anomaly

(a) Use Case 1: Anomaly Detection on Known Objects

Use Case 2: Classify an Unknown Object

Goal: Classify an ***unknown*** space object, based on its photometric signature, as ***debris/fragment***, ***rocket-body***, or ***active payload***

Primary Actor: Operator/Analyst

Pre-Condition: *Athena* has historical (offline) labeled photometric data belonging to object classes in its pre-defined taxonomy, and *Athena* has an established and tested connection with the data sources and operator console

Trigger: *Athena* receives a newly collected photometric signature on an ***unknown*** space object

Main Success Scenario (Story):

1. *Athena* receives a new photometric signature on an ***unknown*** space object
2. *Athena* notifies the operator/analyst of the data collection
3. Operator/analyst request *Athena* to classify this new photometric signature as ***debris/fragment***, ***rocket-body***, or ***active payload***
4. *Athena* compiles and returns an analysis report to the operator consisting of the following: (i) the determined object class, (ii) a measure of confidence in the classification decision, (iii) information on what subsets of its historical offline database are similar to the newly collected data, and (iv) auxiliary information to aid operator/analyst in understanding the classification decision

(b) Use Case 2: Classify an Unknown Object

Figure 1. Example Use Cases of Athena

that end, Figure 1 specifies two example use cases of *Athena*. The first use case describes a scenario where an analyst uses the system to determine if a newly collected photometric signature on a *known* RSO is anomalous in some way with respect to historically collected signatures on the same object. The second use case describes a scenario where a new photometric signature data collection on an *unknown* object is used to classify the object into a pre-defined taxonomy of object classes. We note that several variations of these use cases are possible with *Athena*. However, in this paper we focus on presenting results on these two use cases.

ATHENA SYSTEM ARCHITECTURE

We now describe our system architecture that aims to fulfill use cases such as the ones described above. The architecture, shown in Figure 2, can be viewed as a machine learning pipeline. The blocks are divided into two main components: *learning* and *generalization*. The former contains blocks relevant for the training phase of algorithm development, and the latter contains blocks relevant for applying the learned classifiers and anomaly detectors to newly acquired real-time photometric signatures. The training data consists of real photometric signatures (light curves) and their feature representations available in a tagged (labeled) RSO observations database, and (optionally) high-fidelity simulated photometric signatures. For the latter, Numerica has leveraged AFRL's *Ananke* Simulation Software, for which Numerica has served as the software integrator in the past. The *Athena* database is a critical component in the pipeline, and details on it are provided in the next section.

The output of the training phase is a set of learned anomaly detectors, as well as classifiers for object classification. The learned classifiers and unsupervised anomaly detectors then act on newly-

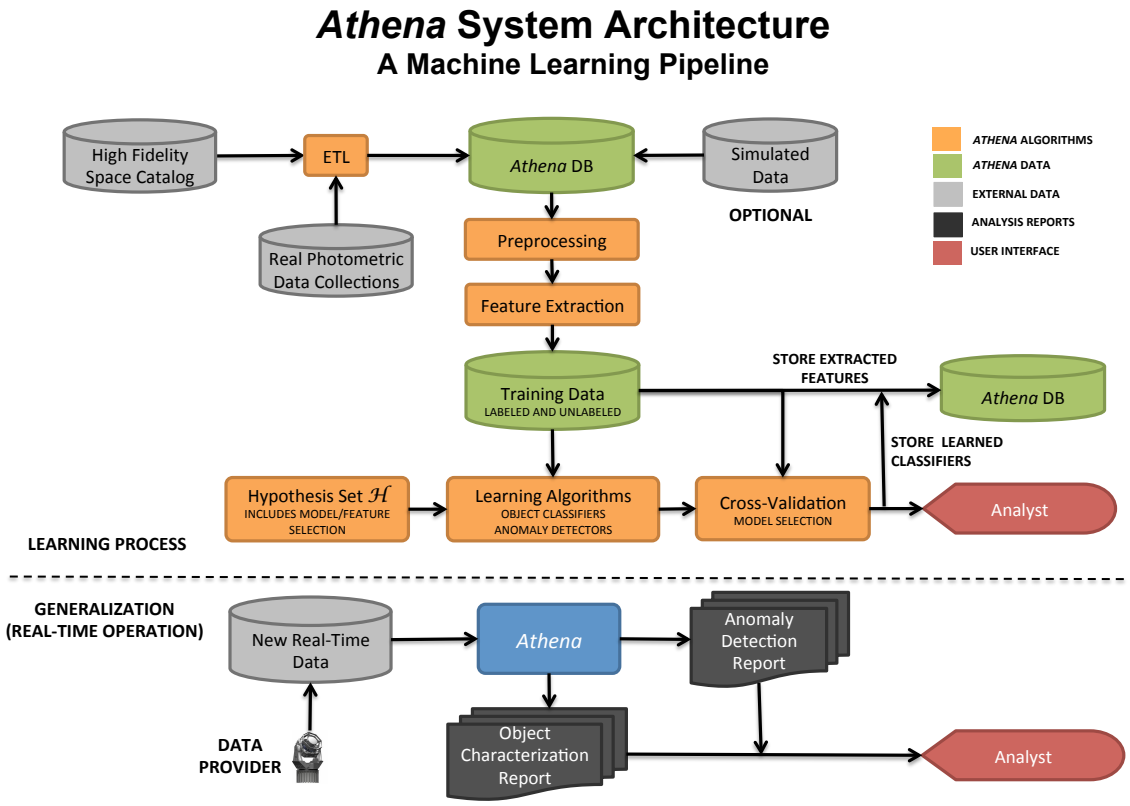


Figure 2. ***Athena*** **System Architecture**

acquired real-time photometric signatures to provide space object characterization and classification outputs, and anomaly/change detections (as I&W outputs). The compiled output reports are then envisioned to automatically update feature-augmented object catalogs, and sent to an analyst dashboard. In this paper, we first describe the *Athena* database, our feature extraction algorithm, an unsupervised anomaly detection algorithm, and object classification framework. We follow the technical discussions with results on feature extraction, anomaly detection, and object classification.

ATHENA DATABASE

To facilitate our data-driven approach, we have developed a database, called the *Athena database*, that facilitates the formation of training, cross-validation, and testing datasets by bringing together data from a variety of sources. The photometric signature datasets that we identified as relevant for the anomaly detection and object classification components of *Athena* consist of a database of labeled, or tagged, "light curves" as a function of time or solar phase angle. For the purpose of object classification, the labels are be assumed to contain, in addition to object ID, an object class and/or status identifier (such as active/inactive, whether the object is a known piece of debris, a rocket body, its operating mode, etc.). For cataloged objects, such status information can be obtained from external sources, such as the European Space Agency's (ESA) Classification of Geosynchronous Objects.[11] Such a database serves the training, cross validation, and testing phases of algorithm development. Currently, the *Athena v1.0* database contains the following historical data sources:

(i) private or closely-held data sources providing raw observations (photometric data), (ii) public data sources providing frequently updated satellite catalog information (such as TLE ephemerides), and (iii) public data sources containing static (occasionally updated) known space object catalog metadata (such as satellite country of origin, launch date, object type, mission, etc).

The database systems used in the *Athena DB* are `SQLite` and `MongoDB`. The `SQLite` database is used for the raw data and enforces a schema. However, the `MongoDB` database allows a flexible data model that allows the storage of more composite entities, such as entire light curves. The primary benefit of forming such a database is that disparate data sources such as those mentioned above are brought together in a unified, queryable location. For example, combining data from different sources allows queries such as the following: "*Retrieve all historical light curves from sensor ABQ1 on all objects classified as 3-axis stabilized which were launched between 1975 and 1990 by XYZ company*", or "*Retrieve all available historical light curves on objects that are classified as being in 'drift mode'.*" Thus, the *Athena* algorithms benefit from several advantages of such a database: (i) ability to cross reference different data sources to form queries not possible using a single data source alone; (ii) ability to easily explore and visualize historical data collections; and (iii) ability to form labeled/unlabeled training datasets, a prerequisite for our machine learning algorithms.

Online (Live, Real-Time) vs. Offline (Historical) Data Sources The *Athena Database* ingests two types of data sources: (i) online, or live and real-time (or near real-time) sources; and (ii) offline or historical data collections. Indeed, *Athena's* data-driven approach relies heavily on historical data collections in the background learning processes. The *offline* data sources consist of historically collected data, available in "batch mode". Such datasets serve in formation of training and cross-validation datasets for the learning algorithms in *Athena*. The *online* data sources consist of new data, available in real-time or near real-time. Usually, the online data would be the data plugged into *Athena* in an operational system or a test-bed, and would be the data that *Athena* would be used to analyze in most cases (although, *Athena* could just as well be used to perform analyses on subsets of the historical data as well). While analyzing online data, *Athena* still uses its offline database in the background learning processes, in order to make inferences on the online data streams.

FEATURE EXTRACTION ALGORITHMS

We now discuss the sparse feature extraction component of *Athena*. Within the machine learning context and, in particular, in the context of processing photometric data, we will use the term *feature* to refer to any quantity derived from the raw photometric signature, including the raw photometric signature itself. The analysis of light curve data from both ground-based and space-based telescopes presents many challenges. In particular, each light curve is a complex amalgamation of numerous affects, including the design of the spacecraft of interest (e.g., the number of solar panels), the spacecraft orientation, and position of the spacecraft relative to the Earth and Sun. Moreover, real data is noisy and often sparse. Accordingly, as we will discuss herein, advanced techniques are required to extract features from such data. Extracting and selecting the right set of features (i.e., the set of features that is best able to discriminate between different types of photometric signatures obtained from different classes of space objects, and are best able to reveal anomalies) are critical in developing space object characterization, classification, and anomaly detection algorithms.

To make our exposition precise let us begin by defining some notation. An individual light curve will be represented by a column vector $y \in \mathbb{R}^{n\times 1}$ and collection of m light curves encoded as the columns of a matrix will be represented by $Y \in \mathbb{R}^{n\times m}$. For feature extraction, we reserve the letter

$D \in \mathbb{R}^{n \times p}$ to be a matrix where each of the p columns is a *basis function* that we use to represent y. Such a matrix D will often be called a *dictionary*. A collection of coefficients that represents y in a dictionary D will be denoted $x \in \mathbb{R}^{p \times 1}$. Accordingly, in the simplest case, we will have that $Dx = y$.

Thus, the collection of all features describing an object can be represented as a vector $x \in \mathbb{R}^p$, which we will refer to as a *feature vector*. By *feature extraction*, we mean the task of computing such a feature vector from the raw photometric signature data. Although the raw signature itself can be considered a feature vector (in which case $D = I$, the identity matrix), for the learning algorithms to work well, more informative and compact features will be required. Said another way, we seek physics-based *sparse* feature representations, by which we mean a representation of a raw photometric signature in a transform domain (defined by D) in which only a few coefficients describe the raw data well. These few coefficients in turn describe the important aspects of signatures. These representations are in turn key in the tasks of characterization and classification of space objects, as well as in revealing anomalies that may be difficult to detect from the raw data alone. A critical component of the *Athena* pipeline is the feature extraction component that *automatically* extracts sparse feature representations in a mixture of transform domains.

Sparse Feature Extraction Using Overcomplete Dictionaries Photometric signatures are known to have embedded in them local characteristics dependent on aspects such as RSO stability state, material composition, orientation, etc. Capturing these features is often difficult due to issues such as noise in the data and insufficient samples. In this section, we describe some techniques to tackle the feature extraction challenges. The methods that we use are based upon recent work in transform domain representation using ℓ_1 techniques,[12] sometimes referred to as *basis pursuit*, and closely related fields such as *compressed sensing*.[13, 12, 14] Such techniques are an active area of current research[15, 16, 17, 18] (and references therein) and revolve around *sparse representations* of data. Our algorithms are focused on using multiple *overcomplete* dictionaries of basis functions for representing photometric data. How does one extract local informative features in a light curve? The underlying principle is that different dictionaries can be chosen that have affinity for particular types of features. As the dictionaries compete to see which provides the "best" representation of the data (i.e., representations that uses the fewest coefficients), those dictionaries that are the "winners" illuminate which features are present. Developing precise definitions of "best" and "winner" basis functions that capture photometric signature characteristics of interest is a key component of our work. If one can find a representation that uses a sparse collection of coefficients across multiple dictionaries, which fits the data well, then that representation provides strong statistical evidence of underlying structure in the light curve.

While various components of the discussion above are known in the existing compressed sensing literature, Numerica's particular techniques provide several novel features, such as treatment of pointwise noise constraints, not found in other methods, that facilitate robust photometric feature extraction from noisy and corrupted data. We describe our specific optimization problem formulation below. While naive searches for sparse representations are computationally very expensive (in fact they are NP-hard), modern techniques exist for solving such problems very efficiently. Building upon the fundamentally important work in Refs. [15, 16, 17, 18] which transforms the NP-hard ℓ_1 optimization problem into a convex ℓ_1 problem, Numerica has made many recent improvements that reduce the required runtime of standard interior point methods[19] by several orders of magnitude, allowing many dictionaries (and therefore many feature classes) to be detected across large

data sets.

One of our key innovations is the development of algorithms that take advantage of pointwise error bounds in a computationally efficient manner while guaranteeing that no error bound is ever violated, thus allowing us to process noisy data where the noise can be time or sensor dependent. Our formulation also allows us to handle *missing data* and *corrupted data* to some extent. We now provide some more details on our optimization problem formulation for extracting sparse feature representations.

Basis Pursuit and Compressed Sensing with Point-Wise Error Constraints The two closely related topics of compressed sensing[12] and basis pursuit[15] have been active areas of research over the past several years. To place the current work in context, we provide a brief overview here. Consider a photometric signal vector $y \in \mathbb{R}^m$. Basis pursuit revolves around the sparse representation of the vector y using a collection of dictionaries D. Specifically, given a matrix $D \in \mathbb{R}^{m \times p}$, with $p > m$, it is of interest to find an $x \in \mathbb{R}^p$ such that $Dx = y$. Again, there are many such solutions x and the principle is again to choose the solution which is the sparsest. Note that while x is high dimensional, if D is chosen "wisely", then there exists a solution x that has very few non-zero entries.

Such a basis pursuit algorithm has several advantages. First, the collection of dictionaries D can be chosen to represent different light curve features of interest. Second, point-wise error constraints are straightforward to guarantee. The point-wise error constraints are important as they help prevent the sparse solution of interest from being contaminated by noise from the sensor. In other words, they bound the size of the perturbations that will be considered meaningful for feature extraction.

Of course, basis pursuit methods require modification for the current context. In particular, classic basis pursuit methods can be computationally expensive. Fortunately, Numerica, as well as many others,[20] have performed substantial work in the area of efficient basis pursuit algorithms, and one of the capabilities that Numerica has developed for the current work is a novel scheme for point-wise constrained basis pursuit problems. In the language of constrained optimization, basis pursuit for a raw 1-D signature/signal y with a dictionary D can be written as the following problem:

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \|x\|_0 \\ \text{subject to} & Dx = y, \end{array} \tag{1}$$

Unfortunately, the problem in (1) is NP-hard, and is not computationally tractable, except for small problems. One of the key insights in Ref. [13] is that (1) can be relaxed to a *convex* problem as

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \|x\|_1 \\ \text{subject to} & Dx = y, \end{array} \tag{2}$$

where $\|\cdot\|_1$ is the ℓ_1 norm (i.e., the sum of the absolute values of the entries of of the vector). The idea is that, under constraints on the sparsity of x and mild constraints on the form of A (A is required to satisfy a condition such as a restricted isometry principle[13]), the solutions to (1) and (2) are the same. Also, since (2) is convex, the problem can quickly and optimally be solved using convex optimization techniques.[21]

The key idea of (1) and (2) is that, among the infinitely many possible representations of y using columns from D, there is a distinguished representation for which x happens to be sparse. Such a parsimonious representation is indicative of the underlying structure of the light curve y. Accordingly, the idea of feature extraction using these techniques is to place into D columns which

represent features of interest. Classically, if too many columns were added to D then one would be in danger of over-fitting the light curve, and masking the true structure of the data. On the other hand, the ℓ_1 regularization prevents over-fitting and emphasizes parsimonious explanations for the observed light curve, encoded in the coefficient vector x. Such sparse sets of features, based upon carefully chosen basis functions, forms the basis of our downstream spacecraft assessment and identification algorithms.

In practice, however, the signal y can be noisy. Thus, we may not want to impose the constraint $Dx = y$ exactly. In this case, the following statistical variation, also known as the LASSO, can be solved:

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \|x\|_1 \\ \text{subject to} & \|Dx - y\|_2 \leq \epsilon, \end{array} \tag{3}$$

While (2) and (3) are quite classic, Numerica has developed techniques for making feature detection in light curves *robust to time-varying noise*. In particular, we have leveraged techniques that control noise on a *point-wise* basis. Such considerations lead to the following problem:

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \|x\|_1 \\ \text{subject to} & |Dx - y| \preceq \epsilon \end{array} \tag{4}$$

where $|\cdot|$ is the *point-wise* absolute value of a vector, $\preceq$ is the *point-wise* less-than-or-equal-to operator for two vectors, and ϵ is a vector of *point-wise* constraints. How can such an ϵ be used in practice? There are many important applications. For example, a large entry in the vector ϵ can be used to indicate missing data. In a similar vein, small entries in ϵ can be used to indicate areas which the user believes is especially important for proper spacecraft identification. Problem (4) can be solved by recasting it as a linear programming problem, but even using advanced interior point techniques such a problem can be very difficult to solve. However, Numerica has developed a fast solver using techniques such as Bregman iterations and ADMM, which solves problem (4), and thus one that allows one to quickly and simultaneously extract photometric data features in the presence of time varying noise.

Designing an Overcomplete Dictionary with Multiple Bases Having defined our optimization problem for feature extraction in (4), our major task is to design the dictionary D in the problem formulation. If the dictionary contains a mixture of bases, then the matrix D will be 'fat' and can be referred to as 'overcomplete'. Indeed, to find a sparse and efficient representation x, we would want an overcomplete dictionary. A pictorial representation is shown in Figure 3.

Readers familiar with compressed sensing might wonder if signal recovery guarantees hold in the case of dictionaries that are redundant (perhaps even highly so) and overcomplete. Indeed, the original compressed sensing guarantees were proved only for orthonormal and incoherent dictionaries. However, it turns out that an overcomplete and highly redundant definition of D can also satisfy properties needed for compressed sensing guarantees for accurate signal recovery to hold. We will skip the details as they are out of scope of this paper, and refer the reader to Ref. [8].

UNSUPERVISED LEARNING ALGORITHMS FOR ANOMALY DETECTION

In this section, we briefly discuss the unsupervised anomaly detection component of *Athena*. In the present context, suppose we are given a large pool of photometric signatures, appropriately calibrated and normalized, collected by one or more optical sensors. Can we determine if there is a

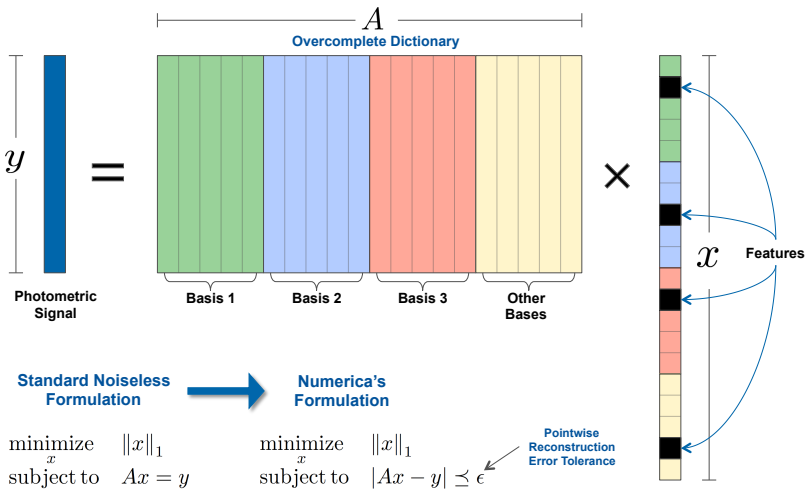


Figure 3. Illustration of Basis Pursuit for Sparse Feature Extraction

subset of light curves that is in some sense different, or "anomalous", when compared to the rest? If so, can we determine this subset, *even if we do not know a priori what an anomalous signature looks like*? Notice that this is an *unsupervised learning* problem, since we are not told what normal or anomalous behavior looks like. We do not attempt to match an *a priori* template to any light curve, but instead allow the pattern to reveal itself in the data by way of dependence or independence of observed photometric data. We define an anomaly as an *abnormal or unexpected temporal dependence* in the photometric signature data collected at one or more sensors. Importantly, our definition of *normal* and *anomalous* arises naturally from the data, and is not externally imposed. An important benefit of such an approach is that an adversary cannot gain any advantage by foreknowledge of earmarked pattern templates.

Although the problem is complex and challenging, Numerica has leveraged recent advances in ℓ_1 techniques for *robust matrix completion, compressed sensing, robust principal component analysis,* and *simple model discovery*, to provide a mathematical foundation to this problem. In addition, Numerica's approach takes into account *uncertainty* in the measurements on which the inferences are based. The details are discussed in Ref. [10], and we only present a brief overview here. A common theme to all these areas is recovery of *low-rank* and *sparse* structures from surprisingly few and noisy measurements.

We formulate the problem by assuming that the extracted feature vectors of all photometric signatures from historical collections on a particular object are represented as a matrix $Y \in \mathbb{R}^{m \times n}$, the rows of which each represent a feature representation of light curves. Each of the m feature vectors is of length n extracted from a signature collected by an optical sensor (this can be achieved after appropriate feature extraction, as we have already discussed). Thus, Y represents a collection of m *events*, where each event is a single light-curve data collection. We then analyze Y using a *latent signal model* and examine the normalized cross correlation matrix $M = YY^T \in \mathbb{R}^{m \times m}$, which is a correlation matrix of the data collection events. We base our calculations on a latent signal model which includes *low rank (L)* and *sparse (S)* components. In this way, we simultaneously discern the low-dimensional structure of the underlying background behavior (the low-rank part), as well as the few anomalous data entries that depart from the nominal description (the sparse part). The

optimization problem we solve for extracting the low-rank and sparse components is

$$\begin{aligned} \underset{L,S}{\text{minimize}} \quad & \|L\|_* + \lambda \|S\|_1 \\ \text{subject to} \quad & |\mathcal{P}_\Omega(M) - \mathcal{P}_\Omega(L+S)| \preceq \epsilon, \end{aligned} \tag{5}$$

where $\|\cdot\|_*$ is the nuclear norm, $\|\cdot\|_1$ is the sum of the absolute values of the entries of the matrix, P_Ω is the projection operator of M onto the index matrix Ω (i.e., M_{ij} is observed when $\Omega_{ij} = 1$ and is not observed when $\Omega_{ij} = 0$), and ϵ is a given constant matrix of bounds on the magnitudes of the entry-wise error. Thus, our formulation allows to deal with partially observed data, and also allows us to ascribe different error tolerances to different sensors. Henceforth, we refer to this optimization problem as the *point-wise error constrained robust principal component analysis* problem (or, in short, the *eRPCA* problem). *Athena* includes a fast solver for solving eRPCA, and example applications of this approach to anomaly detection in simulated photometric data are described in the Results section below.

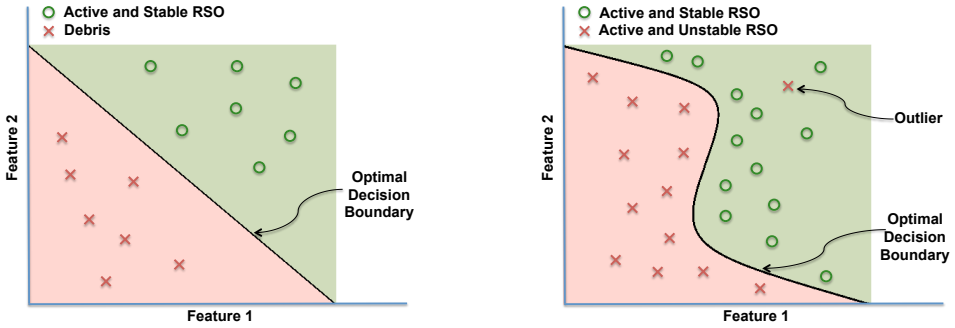
SUPERVISED LEARNING ALGORITHMS FOR OBJECT CLASSIFICATION

In this section, we present some details on the supervised learning component of *Athena*. In the context of supervised learning, we assume the availability of a training dataset, consisting of a set of m *training examples*, $\{(x^{(i)}, y^{(i)})\}_{i=1}^m$. The superscript "$(i)$" is an index into the training set. Let $\mathcal{X}$ denote the space of input or feature vectors (also called the *feature space*) obtained using the feature extraction techniques described above, and $\mathcal{Y}$ denote the output space. In supervised learning, given such a *labeled* training set (i.e., data in which we are given the output y associated with the input feature vector x), our goal is to learn a function $h : \mathcal{X} \to \mathcal{Y}$, such that $h(x)$ is a good predictor for y. The *Athena* database is critical in enabling the construction of such training datasets. The function h is called a *hypothesis*, chosen by the learning algorithm from a set $\mathcal{H}$ of candidate hypotheses. A learning problem is called a *classification* problem if y can take on values only in a finite categorical set. For example, suppose for a space object classification problem setup, y takes on values in the set $\{$DEBRIS, ACTIVE$\}$. Further suppose we have a two-dimensional feature representation of the data collections, perhaps computed using a sparse feature representation framework such as ours. A hypothetical distribution of the training examples in this feature space is shown in Figure 4. The goal of a classification algorithm would be to arrive at an optimal decision boundary that separates the two classes. In this example, there are many possible decision boundaries, and different classification algorithms will arrive at different solutions.

Some examples of supervised machine learning-based classification algorithms include logistic regression (and its multi-class version, softmax regression), linear/Gaussian discriminant analysis, flexible discriminant analysis, k-nearest neighbors, neural networks, naive Bayes, decision trees and random forests, and support vector machines (SVM).

RESULTS

We now present some results that demonstrate the *Athena* pipeline. We first provide an example result of our feature extraction algorithm using a sample of real photometric data. Next, we provide some anomaly detection results using high-fidelity simulated data. Finally, we provide a result that demonstrates the object classification use case. We note that although we have processed and demonstrated anomaly detection in real photometric data via the *Athena* pipeline, for the purposes of this paper we focus on simulated data due to the sensitivity of the (anomaly detection) results obtained on real data.



(a) Classes are linearly separable (b) Classes are non-linearly separable, with an outlier

Figure 4. Example of Classifying Space Objects into One of Two Classes

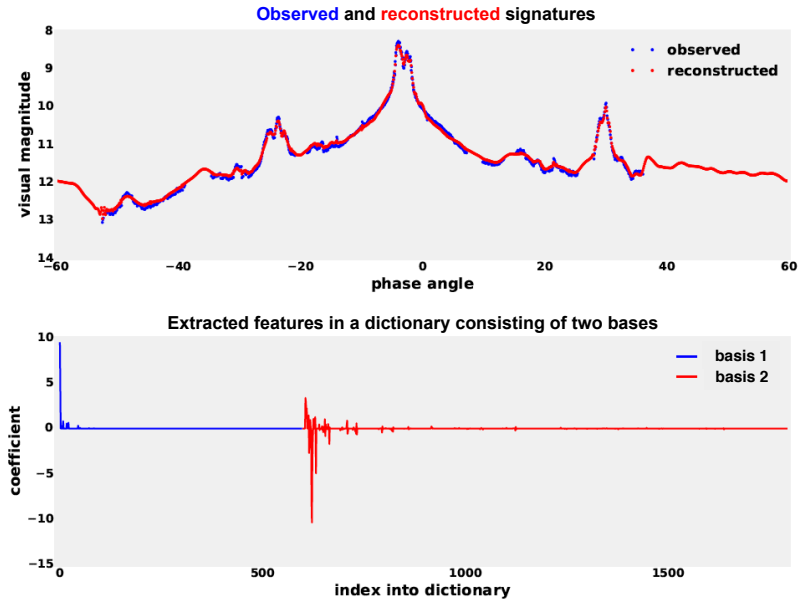


Figure 5. An Example of Sparse Feature Extraction on Real Data (3-Axis Stabilized GEO Satellite)

Feature Extraction Example using Real Data

Feature extraction is a critical prerequisite for downstream *Athena* components such as anomaly detection and object classification, since it is the choice of computed features that enables detection of subtle anomalies or separation of different object classes in feature space. Thus, we first present an illustrative result on sparse feature extraction using a sample of real data. Figure 5 shows feature extraction for photometric data collected on a 3-axis stabilized GEO satellite. The top plot shows (in blue) the observed signature (apparent visual magnitude as a function of the solar phase angle). The red curve is a reconstruction of the signature using the extracted feature representation, which is shown in the bottom plot. In particular, the bottom plot shows the sparse coefficient vector x computed by solving the optimization problem described in (4). In this case, the dictionary matrix D consists of two carefully chosen basis sets. The coefficients corresponding to the first basis are shown in blue, and those corresponding to the second basis are shown in red. As can be seen,

the extracted coefficients are mostly zero, with a few non-zero entries able to capture the original signature well, judging from the reconstruction shown in the top plot (in red). Note that our methods also allow us to also reconstruct interpolated values of the signature in portions where there are no observations. By computing a sparse feature representation of a photometric signature as shown in Figure 5, we are able to capture the important aspects of a noisy and possibly undersampled photometric signature in an parsimonious and efficient manner. The ultimate beneficiaries of such a representation are downstream anomaly detection and object classification components, as we demonstrate next.

Anomaly Detection using Simulated Data

In order to assess the performance of *Athena* algorithms in a reliable manner, it was necessary to be able to generate physically realistic simulated data, such that we have "truth" information available for evaluating the algorithms. Toward that end, we made several updates to AFRL's *Ananke* simulation software, including the following: (i) command line interface for faster light curve simulation, (ii) Monte Carlo simulations support, (iii) ability to vary reflectance properties of different facets in BRDF models, and (iv) ability to simulate ground-based observations. In addition, simulation of anomalies was deemed necessary to test the anomaly detection algorithm performance. Two types of anomaly simulations were implemented: (i) attitude change anomalies, and (ii) uncontrolled spin state anomalies.

We now present several results for anomaly detection on simulated data using the unsupervised learning framework described earlier. The main type of result we show in this paper is the ability to detect anomalous photometric signatures amongst a pool of signatures on the same object or the same type of object. Importantly, no prior knowledge of types of anomalies is assumed here. Instead, anomalous patterns are allowed to emerge automatically by the separation of the raw data into a low-rank and sparse component, as described in our anomaly detection framework discussion above. To illustrate the anomaly detection use case, we now describe the results on four test cases:

- *Test Case 1*: In this test case, 20 lightcurves from similar *uncontrolled and drifting* GEO objects obtained from a ground-based sensor were simulated. These lightcurves are shown in Figure 6 as a function of the equitorial signed phase angle. Three of these signatures, however, are anomalous in the sense that their spin rates are significantly higher than all other data points. These anomalous signatures are circled in red. Figure 7 shows the result of solving our anomaly detection optimization problem, in the form of a heat map of the low-rank (L) and sparse (S) matrices obtained via decomposition of the feature vector cross-correlation matrix. The reader should focus their attention on S matrix, where the anomalous signatures reveal themselves as the high-magnitude components of the matrix (matrix row/column indices 3, 7, and 17, which correspond to signatures 2, 6, and 16 in Figure 6). Clearly, a simple thresholding scheme with an appropriately tuned threshold would detect these signatures as anomalous cases within this pool of signatures.

- *Test Case 2*: The setup here is similar. In this test case, 20 lightcurves from similar *Nadir-pointing* GEO satellites obtained from a ground-based sensor were simulated. These lightcurves are shown in Figure 8 as a function of the equitorial signed phase angle. Two of these signatures, however, are anomalous in the sense that they contain uncontrolled behavior in portions of the signature, generated by artificially removing their control states. These anomalous signatures are circled in red. Figure 9 shows the result of solving our anomaly detection optimization problem, in the

form of a heat map of the low-rank (L) and sparse (S) matrices obtained via decomposition of the feature vector cross-correlation matrix. The anomalous signatures reveal themselves as the high-magnitude components of the S matrix (in this case, signatures 2 and 6). Again, a simple thresholding scheme would detect these anomalies.

- *Test Case 3*: In this test case, 20 lightcurves from similar *Sun-pointing* GEO satellites obtained from a ground-based sensor were simulated. These lightcurves are shown in Figure 10 as a function of the equitorial signed phase angle. All of these cases include an attitude change maneuver in the beginning period of the data collection. Two of these signatures, however, are anomalous in the sense that they contain a rare maneuver in the latter portion of the signature. These anomalous signatures are shown in red. Figure 11 shows the result of solving our optimization problem, in the form of a heat map of the low-rank (L) and sparse (S) matrices obtained via decomposition of the feature vector cross-correlation matrix. The anomalous signatures reveal themselves as the high-magnitude components of the S matrix (in this case, signatures 3 and 17). Importantly, note that if a simple thresholding scheme is applied, then the routine maneuver-induced spikes in the beginning of each case would not be declared as an anomaly, but the rare maneuver states in cases 3 and 17 would be. This determination would be made not from a prior model of what an anomaly looks like, but inferring what is normal and what is anomalous from the data itself.

- *Test Case 4*: In this final test case, again, 20 lightcurves from similar *Sun-pointing* GEO satellites obtained from a ground-based sensor were simulated. These lightcurves are shown in Figure 12 as a function of the equitorial signed phase angle. All of these cases include an uncontrolled state period in the latter portion of the signature. Two of these signatures, however, are anomalous in the sense that they contain a controlled maneuver in the beginning period of the data collection. These anomalous signatures are shown in red. Figure 13 shows the result of solving our anomaly detection optimization problem, in the form of a heat map of the low-rank (L) and sparse (S) matrices obtained via decomposition of the feature vector cross-correlation matrix. The anomalous signatures reveal themselves as the high-magnitude components of the S matrix (in this case, signatures 7 and 15). The difference between this test case and Test Case 3 is that the maneuver in the beginning of cases 7 and 15 is the rare event, and the unstable state in the latter portion of the lightcurves is the common routine event (the situation in Test Case 3 was reversed). A simple thresholding scheme would reveal signatures 7 and 15 as anomalous.

Finally, we note that in all of the above test cases, the use of the feature data matrix (instead of the raw photometric signatures) yields improved detection performance from our algorithms. That is, although not shown in the above results, if we were to perform the eRPCA decomposition on the raw photometric data cross-correlation matrices, then the anomalous signatures do not 'light-up' as conspicuously as they do when using our feature representation instead.

Object Classification Results

In section we present results on object classification on simulated data. The simulation framework used here is as follows. First, the *Ananke* Simulation Software was used to simulate light curves of objects in the following classes: (i) *debris fragment*, (ii) *rocket body*, (iii) *sun-pointing payload*, and (iv) *"M-Shape" payload* (an object type with two solar panel-like facets). Next, 50 Monte Carlo samples were simulated for each class, where each sample is a full light curve. A random subset of the samples was then used as training data for learning classifier, and the rest set aside as the hold-out test set. Finally, the supervised learning pipeline depicted in Figure 15 was executed.

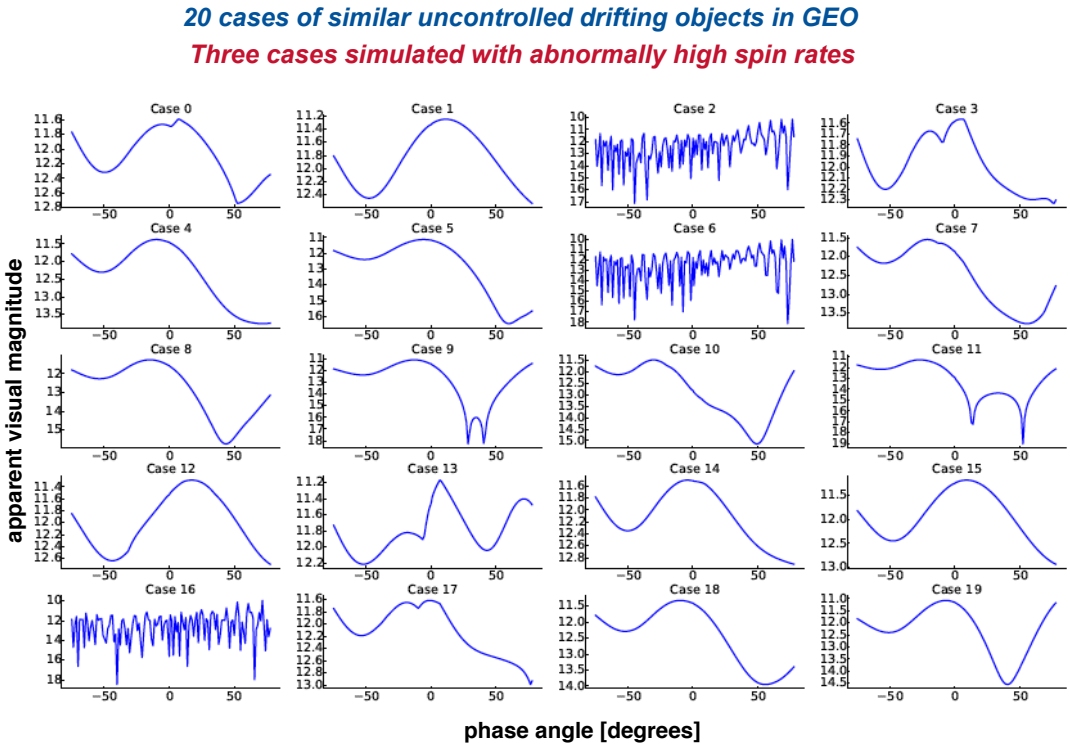


Figure 6. Anomaly Detection Test Case 1: Data

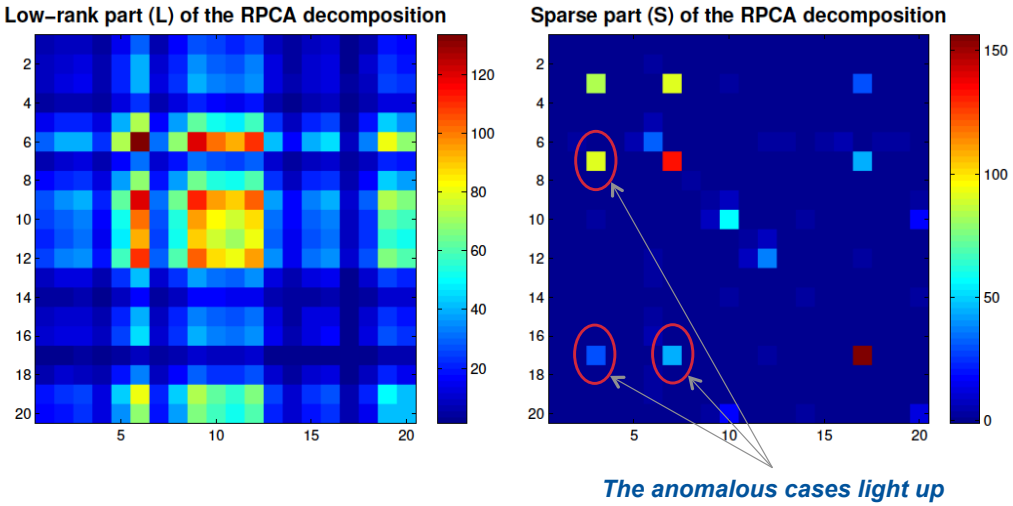


Figure 7. Anomaly Detection Test Case 1: eRPCA Decomposition

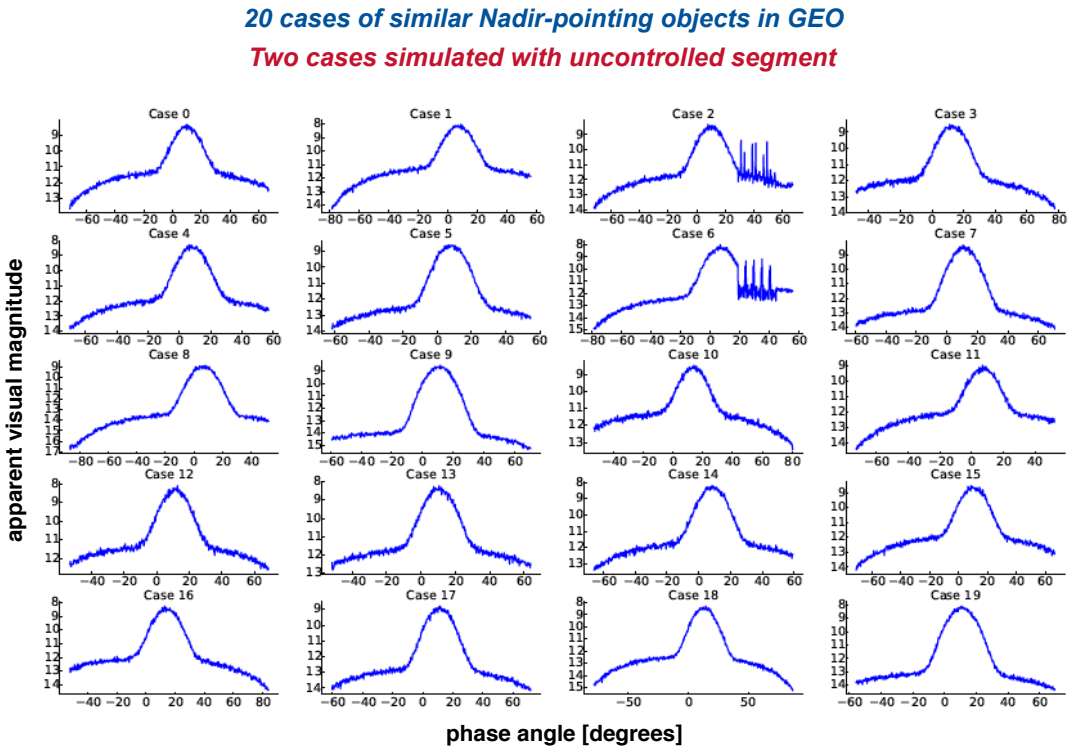


Figure 8. Anomaly Detection Test Case 2: Data

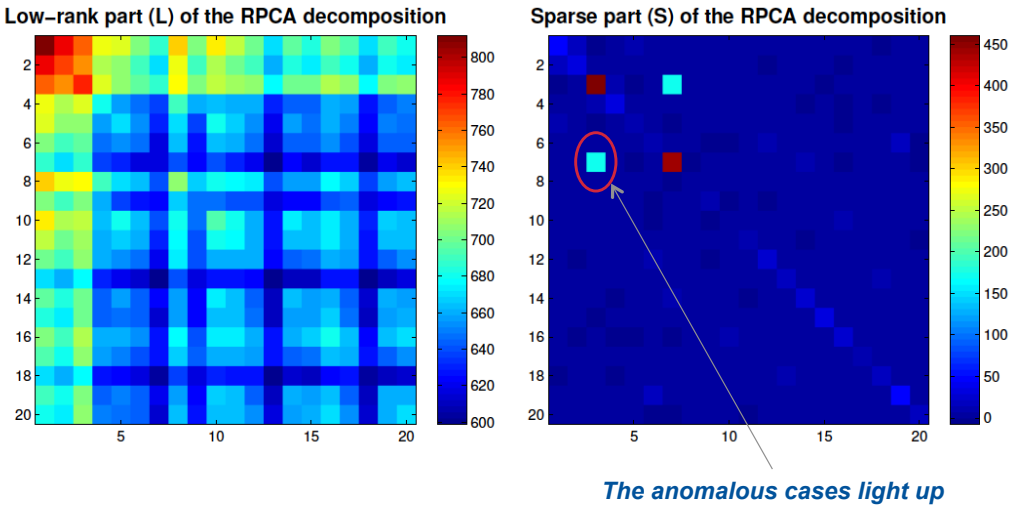


Figure 9. Anomaly Detection Test Case 2: eRPCA Decomposition

20 cases of similar Sun-pointing objects in GEO.

All cases conduct a routine attitude change maneuver during the beginning portion (less than zero phase angle). However, notice two cases simulated with rare maneuver.

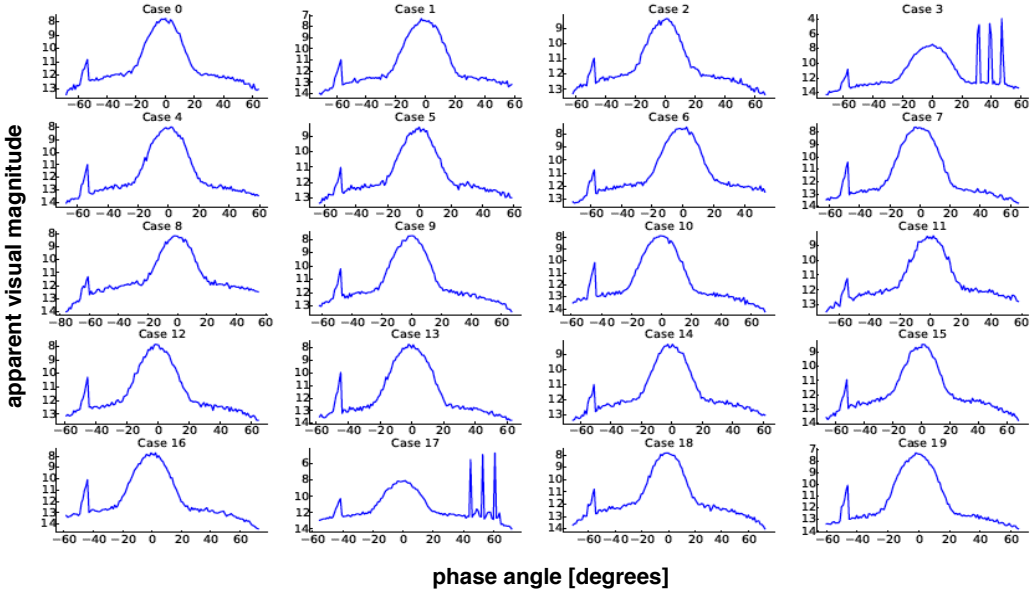


Figure 10. Anomaly Detection Test Case 3: Data

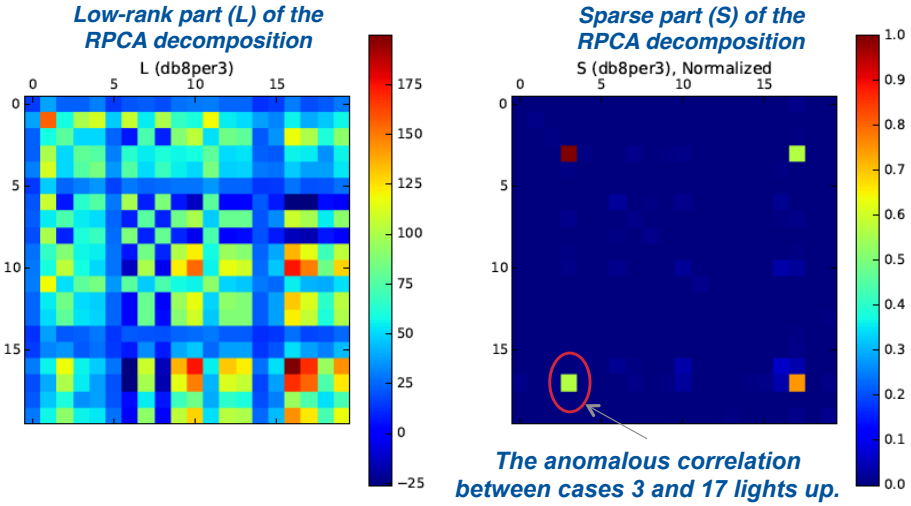


Figure 11. Anomaly Detection Test Case 3: eRPCA Decomposition

20 cases of similar Sun-pointing objects in GEO.

All cases conduct a routine long maneuver in the latter stage (greater than zero phase angle). However, notice 2 cases simulated with rare maneuver.

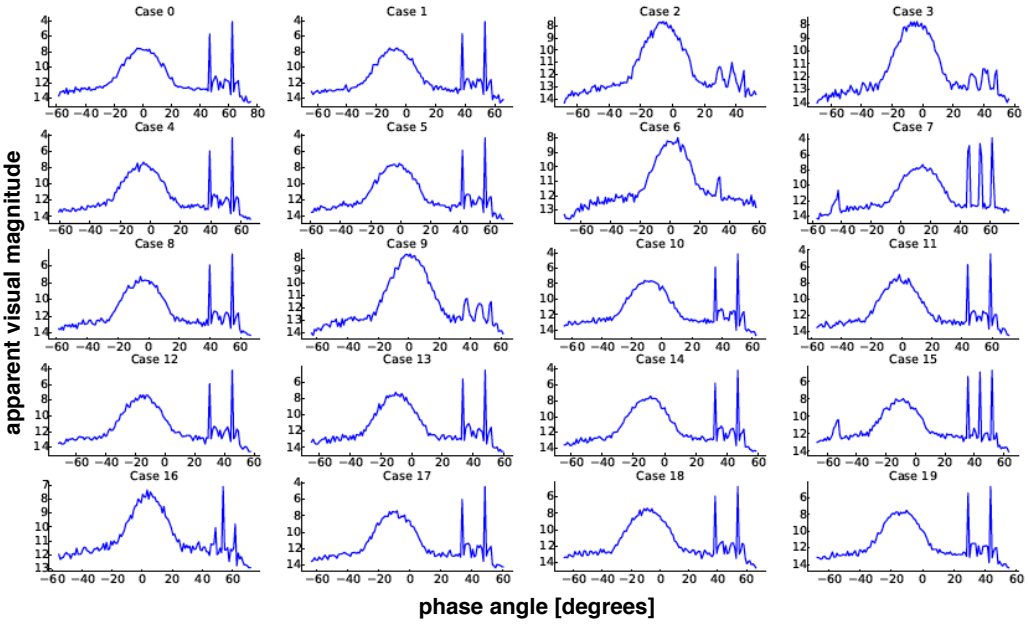


Figure 12. Anomaly Detection Test Case 4: Data

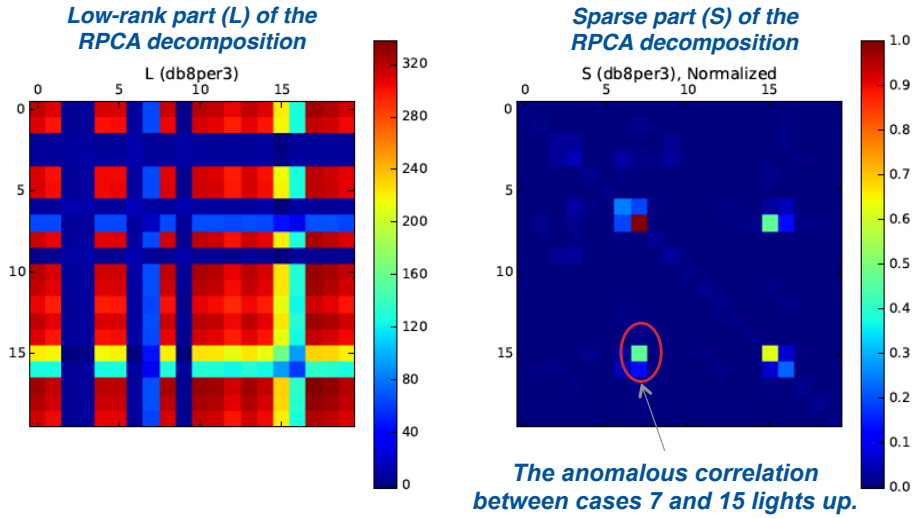


Figure 13. Anomaly Detection Test Case 4: eRPCA Decomposition

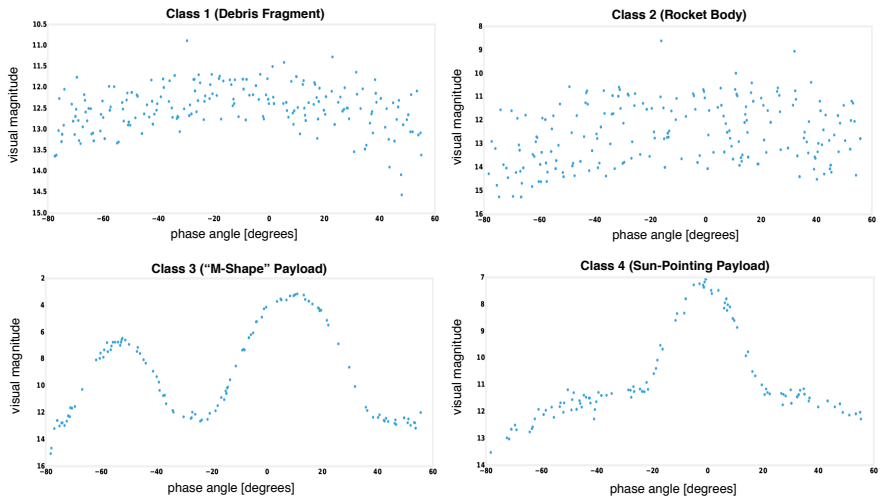


Figure 14. Example Photometric Signatures Simulated for Four Object Classes

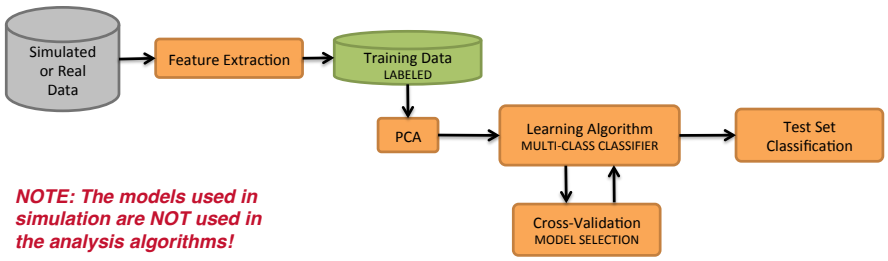


Figure 15. Object Classification Pipeline

All data was simulated with ground-based sensors, and all objects are in GEO. Figure 14 shows examples of photometric signatures simulated for each of the object classes. As discussed earlier in our technical discussion, after feature extraction, the next step in the *Athena* pipeline is learning a classifier that uses these features to discriminate (i.e., learn decision boundaries) between objects of different classes. Given the photometric signatures for the four different object classes, can we learn a classifier that automatically recognizes a newly acquired signature as belonging to one of these classes (or to a different unrepresented class)?

Figure 16 shows the result of running a multi-class support vector machine (SVM), operating on a PCA-reduced representation of the sparse feature representation of the raw data. In particular, as can be seen in the figure, the four object classes appear to be separated in the feature space obtained by keeping the top three principal components. Although not shown, if we had run PCA on the raw signatures directly, we do not obtain such a separation. The decision boundaries learned by the SVM are not shown, since they are difficult to visualize. These boundaries provide optimal separating manifolds that separate the four clusters of space objects. Thus, in this example, the SVM achieves perfect classification performance. We note, however, that as the training data is reduced and the data is simulated with higher noise levels, the separation between the classes reduces and the classification error increases.

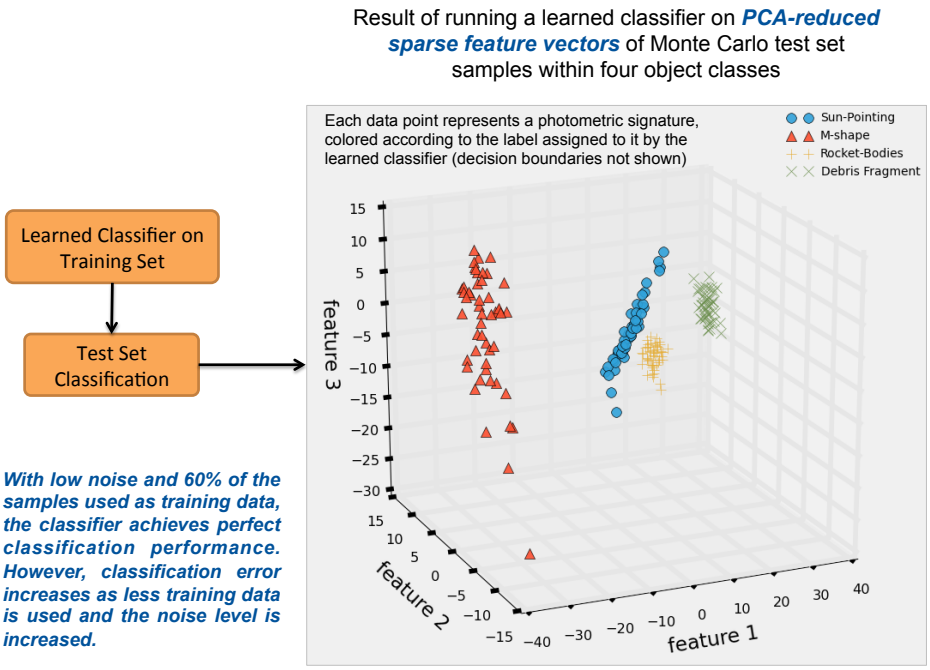


Figure 16. Object Classification Result Visualization

SUMMARY AND CONCLUSIONS

In this paper, we described the main components of the *Athena v1.0* data flow pipeline, including the compressed sensing-based feature extraction algorithms, unsupervised anomaly detection framework, overview of supervised learning for object classification, and the *Athena* database which enables the formation of training and test datasets. We first showed how our feature extraction algorithms can capture the important aspects of real photometric signatures in a faithful yet parsimonious manner. We then showed, via several simulated test cases, the power of our feature representation combined with our unsupervised learning framework, for detecting various type of anomalous events amongst a pool of photometric signature data collections, as well as for separating different object classes of interest in an optimal feature space for doing so. Importantly, our test cases show how our framework does not assume a prior model of normality or anomalous behavior, and instead lets the anomalous events be treated as rare events that reveal themselves by way of a rigorous decomposition of the event cross-correlation matrix. We emphasize that the anomalous events are detected more easily using the sparse feature representation as compared to using the raw photometric data alone. Similarly, it is the feature representation that results in the separation (in feature space) of the historical photometric signatures belonging to the different object classes of interest, which ultimately leads to improved classification performance. Thus, as photometric signature data collections increase within the SSA community in the near future, the primary benefit of the system to operators and analysts is that thousands of signatures can be screened efficiently and automatically for detecting anomalous events and performing an initial object classification into a pre-defined taxonomy, in order to facilitate detailed downstream I&W analysis on only a small subset of large data collections.

ACKNOWLEDGMENTS AND DISTRIBUTION

This work was funded by a Phase I SBIR contract (FA9453-14-M-0155) from the Air Force Research Laboratory, Space Vehicles Directorate. This work has been approved for public release (Case Number OPS-16-1052).

REFERENCES

[1] T. Payne, D. J. Sanchez, S. A. Gregory, L. G. Finkner, E. Caudill, D. M. Payne, L. Kann, and C. K. Davis, "Color photometry of GEO satellites," *Proceedings of the 1999 Space Control Conference*, MIT Lincoln Laboratory, 1999.

[2] A. Chaudhary, C. Birkemeier, S. Gregory, T. Payne, and J. Brown, "Unmixing the materials and mechanics contributions in non-resolved object signatures," *Proceedings of the 2008 Advanced Maui Optical and Space Surveillance Technologies Conference*, Wailea, HI, September 2008.

[3] T. Blake, "Space Domain Awareness (SDA)," *Proceedings of the 2011 Advanced Maui Optical and Space Surveillance Technologies Conference*, Wailea, HI, September 2011.

[4] L. Henderson, P. Goyal, and K. Subbarao, "Inverse Problem Formulation Coupled with Unscented Kalman Filtering for State and Shape Estimation of Space Objects," *Proceedings of the 22nd AAS/AIAA Space Flight Mechanics Meeting*, Charleston, SC, January 2012. Paper AAS-12-115.

[5] R. Linares, M. K. Jah, and J. L. Crassidis, "Inactive space object shape estimation via astrometric and photometric data fusion," *Proceedings of the 22nd AAS/AIAA Space Flight Mechanics Meeting*, Charleston, SC, January 2012. Paper AAS-12-117.

[6] J. M. Aristoff and J. T. Horwood, "Improved Estimation Approaches for High Accuracy Satellite Detection, Tracking, Identification, and Characterization," Final Scientific and Technical Report FA9453-14-M-0051, Air Force Research Laboratory, August 2014.

[7] N. Singh, R. C. Paffenroth, and A. Mont, "Machine Learning using Physics-Based Sparse Feature Representations for Non-Resolved Space Sensing," Final Scientific and Technical Report FA9453-14-M-0155, Air Force Research Laboratory, February 2015.

[8] E. J. Candes, Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Applied and Computational Harmonic Analysis*, Vol. 31, No. 1, 2011, pp. 59–73.

[9] N. Singh, B. A. Miller, N. T. Bliss, and P. J. Wolfe, "Anomalous subgraph detection via sparse principal component analysis," *2011 IEEE Statistical Signal Processing Workshop (SSP)*, 2011, pp. 485–488.

[10] R. Paffenroth, P. Du Toit, R. Nong, L. L. Scharf, A. Jayasumana, and V. Bandara, "Space-time signal processing for distributed pattern detection in sensor networks," *IEEE Journal of Selected Topics in Signal Processing*, Vol. 6, No. 1, 2013.

[11] T. Flohrer, "Classification of Geosynchronous Objects, Issue 15," tech. rep., European Space Agency, February 2013.

[12] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, Vol. 52, 2006, pp. 1289–1306.

[13] E. Candes, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, Vol. 59, 2005, pp. 1207–1223.

[14] E. Candes and T. Tao, "Near-optimal signal recovery from random projections: universal encoding strategies?," *IEEE Transactions on Information Theory*, Vol. 52, 2006, pp. 5245–5406.

[15] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, Vol. 20, No. 1, 1999, pp. 33–61.

[16] D. L. Donoho, Y. Tsaig, I. Drori, and J. L. Starck, "Sparse Solutions of Underdetermined Linear Equations by Stagewise Orthogonal Matching Pursuit," tech. rep., Stanford University, 2006.

[17] E. Candes, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, Vol. 59, No. 8, 2006, pp. 1207–1223.

[18] E. Candes and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, Vol. 23, 2007, pp. 969–986.

[19] J. Dahl and L. Vandenberghe, "CVXOPT,"

[20] W. Yin, S. Osher, D. Goldfarb, and J. Darbon, "Bregman iterative algorithms for ℓ_1-minimization with applications to compressed sensing," *SIAM Journal on Imaging Sciences*, Vol. 1, No. 1, 2008, pp. 143–168.

[21] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, Vol. 3, No. 1, 2010, pp. 1–122.