Kernel Low-Rank Representation for Clustering and Classification
Joseph Wang, Venkatesh Saligrama, David Castanon
joewang@bu.edu, srv@bu.edu, dac@bu.edu

Abstract
In this work, we are developing new approaches for robust feature selection methods that will lead to improved automated classification of explosives, both inrosse in inspection and AIT. Many explosives of interest are based on families with similar chemical composition, and can be expected to exhibit similar patterns in observed characteristics such as X-ray attenuation spectra across frequencies. Our proposed approach uses training data to identify low dimensional nonlinear manifolds around which the data points cluster. These manifolds provide a natural, low-dimensional set of features that can be used to design robust classifiers that perform well with limited training data. The algorithms are based on nonlinear extensions of linear subspace clustering techniques. The performance of the algorithms are illustrated with experiments on different data sets, including a data set of X-ray absorption spectra, and demonstrate significant improvement in clustering and classification performance over conventional techniques.

Methodology
Let $X$ be the available data, with each column containing the measurements corresponding to a sample object measurement $\phi(.)$ is a nonlinear mapping from the data space into a high dimensional space
Main idea: Try to separate images of data points $\phi(X)$ into small numbers of linear subspaces
Resulting problem:
\[
\min \frac{1}{2} \|X - \phi(X)Z\|_F + \lambda \|Z\|_F.
\]
$X$: Data
$Z$: Kernel LRR
$\phi(.)$: Nonlinear basis function
$\| \cdot \|_F$: Nuclear Norm

Columns of $Z$ are the low dimensional representations of the data.
Minimization can be solved without explicitly evaluating the expanded basis, $\phi(.)$, but instead by evaluating the inner products of expanded basis, represented as kernel functions. Using an Inexact Augmented Lagrange Multiplier method, we have developed an algorithm to solve this minimization.

Handwriting Recognition
Given a set of labeled samples of the handwritten digits $1$ and $7$ from the USPS dataset, our goal was to classify unlabeled digits from noisy images.

Using a simple linear classifier on the kernelized low-rank representation, we achieved superior performance compared to linear classification on the linear low-rank representation or kernel support vector machine (SVM), whose performance was matched using $1/5$ the training data, demonstrating the ability to classify with a high degree of accuracy with limited training data.

Explosive Detection
Given a set of x-ray linear-attenuation spectrums sampled at 141 energy values for a set of explosive and non-explosive materials, our goal is to classify an unknown material as explosive or non-explosive.

Using a linear SVM on the kernelized low-rank representation, we achieved a lower classification error rate than traditional kernel SVM methods.

Opportunities for Transition to Customer
Results to date show improved classification performance on real data, especially in the case of limited training data. Of particular interest is the improved explosives detection based on absorption at multiple energies over state-of-the-art methods. We are interested in experimenting with AIT data and with multi-modal data sets, in collaboration with National Laboratory partners, and with video data sets for anomaly detection.

Future Work
Extension to video anomaly detection
Experiments with Explosives data sets
Establishment of performance bounds

Publications Acknowledging DHS Support

References

Acknowledging DHS Support: This material is based upon work supported by the U.S. Department of Homeland Security under Award Number 2008-ST-061-ED0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied of the U.S. Department of Homeland Security.