SAR Automatic Target Recognition via Non-negative Matrix Approximations

Vahid Riasati[a], Umamahesh Srinivas[b] and Vishal Monga[b]

[a]MacAulay-Brown Inc., Dayton OH 45430
[b]Dept. of Electrical Engg., Pennsylvania State University, University Park, PA 16802

ABSTRACT

The set of orthogonal eigen-vectors built via principal component analysis (PCA), while very effective for compression, can often lead to loss of crucial discriminative information in signals. In this work, we build a new basis set using synthetic aperture radar (SAR) target images via non-negative matrix approximations (NNMAs). Owing to the underlying physics, we expect a non-negative basis and an accompanying non-negative coefficient set to be a more accurate generative model for SAR profiles than the PCA basis which lacks direct physical interpretation. The NNMA basis vectors while not orthogonal capture discriminative local components of SAR target images. We test the merits of the NNMA basis representation for the problem of automatic target recognition using SAR images with a support vector machine (SVM) classifier. Experiments on the benchmark MSTAR database reveal the merits of basis selection techniques that can model imaging physics more closely and can capture inter-class variability, in addition to identifying a trade-off between classification performance and availability of training.

Keywords: Automatic target recognition, non-negative matrix approximations, principal component analysis, synthetic aperture radar

1. INTRODUCTION

The problem of identifying target images from sensed imagery, popularly known as automatic target recognition (ATR), has attracted the attention of the research community for over three decades now.[1] The target images can typically be acquired from a rich variety of sensing sources like synthetic aperture radar (SAR), inverse SAR, hyperspectral imaging (HSI), and forward-looking infra-red (FLIR).

In a supervised learning approach to ATR, a small set of labeled representative features extracted from each target class is first used to train a classifier (or decision engine) to minimize classification error. It is well known that suitably chosen features with reduced dimensionality can often encapsulate most of the information in an image, and this is particularly useful for efficient computations in many real-time applications. Accordingly, the process of feature extraction may be modeled as a projection to a lower dimensional *feature space* which is represented by a collection of basis vectors. Mathematically, the feature $\boldsymbol{x} \in \mathbb{R}^n$ corresponding to a target image $\boldsymbol{y} \in \mathbb{R}^m, (n < m)$, is obtained by solving the following optimization problem:

$$\boldsymbol{x} = \arg \min_{\hat{\boldsymbol{x}}} \ \|\boldsymbol{y} - \boldsymbol{A}\hat{\boldsymbol{x}}\|_2, \tag{1}$$

where the projection matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is a collection of n basis vectors, each in $\mathbb{R}^m$.

In literature, principal component analysis (PCA) is perhaps one of the most widely used feature extraction techniques for image classification tasks. In ATR, PCA-based approaches have led to the use of eigen-templates[2] as features. Other popular feature extraction techniques in ATR include geometric descriptors such as robust edges and corners,[3] projection to and selective retention of transform domain coefficients, particularly wavelets,[4] and estimation-theoretic templates.[5] In parallel, a variety of decision engines have been proposed and some of the more recent choices have been inspired by advances in machine learning theory, like neural networks,[6] support vector machines (SVMs)[7] and boosting.[8] Of active related research interest is the problem of fusing information for robust classification, leading to the development of ensemble classifiers,[9] meta-classifiers[10] and principled strategies for feature fusion.[11]

Although the eventual class decision is made by the decision engine, the discriminative capability of the features can significantly influence the success of classification. In this paper, we compare the effect of different basis representations, i.e. the projection matrices $\boldsymbol{A}$ in (1), on ATR classification performance. In particular, we use training SAR images that span the various target classes and build a new basis set via a recent linear algebra advance known as non-negative matrix approximations (NNMAs), which involves an explicit non-negativity constraint. The fundamental insight is that owing to underlying physics, we expect non-negative basis and an accompanying non-negative coefficient set to be a more accurate generative model for SAR target images than the PCA basis which lacks direct physical interpretation. The NNMA basis vectors while not orthogonal capture discriminative local components of SAR images, thus offering much greater discriminative capability. Features for classification are extracted by projecting the SAR target images onto the NNMA basis sub-space. The feature vectors are then fed into a support vector machine classifier which performs the eventual class assignment. We evaluate the different basis representation schemes on the benchmark MSTAR database.[12] Experiments reveal the benefits of using basis representations that model the imaging physics better, when compared to the PCA-based approach[2] as a baseline.

2. BACKGROUND

In this section, we briefly review the key aspects of principal component analysis and non-negative matrix approximations.

2.1 Principal Component Analysis

Principal component analysis[13] is primarily used as a tool to build a reduced dimension basis from a set of statistically representative training data samples. A simplistic yet often powerful assumption to understand any data set that captures a physical phenomenon is that there exists a collection of basis vectors whose linear combinations can model any observation of the particular phenomenon. Often, the high dimensionality of data can mask the underlying, possibly simpler models that can explain the phenomena represented by the data. The goal of PCA is to identify a new basis representation, in terms of a *linear* combination of the original basis vectors, which leads to more meaningful interpretation of the data.

The PCA technique is closely related to the more general technique of singular value decomposition (SVD). Consider a collection of T vectors, each in $\mathbb{R}^m$. Let $\boldsymbol{X}$ be the $m \times T$ matrix consisting of the T vectors stacked as columns. The singular value decomposition of $\boldsymbol{X}$ gives:

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^t, \tag{2}$$

where $\boldsymbol{U} = [\boldsymbol{u}_1 \ \boldsymbol{u}_2 \ \cdots \ \boldsymbol{u}_m] \in \mathbb{R}^{m \times m}$ is the matrix of eigenvectors of $\boldsymbol{X}\boldsymbol{X}^t$, $\boldsymbol{V} = [\boldsymbol{v}_1 \ \boldsymbol{v}_2 \ \cdots \ \boldsymbol{v}_T] \in \mathbb{R}^{T \times T}$ is the matrix of eigenvectors of $\boldsymbol{X}^t\boldsymbol{X}$ and the diagonal entries of $\boldsymbol{\Lambda} \in \mathbb{R}^{m \times T}$ are the singular values arranged in descending order. The r most relevant singular vectors can be chosen to reconstruct an r-rank approximation of the data matrix. An important property of SVD is that, in fact, this approximation is the optimal r-rank approximation to $\boldsymbol{X}$ in the sense of minimum Frobenius norm reconstruction error. This property naturally leads to a method of achieving dimensionality reduction by identifying the $r < n$ most significant singular vectors (or, equivalently principal components) onto which any new vector may be projected. These principal components (i.e. the r singular vectors) are necessarily orthogonal by design.

Eigen-templates[2] based on the SVD technique have been employed for feature extraction in ATR with considerable success. The first r columns of $\boldsymbol{U}$ are chosen as the projection matrix $\boldsymbol{A}$, and feature vectors are obtained by projecting the target images onto the r-dimensional space spanned by $\boldsymbol{A}$. This set of orthogonal eigen-vectors while very effective for compression can often lead to loss of discriminative information in signals which could be useful for distinguishing between signal classes.

2.2 Non-negative Matrix Approximations

Non-negative matrix factorization (NMF) is a dimensionality reduction technique proposed by Lee et al.[14] Unlike SVD, NMF performs matrix factorizations (more accurately, approximations) under an element-wise nonnegativity constraint on the factor matrices. The success of NMFs stems from the fact that for many physical stimuli,

including image pixel intensities/colors, the underlying generative model is indeed a linear combination of basis functions with element-wise non-negative components. Orthogonality as in SVD is therefore often unnatural - unless compression/accurate reconstruction from a small number of coefficients is the ultimate goal. In ATR, the goal is to achieve high classification rates, which is very distinct from compression/reconstruction.

The goal of NMF is to factorize the data matrix $\boldsymbol{X}$ into two *non-negative* matrices:

$$\boldsymbol{X} \approx \boldsymbol{WH}, \tag{3}$$

where $\boldsymbol{W} \in \mathbb{R}^{m \times r}$ and $\boldsymbol{H} \in \mathbb{R}^{r \times T}$. Each column $\boldsymbol{x}$ of $\boldsymbol{X}$ can be written as $\boldsymbol{x} \approx \boldsymbol{Wh}$, where $\boldsymbol{h}$ is the corresponding column of $\boldsymbol{H}$. As a result, we obtain an r-dimensional feature representation of each target image (r is typically chosen to be smaller than m and T). In contrast to the SVD technique discussed in Section 2.1, the projection matrix $\boldsymbol{A}$ is now chosen to be the matrix $\boldsymbol{W}$, which can be interpreted as a (dimensionality-reducing) basis for feature extraction.

For a collection of training target vectors $\boldsymbol{X}$, $\boldsymbol{W}$ and $\boldsymbol{H}$ can be obtained by solving the following optimization problem:

$$\text{minimize } \|\boldsymbol{X} - \boldsymbol{WH}\|^2 \text{ w.r.t. } \boldsymbol{W} \text{ and } \boldsymbol{H}, \text{ subject to } \boldsymbol{W}, \boldsymbol{H} \geq 0, \tag{4}$$

which can be solved via the Euclidean update rule. An alternate formulation of the NMF problem is considered in this paper:

$$\text{minimize } D(\boldsymbol{X}\|\boldsymbol{WH}) \text{ w.r.t. } \boldsymbol{W} \text{ and } \boldsymbol{H}, \text{ subject to } \boldsymbol{W}, \boldsymbol{H} \geq 0, \tag{5}$$

where the *divergence* term is defined as:

$$D(\boldsymbol{P}\|\boldsymbol{Q}) = \sum_{i,j} \left(\boldsymbol{P}_{ij} \log \frac{\boldsymbol{P}_{ij}}{\boldsymbol{Q}_{ij}} - \boldsymbol{P}_{ij} + \boldsymbol{Q}_{ij} \right), \tag{6}$$

and subscripts represent the element indices. Corresponding to this formulation, we use an iterative divergence update rule to solve the problem. Once $\boldsymbol{W}$ is obtained using the training target vectors, the feature vector corresponding to any new target vector $\boldsymbol{y}$ is obtained by solving the following optimization problem:

$$\boldsymbol{h} = \min_{h} \|\boldsymbol{y} - \boldsymbol{Wh}\|_2, \text{ subject to } \boldsymbol{h} \geq 0, \tag{7}$$

which is the well-known non-negative least squares problem.

3. PROPOSED CLASSIFICATION FRAMEWORK

The overall classification framework is illustrated in Fig. 1. Target classification is a two-stage process. First, in the *training* stage, lexicographically-ordered versions of training images from all classes are collected in a data matrix $\boldsymbol{X} \in \mathbb{R}^{m \times T}$. Features are generated from this set of labeled training vectors and are used to train the SVM-based classifier. The aspect of feature extraction is tied to the choice of projection matrix. For our experiments, we obtain the projection matrix $\boldsymbol{A}$ in two different ways:

1. Using the r-rank orthogonal projection $\boldsymbol{U}$ from the SVD procedure (see Section 2.1),

2. Using the non-negative basis matrix $\boldsymbol{W}$ obtained from NMFs (see Section 2.2).

For each choice of projection, features are extracted from target vectors by solving (1).

Next, in the *test* stage, features from new test images are assigned target class labels by the SVM. So far, we have considered the feature extraction process in detail. We now briefly introduce the SVM classifier.
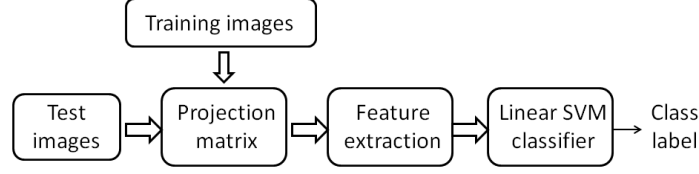
Figure 1. Schematic of the overall classification framework.

3.1 Support Vector Machines

SVM[15] is a popular machine learning technique to classify data, which determines the margin maximizing hyperplane in a feature space. The decision function of a linear SVM classifier (for a two-class problem) may be written as:

$$f(\boldsymbol{x}) = \sum_{i=1}^{N} \alpha_i y_i (\boldsymbol{s}_i^T \boldsymbol{x}) + b, \tag{8}$$

where $\boldsymbol{x} \in \mathbb{R}^n$ is the feature vector to be classified, the $\boldsymbol{s}_i$ are support vectors, N is the number of support vectors, and $y_i \in \{-1, 1\}$ are the binary classifier decisions. Parameters α_i and b are determined by solving the ensuing constrained optimization problem for margin maximization. A binary classification decision for $\boldsymbol{x}$ is made depending on whether $f(\boldsymbol{x}) > 0$ or otherwise. The linear SVM assumes that the feature data is linearly separable, i.e. there exists an optimal hyperplane that separates the two classes. This technique can be extended to learn more elaborate non-linear separators between classes via the kernel trick, by considering kernels which map features into a higher-dimensional space where they are more likely to be linearly separable. The corresponding decision function for a kernel-based classifier is:

$$f(\boldsymbol{x}) = \sum_{i=1}^{N} \alpha_i y_i K(\boldsymbol{s}_i, \boldsymbol{x}) + b, \tag{9}$$

where $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \langle \phi(\boldsymbol{x}_i), \phi(\boldsymbol{x}_j) \rangle$ and ϕ represents a nonlinear mapping of target data. For multi-class classifiers, the one-versus-all approach is employed to decide the target category. The use of SVMs in ATR problems has proliferated and it has been widely accepted to be representative of state-of-the art classifiers.

3.2 Feature extraction: NNMA vs. PCA

The projection $\boldsymbol{A}$ captures some information about the overall distribution of target images from all classes. In PCA, the goal is to capture this information by identifying the principal components from the eigenvectors of the sample covariance matrix. In NNMA, the idea is to factorize the data matrix into basis and coefficient matrices, both of which have non-negative elements. In both these cases, it is also well known that low-rank approximations can represent the original data reasonably well, leading to dimensionality reduction that can in turn reduce computational complexity. The key assumption in this process of projection to a suitable basis is that a handful of principal components or the low-rank NNMAs are sufficient to capture most of the variations between the different target classes. The first few principal components will be useful for classification when inter-class variations are more dominant than intra-class variations.

Our choice of NNMAs for feature extraction is motivated by the physics underlying the generative model for SAR target image capture. The intensity value represented by each SAR image pixel is non-negative. However, there is no empirical justification for the choice of an orthogonal set of basis vectors to represent the feature space. This argument motivates the claim that a non-negative projection matrix (along with the corresponding nonnegative feature vectors) is a more reasonable representation of the feature space compared to a representation based on orthogonal PCA (or equivalently, SVD) feature vectors. The NNMA basis vectors while not orthogonal capture discriminative local components of SAR images, thus offering much greater discriminative capability. The results in the next section validate our claim.

Table 1. Confusion matrix for shared basis using PCA.

Class	BMP-2	BTR-70	T-72	ZIL131	D7
BMP-2	**0.84**	0.06	0.04	0.02	0.04
BTR-70	0.05	**0.87**	0.03	0.02	0.03
T-72	0.03	0.07	**0.83**	0.03	0.04
ZIL131	0.05	0.03	0.02	**0.84**	0.06
D7	0.06	0.02	0.04	0.06	**0.82**

Table 2. Confusion matrix for shared basis using NNMA.

Class	BMP-2	BTR-70	T-72	ZIL131	D7
BMP-2	**0.86**	0.05	0.02	0.05	0.02
BTR-70	0.07	**0.88**	0.04	0.01	0.0
T-72	0.03	0.04	**0.86**	0.02	0.05
ZIL131	0.01	0.06	0.05	**0.87**	0.01
D7	0.04	0.02	0.06	0.04	**0.84**

4. EXPERIMENTS AND RESULTS

Experiments are performed on a set of benchmark SAR images obtained from the Moving and Stationary Target Acquisition and Recognition (MSTAR) program,[12] which has released a large set of SAR images in the public domain. Five target classes from the database are used: T-72 tanks, BMP-2 infantry fighting vehicles, BTR-70 armored personnel carriers, ZIL trucks and D7 tractors. The T-72 and BMP-2 classes have three different serial number variants each, while the other three classes have a single serial number variant each. For training, 150 images are selected from each class, while a total of 1913 images are used for testing. Training and test images are acquired with 17° and 15° depression angle respectively. Minimal pre-processing of data in the form of speckle denoising and target centering is performed prior to classification, similar to other algorithms in literature.[16] The size of the target chip - which is the processed image input to the basis selection scheme - is normalized to 64×64, i.e. the vectorized target image lies in $\mathbb{R}^{4096}$.

The classification performance is represented using confusion matrices. Each row of the confusion matrix indicates the true identity of target images, while each column is the predicted target class label as a result of our proposed scheme. Tables 1-2 show the confusion matrices. We select the 750 largest singular vectors to build $\boldsymbol{A}$ in case of PCA, and choose a rank-r approximation using NNMA where $r = 750$. We observe that using NNMAs improves classification performance over the baseline PCA, pointing to the benefits of selecting the basis based on characteristics of the image generative model.

5. CONCLUSION

Orthogonal feature space representations based on principal component analysis have been used for feature extraction with considerable success in SAR automatic target recognition. In this paper, we propose a novel feature space representation based on non-negative matrix approximations, which is intuitively motivated by the underlying imaging physics. In particular, we use the training SAR target data to obtain a projection matrix and feature vectors with strictly non-negative elements, by solving a constrained optimization problem. This choice of non-negative projection matrix simultaneously achieves dimensionality reduction as well as captures the key inter-class variations, leading to better classification performance than PCA-based classifiers. As part of future work, we will investigate the design of basis matrices that can capture inter-class variability better while offering better robustness to choice of training.

REFERENCES

[1] Bhanu, B. and Jones, T. L., "Image understanding research for automatic target recognition," *IEEE Aerosp. Electron. Syst. Mag* **8**, 15–23 (Oct. 1993).

[2] Bhatnagar, V., Shaw, A., and Williams, R. W., "Improved automatic target recognition using singular value decomposition," in [*IEEE Int. Conf. Acoust., Speech, Signal Processing*], 2717–2720 (1998).

[3] Olson, C. F. and Huttenlocher, D. P., "Automatic target recognition by matching oriented edge pixels," *IEEE Trans. Image Processing* **6**, 103–113 (Jan. 1997).

[4] Casasent, D. and Smokelin, Y., "Wavelet and Gabor transforms for target detection," *Optical Engineering* **31**(9), 1893–1898 (1992).

[5] Grenander, U., Miller, M. I., and Srivastava, A., "Hilbert-Schimdt Lower Bounds for Esimators on Matrix Lie Groups for ATR," *IEEE Trans. Pattern Anal. Machine Intell.* **20**, 790–802 (Aug. 1998).

[6] Daniell, C. E., Kemsley, D. H., Lincoln, W. P., Tackett, W. A., and Baraghimian, G. A., "Artificial neural networks for automatic target recognition," *Optical Engineering* **31**(12), 2521–2531 (1992).

[7] Zhao, Q. and Principe, J., "Support vector machines for SAR automatic target recognition," *IEEE Trans. Aerosp. Electron. Syst.* **37**, 643–654 (Apr. 2001).

[8] Sun, Y., Liu, Z., Todorovic, S., and Li, J., "Adaptive boosting for SAR automatic target recognition," *IEEE Trans. Aerosp. Electron. Syst.* **43**, 112–125 (Jan 2007).

[9] Rizvi, S. A. and Nasrabadi, N. M., "Fusion techniques for automatic target recognition," in [*Applied Imagery Pattn. Recogn. Workshop*], 27–32 (2003).

[10] Srinivas, U., Monga, V., and Raj, R. G., "Meta-classifiers for exploiting feature dependencies in automatic target recognition," in [*IEEE Intl. Radar Conf.*], 147–151 (2011).

[11] Srinivas, U., Monga, V., and Raj, R. G., "Automatic target recognition using discriminative graphical models," in [*Proc. IEEE Intl. Conf. Image Processing*], 33–36 (2011).

[12] "*The Airforce Moving and Stationary Target Recognition Database.*"

[13] Jolliffe, I. T., [*Principal Component Analysis*], Springer, New York (1986).

[14] Lee, D. D. and Seung, H. S., "Algorithms for non-negative matrix factorization," *Advances Neural Inform. Processing Syst.* (2001).

[15] Vapnik, V. N., [*The nature of statistical learning theory*], New York, USA: Springer (1995).

[16] Ross, T., Worrell, S., Velten, V., Mossing, J., and Bryant, M., "Standard SAR ATR evaluation experiments using the MSTAR public release data set," *SPIE Conf. on Algorithms for Synthetic Aperture Radar Imagery V* **3370**, 566–573 (Apr. 1998).