Naval Research Laboratory

Washington, DC 20375-5320

NRL/MR/5650--09-9206

Intelligent Data Fusion Using Sparse Representations and Nonlinear Dimensionality Reduction

JONATHAN M. NICHOLS
FRANK BUCHOLTZ

Photonics Technology Branch
Optical Sciences Division

JOSEPH V. MICHALOWICZ

Global Strategies Group
Crofton, Maryland

September 28, 2009

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
28-09-2009	Memorandum	15 April 2007 - 30 Sept 2009

4. TITLE AND SUBTITLE

Intelligent Data Fusion Using Sparse Representations and Nonlinear Dimensionality Reduction

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
ONR 61135N

6. AUTHOR(S)

Jonathan M. Nichols, Frank Bucholtz, and Joseph V. Michalowicz*

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
WU EW014-02-41-8955

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Research Laboratory
4555 Overlook Avenue, SW
Washington, DC 20375-5320

8. PERFORMING ORGANIZATION REPORT NUMBER

NRL/MR/5650--09-9206

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Naval Research
One Liberty Center
875 North Randolph Street
Arlington, VA 22203-1995

10. SPONSOR / MONITOR'S ACRONYM(S)

ONR

11. SPONSOR / MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

*Global Strategies Group, Crofton, Maryland

14. ABSTRACT

We propose a new method for performing data fusion and subsequent classification in an information-efficient manner. We argue that an algorithm that can find sparse, low-dimensional representations of data is an excellent candidate for data fusion and classification. Two recent developments in signal processing are investigated: 1) The use of over-determined dictionaries (e.g., frames), and 2) the use of so-called nonlinear dimensionality reduction techniques.

15. SUBJECT TERMS

Data fusion
Image fusion
Frames
Nonlinear dimensionality reduction
Sparse representations

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Frank Bucholtz
Unclassified	Unclassified	Unclassified	UL	13	**19b. TELEPHONE NUMBER** *(include area code)* (202) 767-9342

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Executive Summary

Advances in sensing technology have yielded large complex data sets that often outpace our ability to reliably and efficiently analyze the data. Numerous sensing modalities can be employed in a battlefield environment, each collecting different information pertinent to a given task (e.g. target detection and acquisition). For example, third generation Forward Looking Infrared (FLIR) sensors can record data from a scene in both the Mid-wave IR (MWIR) and Long-wave IR (LWIR) regimes. One would conceivably want to combine the information from these two wavelength regimes in such a way as to improve target detection and classification. More generally, what is needed are the algorithms that can integrate data from different sensing modalities to produce useful information for the warfighter.

This work identifies an appropriate framework for integrating, or fusing, data from multiple sources to produce actionable intelligence. Central to this process is the concept of *dimensionality reduction*. Simply stated, dimensionality reduction is the process of taking a collection of high-dimensional data vectors (e.g., a collection of images) of dimension M and applying an appropriate transformation that results in data vectors of dimension $D<<M$. If the transformation is performed correctly, the useful information in the original M data vectors is preserved in the much-reduced data space of dimension D. **We believe that fusion and classification of high-dimensional data is greatly improved if a proper, low-dimensional representation can be found.**

To achieve this goal we propose using two recent developments in signal processing. Both approaches seek better data representations (models) and, hence, an improved ability to reduce high-dimensional data while preserving the useful information. The first such approach involves using over-determined dictionaries or "frame" representations for the data. This approach transforms the data into just a few appropriately chosen vectors and results in a large information compression. The second technique comprises a class of intelligent data reduction algorithms known collectively as *Nonlinear Dimensionality Reduction* (NLDR). These algorithms have had great success in a limited number of applications where traditional, linear techniques fail.

Manuscript approved August 12, 2009.

In short, both the method of frames and NLDR approaches provide better models for the data. Better data models lead directly to 1) a large reduction in data dimensionality and 2) improved data classification.

Background

The purpose of any dimensionality reduction technique is to intelligently reduce the size of large, complex data sets so that information of interest can be identified and classified quickly and accurately while redundant or unnecessary information is ignored. This is accomplished by transforming the original data to a smaller space that contains only the information of interest. Extraneous information is discarded. This not only improves the process of information extraction but also significantly reduces computational effort. The exact process chosen to accomplish the reduction, however, can lead to vastly different results. We believe that two newly developed techniques, sparse representations (modeling) and nonlinear dimensionality reduction (NLDR), can offer significant improvements in information extraction, fusion and classification over conventional approaches.

In this section we provide a brief introduction to data analysis using overdetermined dictionaries and Nonlinear Dimensionality Reduction (NLDR) techniques. Also discussed is the Support Vector Machine (SVM), a well-known algorithm that can be employed for data classification.

Sparse representations

It is quite common in signal processing to represent a received signal $\vec{y} = \left(y_1, y_2, \ldots, y_M\right)$, consisting of *M* discrete observations, as a linear combination of some basis

$$\vec{y} = \mathbf{A}\vec{x} \qquad \textbf{(1.1)}$$

where **A** is the *M* x *M* set of basis vectors and $\vec{x}$ is the coefficient vector associated with the decomposition. For example, if **A** is the cosine basis the values in $\vec{x}$ are the real parts of the Fourier Transform. If the original signal $\vec{y}$ is taken as a sinusoid then the vector $\vec{x}$ will contain only 1 non-zero component corresponding to the vector in **A** with the same frequency as $\vec{y}$. The decomposition in this case is said to be sparse in the sense that a single number (the non-zero

amplitude of the sinusoid) captures all the information in $\vec{y}$. A proper choice of basis has yielded a reduction in dimensionality from *M* to 1. The reason for such a large dimensionality reduction in this case is that we chose the correct signal model, i.e. the cosine basis. In general, however, we do not know the signal model *a priori*.

The decomposition given by Eqn. 1.1 is actually more restrictive than need be. There is nothing to prevent us from specifying more than *M* (spanning) vectors in modeling the signal. In fact, there has been a recent explosion of literature devoted to the concept of using *overdetermined dictionaries* to represent data[1]. Sometimes referred to as "frames" (provided they meet certain mathematical criteria), these dictionaries require that a constraint be placed on the problem 1.1 as it no longer possesses a unique solution. The constraint must be chosen so that "desirable" solutions can be found. Let $A \in \mathbb{R}^{MxK}$ be an overdetermined dictionary of vectors with $M<K$. We may now write 1.1 as

$$\min_{\vec{x}} \|\vec{x}\|_L \text{ subject to } \vec{y}=\mathbf{A}\vec{x} \qquad \textbf{(1.2)}$$

i.e., solve the overdetermined system of linear equations subject to the constraint that the vector $\vec{x}$ have the smallest possible *L*-norm. In the spirit of the above mentioned Fourier analysis example, the "desirable" solution minimizes the $L=0$ norm, defined as

$$\|\vec{x}\|_0 = \#\{i : x_i \neq 0\} \qquad \textbf{(1.3)}$$

where we simply count the number of non-zero coefficient vectors that result from the decomposition. Heuristically this makes sense as we are attempting to find the representation that takes all *M* observations and reduces their information content to a single number (again, as with the Fourier example). Solutions $\vec{x}$ that are found in this fashion are said to be sparse. Algorithms for finding sparse solutions are currently available including Matching Pursuits[2] or Basis Pursuits and their variations[1].

Typically, the problem is more complicated than the one given by 1.2. We are often in the position of needing to find the best possible dictionary **A** and the associated sparse coefficient vector that results from the dictionary choice. It is also frequently the case that we will have

access to multiple training samples $\vec{y}_i \;\; i = 1 \ldots N$ as opposed to a single piece of data as is implied by 1.2. For this more complicated problem, the optimization becomes

$$\min_{A,\{\vec{x}_i\}_{i=1}^{N}} \sum_{i=1}^{N} \left\|\vec{x}_i\right\|_0 \;\; \text{subject to } \left\|\vec{y}_i - A\vec{x}_i\right\|_2 < \varepsilon, \; 1 \le i \le N \qquad \textbf{(1.4)}$$

That is to say, find a dictionary that minimizes the 2-norm in the reconstruction and, given that dictionary, find the sparsest possible solution. Solutions to this problem may be found in an iterative fashion, first solving for the coefficients $\vec{x}_i$, then solving for the dictionary **A**. As an example, consider the image compression application demonstrated in Fig. 1.

Fig. 1: Results of applying several different compression algorithms to a 180x220 pixel image. Images taken from[1]

The original image is reconstructed using a cosine (JPEG) and wavelet (JPEG-2000) basis as well as the traditional covariance-based PCA algorithm. The final frame shows the results of applying the so-called K-SVD algorithm[1]. This algorithm is one approach to solving 1.4 thereby finding an overcomplete dictionary representation that admits sparse solutions[1]. Shown beneath the figures are the peak signal-to-noise ratios achieved for a given level of compression. Clearly the sparse representations (and corresponding dictionaries) found by solving 1.4 can significantly outperform the results of applying a standard basis.

What is attractive from the data fusion perspective is the ability of these solutions to significantly compress information from very high-dimensional data sources. Again, we believe that fusion is much more easily accomplished in a low dimensional space. Our conceptual view of how the sparse data modeling would lead to data fusion is found in Fig. 2.

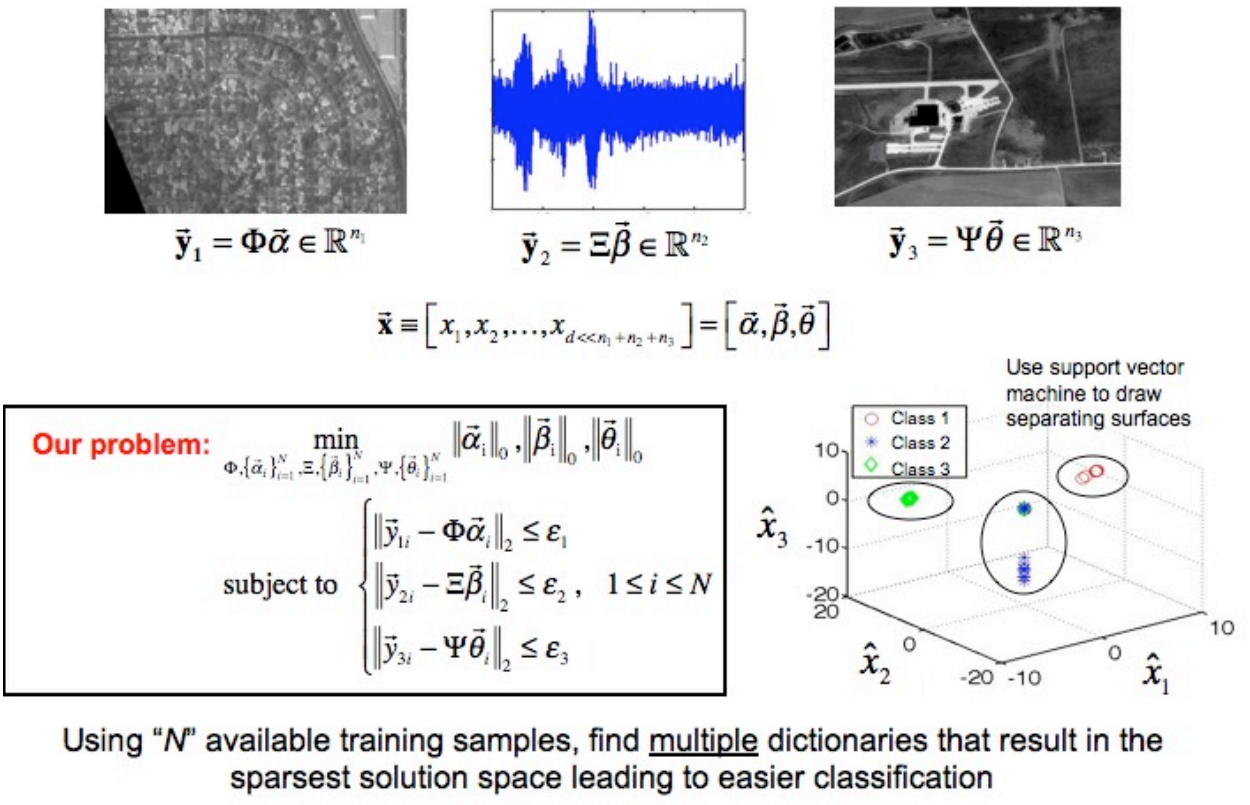

Fig. 2: NRL approach to data fusion. For each individual piece of n_i dimensional data, find the dictionary and coefficient vector that allows for a sparse representation. The concatenated coefficient vector of dimension $d<<n_1+n_2+n_3$ can then be used to classify the data in the low-dimensional space. Support Vector Machines (SVM) are one well-known approach to drawing decision surfaces for classification purposes

For each piece of data we find an associated low-dimensional representation. The concatenation of these representations occupies a greatly reduced data space while still capturing the information present in the original data vectors. Classification efficiency will be greatly improved by working in the reduced, fused space.

To this point we have not discussed the specific classifier used to operate on the low-dimensional space. What is needed for the classification problem is a means of drawing separating surfaces between the different classes of data as shown in Fig. 2. For example, we may wish to distinguish data corresponding to military targets from those corresponding to civilian objects. Assuming that the low-dimensional representation has effectively captured the differences between these two classes and placed them on separate parts of the manifold, we seek a method that can draw a dividing or "decision" surface between the two. This allows for future points to

be classified as belonging to either the military or civilian classes. One effective approach for dividing up a space of arbitrary dimension is the Support Vector Machine (SVM)[3]. The SVM takes training data with known relationships between data and class, and solves the optimization problem of finding the hyper-surface that maximizes separation between nearby vectors (the support vectors) of the different classes. The SVM algorithm is now standard and software for various platforms is readily available[4]. Although not the focus of this work, the SVM is a crucial component to making the final assessment as to which class the data belong. The SVM is illustrated schematically in Fig. 2.

Nonlinear Dimensionality Reduction (NLDR)

A second, complementary approach to sparse signal representations is the idea of nonlinear dimensionality reduction (NLDR). The goal of the NLDR techniques is also to find a low dimensional representation of complex high-dimensional data. Consider N samples of an M-dimensional data vector $\vec{y}_i \in \mathbb{R}^M \quad i = 1 \ldots N$. Each of these vectors can be thought of as a point in M-dimensional Euclidean space. What if we were able to find a much lower D-dimensional representation of the data, $D << M$ such that the data vectors maintain the same relationship to one another as in the high-dimensional space? This is the goal of NLDR approaches.

One of the earliest NLDR techniques was developed in 2000 by Roweis & Saul and is referred to as Locally Linear Embedding (LLE)[5]. This approach begins by building linear models to describe *local* geometric relationships among data points in the original, high-dimensional data space. The new, low-dimensional space is then obtained by projecting the original data in such a way as to preserve this local geometry. Another approach developed around the same time was the ISOMAP approach of Tenenbaum *et al.*[6] Subsequently developed techniques include diffusion maps[7], Hessian eigenmaps[8] and the Laplacian kernel approach of Jones *et al.*[9] In each of the approaches the general goal is the same: construct some measure of local manifold geometry in the high-dimensional space and preserve that measure in projecting down to the low dimensional space. Because these approaches are based on local geometric considerations rather than global linear mappings, they are appropriate in situations for which the high-dimensional data are not linearly separable (an implicit assumption made by traditional approaches).

An important advantage of NLDR compared to conventional approaches concerns how the data are treated mathematically. Conventional approaches typically produce a new, smaller data space from linear combinations of the original data. One common example is the Principal Component Analysis (PCA) approach which seeks linear combinations of the original data axes along which the data shows highest variance, next-highest variance, etc. The assumption of linearity is a severe constraint since *there is no reason to believe that the key pieces of information to be extracted from the data are linearly separable from the noise and clutter*. NLDR approaches recognize this fact and allow the data to be nonlinearly related. The result is a data reduction approach that much more accurately captures the proper information relationships among the data thus allowing for accurate classification. A simple example is shown in Figure 3.

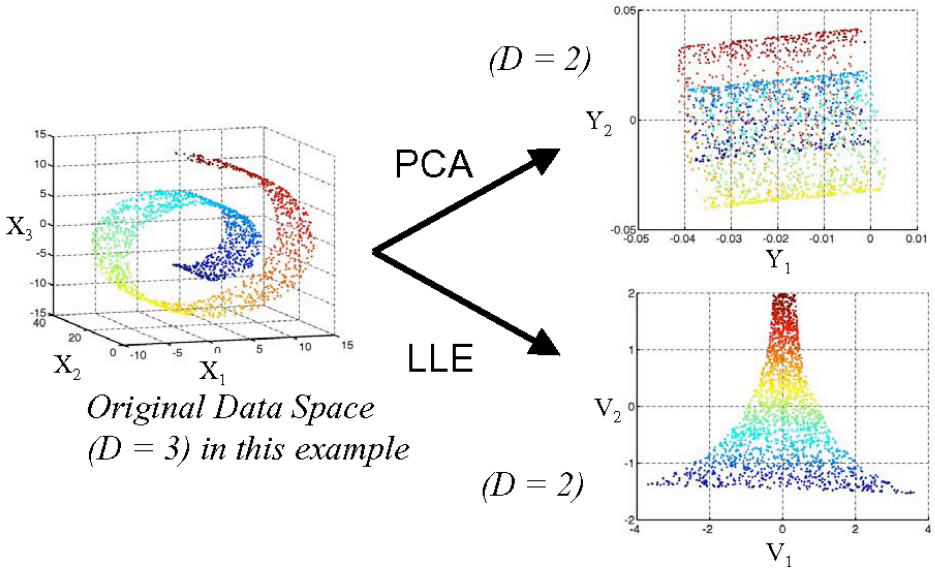

Fig. 3.Comparison of conventional approach to data reduction (PCA) and one of the NLDR approaches, Local Linear Embedding (LLE). The PCA-based reduction cannot resolve the true relationship among the data points and the end result would be a large number of false alarms. LLE reduction preserves the correct information relationships among the data. Here, the PCA approach did NOT simply "squash" the original data onto the x1-x2 plane, rather it formed a linear combination of the x1-x2-x3 axes to obtain new y1-y2 axes depicted in the upper right plot.

In this simple example, the original data lives on a manifold known as the "Swiss roll" – a manifold shape that is particularly useful for illuminating the differences between linear and nonlinear approaches to dimensionality reduction. Here both PCA and LLE algorithms were

applied to obtain a 2-D reduced dimensionality space. Both PCA and LLE mapped points that were close together in the original data to points that are close together in the reduced dimensionality space – this is good. Unfortunately, PCA also maps points that are far away from each other in the original space (dark red and dark blue, for example) on top of each other in the reduced ($D = 2$) space. This will inevitably lead to confusion in the reduced space concerning the information relationship among these points. On the other hand, LLE clearly maintains the proper relationship among the red, yellow, and blue points in the reduced space and classification will be much more accurate in this case.

In general, the NLDR techniques work by first forming a connectivity matrix, or Markov matrix, describing how the high-dimensional data relate to one another geometrically. For example, in the diffusion map technique the matrix is formed as

$$A_{ij} = K(\sigma) = e^{-\|\vec{y}_i - \vec{y}_j\|/\sigma^2} \quad \textbf{(1.5)}$$

where a Gaussian kernel (parameterized by σ)is used to define distances between each sample *i* and every other sample *j*. It can be shown that after proper normalization, the eigen-vectors of this matrix are a geometry preserving, low-dimensional embedding[7]. Other NLDR techniques work in the same fashion: find a sparse matrix that captures local geometric information among the data vectors and take the first "*D*" eigenvectors as the new, reduced dimensionality coordinates.

There are a few ways in which one could conceivably use the NLDR approaches to fuse data from different modalities. The first of these is illustrated in Fig. 4. Data from the individual sensing modalities are first reduced to occupy a low dimensional space. These low dimensional spaces can then be joined together to form the "fused" low-dimensional space. Again, a SVM is envisioned as a good way to divide the low-dimensional space into separate classes.

A second approach would be to combine the data *before* applying the NLDR approaches. In this approach one would effectively be letting the NLDR method perform the fusion implicitly. This approach is illustrated schematically in Figure 5.

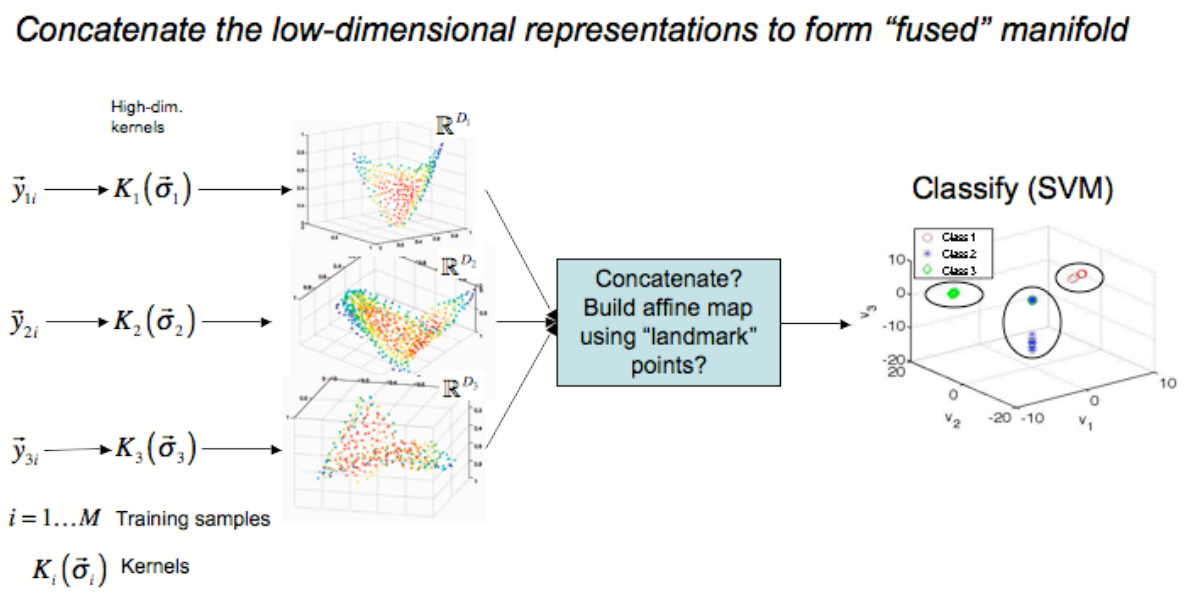

Fig. 4: One approach to using NLDR techniques for data fusion. Each piece of data is reduced to a low-dimensional manifold using the appropriate NLDR technique and appropriate choice of kernel. The sub-spaces are then concatenated together for classification purposes.

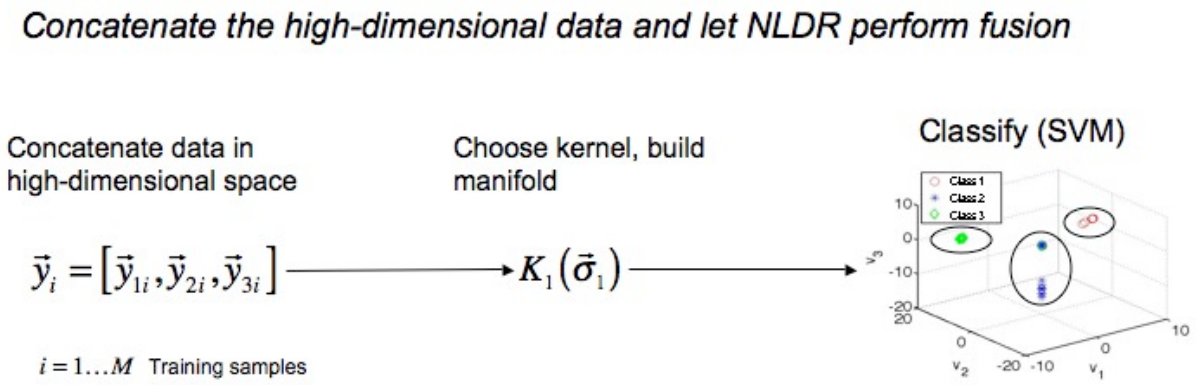

Fig. 5: Schematic showing second possible application of NLDR to data fusion. The data from the different sensing modalities are first concatenated into one large data vector. These samples are then reduced (and fused) using the NLDR approach to produce the low-dimensional space for classification.

One could also conceive of a hybrid approach whereby sparse representations, as found using overdetermined dictionaries, form a relatively low-dimensional manifold. The NLDR techniques could then operate on this manifold in order to get still better classification performance.

Clearly there are many ways to perform dimensionality reduction – nearly all of which have some merit. The eventual choice will, as always, depend on the specific application under study and the sensing modalities involved. The general framework, as presented here, illustrates some of these possibilities and describes how they might be used in problems of high-dimensional data fusion.

Summary

We argue that any algorithm that can find sparse, low-dimensional representations of data is an excellent candidate for data fusion and classification. By capturing the key information in a piece of data in only a few coordinates, one can greatly reduce the amount of information that needs to be processed. In effect, these approaches are designed to discard redundant and unnecessary information or clutter, thus improving both the accuracy and speed of classification. A number of important applications can be aided by such an approach. Analysis of multi-spectral data, combinations of ground-based and air-based sensing modalities, or even fusion of time-series and image data are all areas where the above described methods could be valuable.

References

[1] D. L. Bruckstein, D. L. Donoho, and M. Elad, "From Sparse Solutions of Systems of Equations to Sparse Modeling of Signals and Images", *SIAM Review*, **51**(1), 34-81, 2009.

[2] G. K. Rhode, J. M. Nichols, and F. Bucholtz, "Chaotic Signal Detection and Estimation Based on Decompositions Over Attracting Sets: Applications to Secure Communications. *Chaos* **18**, 013114, 2008.

[3] V. Vapnik, "Statistical Learning Theory", Wiley, New York, 1998.

[4] Chih-Chung Chang and Chih-Jen Lin, LIBSVM : a library for support vector machines, 2001. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

[5] S. T. Roweis and L. K. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding", *Science* **290**, 2323-2326.

[6] J. B. Tenenbaum, V. de Silva and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction" *Science* **290** 2319-2323, 2000.

[7] R. R. Coifman and S. Lafon, "Diffusion Maps", *Applied Computational Harmonic Analysis* **21**, 5-30, 2006.

[8] D. L. Donoho and C. Grimes, "Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data", *Proceedings of the National Academy of Sciences USA* **100**(10), 5591-5596, 2003.

[9] P. W. Jones and M. Maggioni and R. Schul, "Manifold Parametrizations by Eigenfunctions of the Laplacian and Heat Kernels" *Proceedings of the National Academy of Sciences USA* **105**(6), 1803-1808, 2008.