Multiple Dimensions of LSST Transient Detection: How do we detect things that go bump in the night that we have not yet thought of?

W.T. Vestrand (LANL), A. Becker (UW), S. Perkins (LANL), J. Theiler (LANL), J. A. Tyson(UCD), A. Rest(CTIO), C. Smith(CTIO), C. Stubbs(Harvard), N. B. Suntzeff(CTIO), and P. Wozniak(LANL)

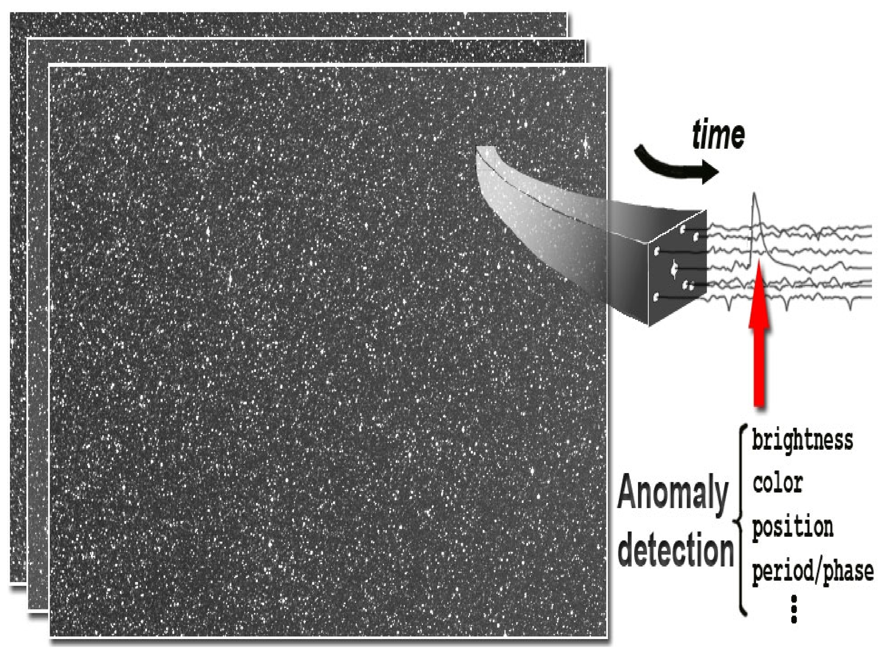
A salient challenge for the Large Synoptic Survey Telescope (LSST) is how to recognize important transients, in real time, in a scene full of normal variations. The data stream will simply be too large for efficient transient identification by human analysts. The broad continuum of properties for both extraneous artifacts and interesting transients make them difficult to deal with on a piecemeal basis with hard-wired code. Further, understanding of the time domain is too incomplete to predict confidently the properties of important changes. We examine the potential of modern Machine Learning (ML) techniques for solving this problem. In particular, we discuss the application of ML techniques for automated anomaly detection that can identify transients without an a priori description. Many anomalies will be instrumentation errors; automating their identification will allow prompt action to maintain LSST data quality. But some of the anomalies are likely to be things that go bump in the night that we have not yet thought of.

•The LSST Opportunity

The huge etendue ($\Omega A \sim 318$ deg^2-m^2) of the LSST will yield fast cadence monitoring of the sky to cosmologically interesting depths.

This unprecedented capability will allow the LSST to detect rare, ephemeral, phenomena ---things that go bump in the night that we have not yet thought of.

To fully exploit this capability, the LSST must be able to recognize and identify important transients in real time so that other instrumentation, like high energy satellites, can be cued to promptly make otherwise impossible observations while the ephemeral behavior is still present.



•The Challenge

The LSST will push our current techniques for real time knowledge extraction **beyond the tipping point.**

•It will generate more than 50 Petabytes of data (~2 Terabyte per hour) that must be mined in real time.

•Real time monitoring will be important for maintaining the health of the instrumentation as well as time domain science.

•More than 10 Billion objects will be monitored for important variations in real time.

•Humans lack attention span, response time, and memory required to monitor the data, recognize important variations, and respond.

•A new automated approach must be developed for knowledge extraction in real time.

•A Solution

To optimize the LSST capability for discovery in Time Domain astronomy, we must develop and integrate the robotic hardware with machine learning technology and context knowledge provided by virtual observatories to create an autonomous "Thinking" LSST.

A key part of this approach is developing methods for applying of three types of machine learning to astronomy:

Automated Feature Extraction: Real time identification of artifacts and transients in direct and difference images.

Classifiers: Automated classification of celestial objects based on temporal and spectral properties.

Anomaly Detection: Real time recognition of important deviations from normal behavior for persistent sources.

Robotic Hardware
LSST Telescope and Camera
Wide-field Monitoring
Real-Time Photometry
and Astrometry Pipeline

Machine Learning
Supervised Leaning
Unsupervised Learning,
ML Classifiers,
Anomaly Detection

Context Knowledge
Record of
Sky Variability,
Massive Distributed
Disk Array

An Autonomous "Thinking" LSST
A powerful Engine for Discovery
in the Time Domain

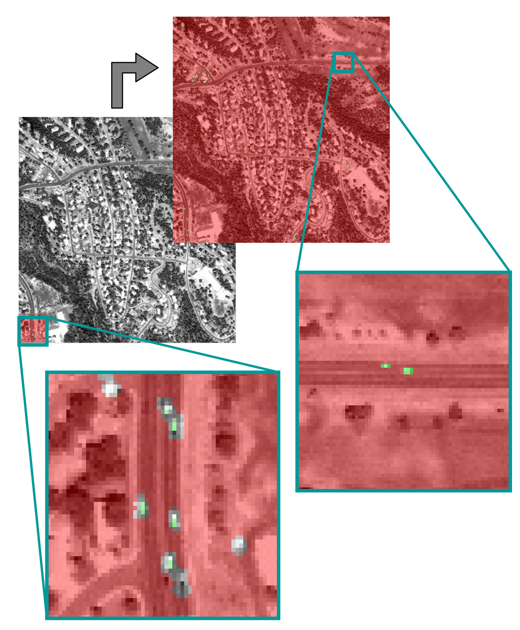
Supervised Machine Learning

Easier to show a machine what to find…

- –Machine Learning derives classification algorithms directly from examples of data

•…than to tell a machine how to find it

- –Requires domain expertise
- –Involves software development
- –Demands careful attention to statistical characterization
- –Entails substantial amount of *trial and error*



What is an Anomaly?

Dictionary Gives two Definitions—

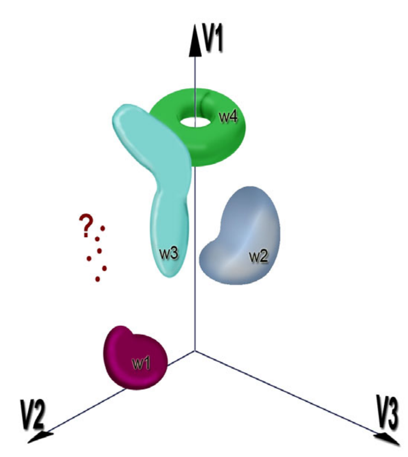
1) Deviation or departure from the normal or common order, form, or rule

2) One that is peculiar, irregular, abnormal, or difficult to classify.

•Anomalies do not permit positive definition of their properties--If we knew what they were, we wouldn't call them anomalies.

Anomaly detection is

an established branch of machine learning that is a fundamentally unsupervised learning method. Using unmarked data samples (from sub-samples of the archive itself), the anomaly detection algorithm provides a "simplest" specification of the data. As illustrated in the figure to the right, similar objects cluster in complex regions of n-dimensional parameter space. Objects in the various feature regions can therefore be flagged as artifacts unworthy of follow-up or as real objects with a priority ranked by the astronomer for follow-up. Data not described by this specification (the question mark points in figure) are treated as anomalies and identified for follow-up studies to determine its properties. Some anomalies will be instrumentation errors, by automating their identification will allow the system to "bootstrap" its knowledge and improve the response. But some of the anomalies will be rare astrophysical objects that are difficult to find any other way.



Example of anomaly detection in an image.

- Ikonos tile: Los Alamos
- True-color, from 4-band MS imagery
- Uses band resampling method to identify anomalous pixels (indicated as darker)

Summary:

The LSST will find normal source variations and instrumental artifacts in every image, the key to success will be real-time identification of important variations in a "forest" of normal variations. To efficiently identify the important "things that go bump in the night", the LSST will require multiple approaches to automated transient detection. For example, the integration of machine learning techniques and context information provided by virtual observatories with the real-time analysis pipeline will be essential for identifying fast transients while they are still present. An advantage of using machine learning (ML) techniques like anomaly detection algorithms is that they can find transients with properties we have not thought of. ML techniques also allow the system to be trained, both by mining its own data and by interacting with human analysts. This will give the system an ability to bootstrap its capabilities by ignoring artifacts "like that" or by finding more "like this" without generating new hard-wired code. ML techniques can therefore enable the efficient construction of queries for the LSST to act as an autonomous discovery engine that searches the night sky.