SPARSE MANIFOLD LEARNING WITH APPLICATIONS TO SAR IMAGE CLASSIFICATION

V. Berisha [1,2], *N. Shah* [2], *D. Waagen* [2], *H. Schmitt* [2], *S. Bellofiore* [2], *A. Spanias* [1], *D. Cochran* [1]

SenSIP Center, Arizona State University[1] Raytheon Missile Systems[2]

ABSTRACT

Nonlinear data-driven dimensionality reduction techniques have recently gained popularity due to the emergence of high dimensional data sets. The algorithmic complexity and storage requirements of these techniques, however, can make them prohibitive in resource-limited applications. It is therefore beneficial to reduce the number of exemplar samples required for performing an out-of-sample extension to a test point. In this paper, we propose a novel method for selecting a minimal set of exemplars and performing the out-of-sample extension. In the case of two-class target recognition with Synthetic Aperture radar (SAR) data, we compare the efficacy of the proposed approach with other approaches for selecting a subset of the available training samples. We show that the proposed algorithm outperforms the existing methods by providing low-dimensional embeddings that maintain interclass separability using fewer retained exemplars.

Index Terms— classification, reduced complexity Isomap, dimensionality reduction, out-of-sample extension, SAR

1. INTRODUCTION

The emergence of high-dimensional data sets has driven interest in using manifold learning algorithms to extract lower-dimensional representations [1]. In this paper, we are concerned with the storage requirements and computational complexity of dimensionality reduction and out-of-sample extension techniques for target recognition with Synthetic Aperture Radar (SAR) data. The data is taken from the public release portion of the Moving and Stationary Target Acquisition and Recognition (MSTAR) dataset [2], a collection done using a 10 GHz SAR sensor in one-foot resolution spotlight mode at different azimuth and depression angles for a number of different targets. An example to motivate the need for lower-dimensional representations is shown in Fig. 1. The figure shows 50 pixel x 50 pixel SAR images of two targets, a T-72 tank and a BMP-2 infantry fighting vehicle. The classification of individual test images as belonging to the T-72 class or the BMP-2 class requires performing analysis in $\mathbf{R}^{2500}$ space. From the well-known Curse of Dimensionality, it is clear that the density of the data set in this high-dimensional space is sparse and therefore any parameter estimates performed in the high-dimensional space are unreliable. Due to the similarity between images, it is reasonable to assume that, although the data set is represented in $\mathbf{R}^{2500}$, information relevant for the representation and classification of these images is likely to reside in a much lower-dimensional subspace. This motivates the use of dimensionality reduction techniques for learning the underlying informative subspace.

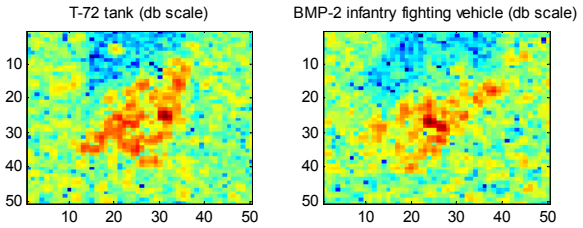

Fig. 1. Sample SAR images of a T-72 tank and a BMP-2 infantry fighting vehicle.

Dimensionality reduction techniques include extraction of features defined by domain experts (e.g., size, shape, contrast, etc.) and data-driven subspace projection/parametrization coordinates. The data-driven techniques can be linear or nonlinear. Linear techniques attempt to maintain consistent Euclidean distances between points in the low-dimensional embeddings, whereas nonlinear techniques attempt to maintain consistent geodesic distances (distances along the manifold) in the low-dimensional embeddings [3] [4] [5]. Linear techniques work well in situations in which the data lies on a linear lower-dimensional subspace, however they grossly overestimate the underlying dimension for data sets lying in nonlinear manifolds.

Although nonlinear dimensionality reduction algorithms (i.e. Isomap [5]) can provide better representations of the underlying manifold geometry, their complexity and storage requirements can make them prohibitive for certain applications. Both complexity and storage increase quadratically with the number of available observations. Although research in the algorithms themselves has been extensive [3] [5], lit-

This work was supported by the Defense Advanced Research Projects Agency under ONR Contract N00014-04-C-0437 and the Arizona State University SenSIP Center. The primary contact is Visar Berisha at visar@asu.edu. Distribution Statement A. Approved for public release; Distribution is unlimited

Table 1. Determining the Number of Training Points

- Compute the vector of the l^{th} largest eigenvalues of $\tilde{\mathbf{M}}$ divided by the number of points, $\frac{\boldsymbol{\lambda}}{N}$
- $\tilde{\mathbf{M}}_0$ = random 10 x 10 symmetric sampling of $\tilde{\mathbf{M}}$
- For each training point not randomly selected during initialization $p = 1 \ldots N - 10$
  - For each unselected training point $n = 1 \ldots N - p - 10$
    - $\tilde{\mathbf{M}}_{p,n}$ = sample $\tilde{\mathbf{M}}$ corresponding to the subset of candidate points and $\tilde{\mathbf{M}}_0$
    - Compute the vector of the l^{th} largest eigenvalues of $\tilde{\mathbf{M}}_{p,n}$ divided by the number of points, $\frac{\boldsymbol{\lambda}_{p,n}}{p+1}$
    - $e_n = \| \frac{\boldsymbol{\lambda}_{p,n}}{p+1} - \frac{\boldsymbol{\lambda}}{N} \|$
  - n^* = minimizer of e_n
  - $\tilde{\mathbf{M}}_p$ = sample $\tilde{\mathbf{M}}$ to also include the n^{*th} point
  - $e_p = \| \frac{\boldsymbol{\lambda}_{p,n}}{p+1} - \frac{\boldsymbol{\lambda}}{N} \|$
  - if $e_p < tol$ then end

tle has been done to address the issue of reducing the number of observations stored for out-of-sample extensions [9]. Since the training of these algorithms is performed offline, the complexity constraints during training are relaxed. However the low-dimensional embedding of test points must be performed in real-time. This places restrictions on the complexity and the storage requirements of the algorithm that performs the embedding. Once the test points are projected into the low-dimensional subspace, classification techniques can be applied.

In this paper, we propose an algorithm for determining an appropriate number of points required to represent a low-dimensional manifold lying in a high-dimensional space, selecting these points, and performing the out-of-sample extension using only the selected points. We show that the proposed technique results in smaller storage requirements without compromising performance. Although the results are given in terms of a SAR image classification example, the algorithm is broadly applicable. In fact, the findings can be generalized to any classification problem.

This paper is organized as follows. Section 2 provides a description of the proposed algorithm. In Section 3 we analyze and compare the storage requirements of the proposed algorithm. Section 4 contains sample results using actual SAR image data and section 5 provides concluding remarks.

2. PROPOSED METHOD

In this section we describe the two main components of the proposed algorithm: determining the number of points required to describe the underlying manifold and performing the out-of-sample (OOS) extension using only the selected points.

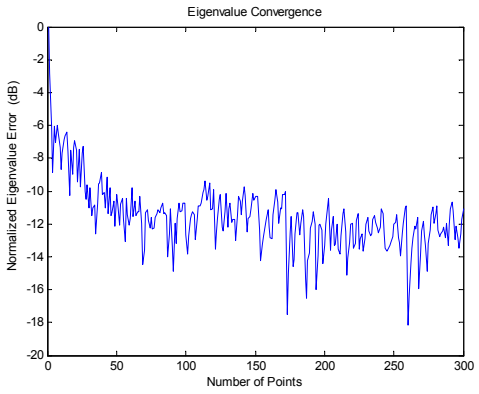

Fig. 2. The convergence of eigenvalues as more training points are selected.

Consider a data set $\{\mathbf{x_1}, ..., \mathbf{x_N}\}$ with N points residing in some m-dimensional space. For $l << m$, the l-dimensional embedding of $\mathbf{x_i}$ is the vector $\mathbf{y_i}$. An NxN neighborhood matrix $\mathbf{M}$ is formed using some two-argument (kernel) function $K(a, b)$ with $M_{ij} = K(\mathbf{x_i}, \mathbf{x_j})$. For Isomap, the neighborhood matrix is the geodesic distance between all points $\mathbf{x_i}$ and $\mathbf{x_j}$. The matrix is then normalized as described in [6] to form $\widetilde{\mathbf{M}}$. Denote the largest positive eigenvalues of $\widetilde{\mathbf{M}}$ and their corresponding eigenvectors by λ_j and $\mathbf{v_j}$.

2.1. Determining the Number of Points

The lower-dimensional embedding, $\{\mathbf{y_1}...\mathbf{y_N}\}$, for a high-dimensional data set, $\{\mathbf{x_1}...\mathbf{x_N}\}$, will eventually converge as more points are added to the set. More specifically, as the data set grows, the eigenvectors, $\mathbf{v_j}$, and the scaled largest positive eigenvalues, $\frac{\lambda_j'}{N}$, converge [7].

The idea behind the proposed technique is to keep the minimum number of samples that provide efficient convergence. The idea is based on a greedy optimization scheme in which we add the points that provide the quickest eigenvalue convergence. Start with an empty geodesic distance matrix and add individual rows and columns corresponding to a particular point. Solve for the eigenvalues of the resulting matrix. These eigenvalues are then compared to the eigenvalues of the original matrix, $\widetilde{\mathbf{M}}$. The training point that provides the quickest convergence at that iteration is added to the set. The algorithm is outlined in Table 1.

We apply the algorithm to the 2500-dimensional SAR images. Figure 2 shows the convergence of eigenvalues for each additional training point in terms of the normalized residual error between the estimated eigenvalue sequence and the eigenvalue sequence determined from the full data set. The appropriate number of training points is determined by setting a tolerance for the convergence, determined for example by seeking a knee in the curve of residual error. In this example, roughly 25% (300 points) of the total data set was

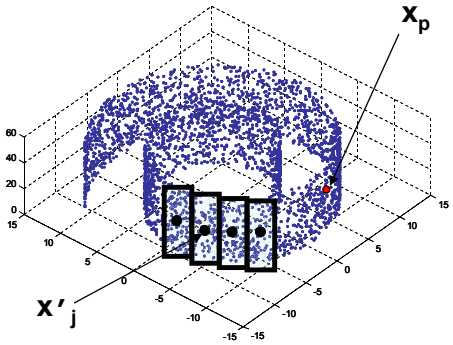

Fig. 3. An example to demonstrate the principles of the proposed OOS extension algorithm.

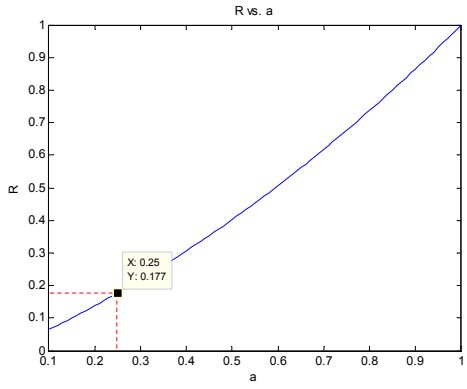

Fig. 4. Storage requirements ratio (R) as a function of the fraction of training data used (a).

selected.

2.2. Out-of-Sample Extension

Assume that we have available the parameters from an Isomap algorithm (the principal eigenvectors and eigenvalues of the geodesic distance matrix) pre-trained with N data points. The $k-th$ component of the low-dimensional embedding, $\mathbf{y_p}$, corresponding to a newly acquired high-dimensional point, $\mathbf{x_p}$, is determined using the Nyström expansion below [6] [10].

$$y_p(k) = \frac{1}{\sqrt{\lambda_k}} \sum_{i=1}^{N} v_{ik} dist(\mathbf{x_p}, \mathbf{x_i}) \tag{1}$$

where $dist(\mathbf{x_p}, x_i)$ denotes the geodesic distance between the out-of-sample point p and every other point in the training set. Equation (1) implies that to embed each additional point we must store the entire training data set, all eigenvalues, all eigenvectors, and the geodesic distance matrix. For large data sets this quickly becomes impractical.

Consider the Swiss roll example shown in Fig. 3. Out of the set of N $(\mathbf{x_1}, ..., \mathbf{x_N})$ total points, we select a subset of N' $(\mathbf{x_1'}, ..., \mathbf{x_{N'}'})$ points using the algorithm described in [8]. These are denoted by the dark points in the figure and the boxed area around each point represents the nearest neighbors of these points. The total space is therefore divided into N' neighborhoods, whose centers (in the nearest-neighbor sense, not a volumetric average) are the selected points, $\mathbf{x_j'}$. We approximate the geodesic distance between the newly acquired point $\mathbf{x_p}$ and all points in neighborhood j by the distance between $\mathbf{x_p}$ and $\mathbf{x_j'}$. This allows us to combine all the eigenvectors associated with each neighborhood in the Nyström expansion. We can formulate the expansion as follows:

$$y_p(k) \approx \frac{1}{\sqrt{\lambda_k}} \sum_{j=1}^{N'} \sum_{i=1}^{\frac{N}{N'}} \tilde{v}_{ijk} dist(\mathbf{x_p}, \mathbf{x_j'}) \tag{2}$$

where $\tilde{v}_{ijk}$ represents the $k-th$ component of the $i-th$ eigenvector associated with the $j-th$ neighborhood. Analyzing equation (2) shows that we can store the sum of the $\frac{N'}{N}$ eigenvectors in each neighborhood rather than each individual eigenvector:

$$\tilde{v}_{jk} = \sum_{i=1}^{\frac{N}{N'}} \tilde{v}_{ijk}. \tag{3}$$

This formulation allows us to approximate the lower-dimensional embedding as follows:

$$y_p(k) \approx \frac{1}{\sqrt{\lambda_k}} \sum_{j=1}^{N'} \tilde{v}_{jk} dist(\mathbf{x_p}, \mathbf{x_j'}). \tag{4}$$

It is clear from the equation above that the storage requirements are significantly less due to the fact that the number of stored points has been reduced from N to N'.

3. STORAGE REQUIREMENTS AND COMPLEXITY

As stated earlier, the training data set, the geodesic distance matrix, and the eigenvectors are the most burdensome in terms of storage. The proposed algorithm reduces each parameter set significantly due to the fact that only a subset of N' points out of the total are required to perform the out-of-sample extension. It is easy to see show that the ratio between the storage requirements of the proposed algorithm and the original storage requirements can be expressed as follows:

$$R = \frac{N'm + N'^2 + N'l + l}{Nm + N^2 + Nl + l} \tag{5}$$

We can approximate the ratio in (5) as a function of the fraction of points used, $a = 0 \ldots 1$. In Fig. 4 we plot R as a function of a and label the point corresponding to $a = 0.25$. The plot shows that by storing 25% of the points we reduce the storage requirements of the parameter set to $\approx 17\%$ of the original amount.

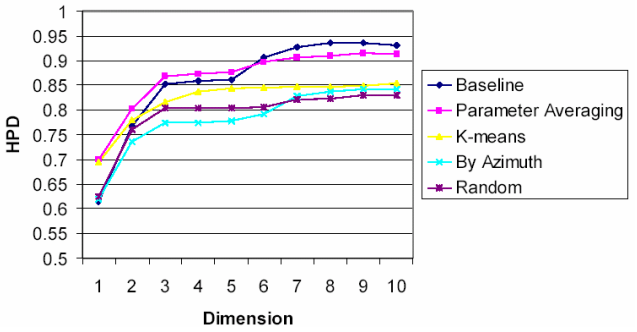

Fig. 5. Separability of two classes as a function of subspace dimensionality for four different training set selection techniques. The baseline is generated by performing the out-of-sample extension with the complete data set.

4. RESULTS

The results presented in this section are in terms of interclass divergence, which we use as a surrogate for classification performance. We use the public release portion of the MSTAR database of SAR images. The SAR images are first pre-processed using a constant false alarm rate (CFAR) algorithm that centers a 50 pixel x 50 pixel frame around the target. We test the efficacy of the proposed algorithm for a two class scenario, namely the T-72 tank and the BMP-2 infantry fighting vehicle. The data set is divided into two groups, the training set (depression angle 17^o) and the testing set (depression angle 15^o). Isomap is used to perform the manifold learning using only a subset of the training data and the out-of-sample extension is performed on the whole test set using the proposed algorithm.

As an alternative to evaluating specific classifier performance, we use the Henze-Penrose Divergence (HPD) for measuring relative interclass separability [11]. The underlying assumption is that independent of any particular classifier, feature sets that exhibit more divergence (or separability) should in general be of greater utility than feature sets that exhibit less divergence (or separability) [12]. For two data sets with an equal number of samples, HPD values range from 0.5 to 1, with 0.5 implying the classes cannot be separated and 1 implying that the two classes are completely separable.

Figure 5 shows the HPD of the two data sets at different subspace dimensionalities, where, for example, dimensionality = 8 means that the first 8 subspace coordinates are retained as features. We compare four different set selection techniques (with each limited to 25% of the original data) to the baseline generated using the complete data set. As the plot shows, our proposed technique (parameter averaging) outperforms the other methods at each of the tested subspace dimensionalities and approaches the baseline for higher subspace dimensionality. The other subset selection techniques include selecting a subset at random, selecting a subset using the k-means clustering algorithm, and selecting a subset using clustering based on $a\ priori$ information about the target azimuth angle. It is important to note that the storage and complexity requirements for out-of-sample extensions are identical for all techniques.

5. CONCLUSION

In this paper, we proposed an algorithm for selecting an appropriate subset of training points for manifold learning and lower-dimensional embedding in the context of classification. The purpose of the algorithm is to reduce the storage requirements for real-time applications without compromising the separability of classes in lower dimensions. An explicit algorithm is presented that determines a sufficient set of points for performing the dimensionality reduction. In addition, an out-of-sample algorithm making use of only the selected points is proposed. Preliminary results (with actual SAR data) suggest that a significant reduction in the training set is possible without a reduction in classification performance, as measured by interclass divergence.

6. REFERENCES

[1] A.K. Jain, R. Duin, and J. Mao, "Statistical Pattern Recognition: A Review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 4-37, January 2000.

[2] E.R. Keydel, "MSTAR Extended Operating Conditions," *SPIE Proceedings*, vol. 2757, pp. 228-242, 1996.

[3] B. Schlkopf, A. Smola, and K. Mller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, pp. 1299-1319, 1998.

[4] J. Ham, D. Lee, S. Mika, and B. Schlkopf, "A kernel view of the dimensionality reduction of manifolds," *Max-Planck Institute for Biological Cybernetics Technical Reports*, No. 110, 2003.

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

[6] Y. Bengio et al., "Out of sample extensions for LLE, ISOMAP, MDS, Eigenmaps, and Spectral Clustering," *Neural Information Processing Systems Conference*, 2003.

[7] J. Shawe-Taylor and C. Williams, "The stability of kernel principal components analysis and its relation to the process eigenspectrum," in S. Becker, S. Thrun, and K. Obermayer, editors, *Advances in Neural Information Processing Systems* 15. MIT Press., 2003.

[8] J.G. Silva, J.S. Marques, and J.M. Lemos, "Selecting Landmark Points for Sparse Manifold Learning," *Neural Information Processing Systems Conference*, 2005.

[9] V. de Silva and J.B. Tenenbaum, "Sparse multidimensional scaling using landmark points," preprint, 2004.

[10] J.C. Platt, "FastMap, MetricMap, and Landmark MDS are all Nyström Algorithms," *Microsoft Research Technical Reports*, No. 2004-26, September 2004.

[11] N. Henze and M.D. Penrose, "On the Multivariate Runs Test," *The Annals of Statistics*, vol. 27, pp. 290-298, 1999.

[12] N. Shah, D. Waagen, B. Corner, and H. Schmitt, "Correlation of Interclass Divergence and Classification Performance," *Combat Identification Systems Conference*, June 19-22, Orlando, FL, 2006.