SAR Automatic Target Recognition via Non-negative Matrix Approximations

Vahid Riasati† Umamahesh Srinivas‡ Vishal Monga‡

†MacAulay-Brown Inc.
Dayton, OH

‡Pennsylvania State University
University Park, PA

2012 SPIE Defense, Security + Sensing: Advances in Algorithms for ATR I

April 24, 2012

Automatic Target Recognition (ATR)

- Exploit imagery from diverse sensed sources for automatic target identification[1]
- Variety of sensors: synthetic aperture radar (SAR), inverse SAR (ISAR), forward looking infra-red (FLIR), hyperspectral
- Diverse scenarios: air-to-ground, air-to-air, surface-to-surface

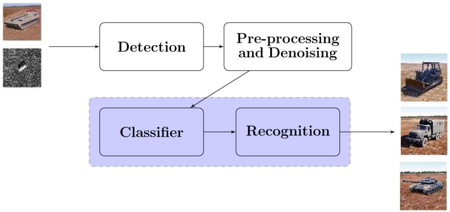


Figure: Schematic of ATR framework. The classification and recognition stages assign an input image/ feature to one of many target classes.

[1] Bhanu et al., IEEE AES Systems Magazine, 1993

Target classification

Two-stage framework:

1. Feature extraction from sensed imagery
    - Geometric feature-point descriptors[2]
    - Eigen-templates[3]
    - Transform domain coefficients - wavelets[4]

---

[2] Olson et al., IEEE Trans. Image Process., 1997

[3] Bhatnagar et al., IEEE ICASSP, 1998

[4] Casasent et al., Neural Networks, 2005

[5] Daniell et al., Optical Engineering, 1992

[6] Zhao et al., IEEE Trans. Aerosp. Electron. Syst., 2001

Target classification

Two-stage framework:

1. Feature extraction from sensed imagery
   - Geometric feature-point descriptors[2]
   - Eigen-templates[3]
   - Transform domain coefficients - wavelets[4]
2. Decision engine which performs class assignment
   - Linear and quadratic discriminant analysis
   - Neural networks[5]
   - Support vector machines (SVM)[6]

---

2 Olson et al., IEEE Trans. Image Process., 1997

3 Bhatnagar et al., IEEE ICASSP, 1998

4 Casasent et al., Neural Networks, 2005

5 Daniell et al., Optical Engineering, 1992

6 Zhao et al., IEEE Trans. Aerosp. Electron. Syst., 2001

Recent research trends: Classifier fusion

- Search for 'best possible' features from a classification standpoint
- Exploit complementary yet correlated information offered by different sets of features/classifiers
  - Product of individual classification probabilities[7]
  - Voting strategy[8]
  - Boosting[9]
  - Meta-classification[10]

---

[7] Paul et al., IEEE ICASSP, 2003

[8] Gomes et al., IEEE Radar Conf., 2008

[9] Sun et al., IEEE Trans. Aerosp. Electron. Syst., 2007

[10] Srinivas et al., IEEE Radar Conf., 2011

Motivation: Feature extraction

- Feature extraction $\rightarrow$ projection to lower dimensional feature space
  1. Inherent low-dimensional space that captures image information with minimal redundancy[11]
  2. Computational benefits for real-time applications

[11] Jolliffe, Principal Component Analysis, Springer, 1986

Motivation: Feature extraction

- Feature extraction $\rightarrow$ projection to lower dimensional feature space
  1. Inherent low-dimensional space that captures image information with minimal redundancy[11]
  2. Computational benefits for real-time applications
- Optimization problem:

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

  - $\boldsymbol{y}$: target image in $\mathbb{R}^m$
  - $\boldsymbol{x}$: corresponding feature vector in $\mathbb{R}^n, n < m$
  - $\boldsymbol{A}$: projection matrix in $\mathbb{R}^{m \times n}$ - collection of n basis vectors, each in $\mathbb{R}^m$

[11] Jolliffe, Principal Component Analysis, Springer, 1986

Motivation: Feature extraction

- Feature extraction → projection to lower dimensional feature space
  1. Inherent low-dimensional space that captures image information with minimal redundancy[11]
  2. Computational benefits for real-time applications
- Optimization problem:

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

  - $\boldsymbol{y}$: target image in $\mathbb{R}^m$
  - $\boldsymbol{x}$: corresponding feature vector in $\mathbb{R}^n, n < m$
  - $\boldsymbol{A}$: projection matrix in $\mathbb{R}^{m \times n}$ - collection of n basis vectors, each in $\mathbb{R}^m$
- How to choose $\boldsymbol{A}$?

[11] Jolliffe, Principal Component Analysis, Springer, 1986

Contribution of our work

- Non-negative matrix approximation (NNMA) for feature extraction
- Performance comparison with traditional principal component analysis-based feature extraction

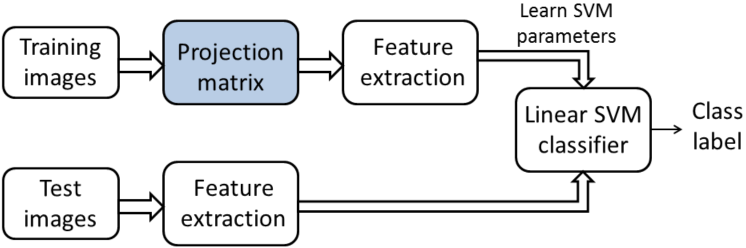
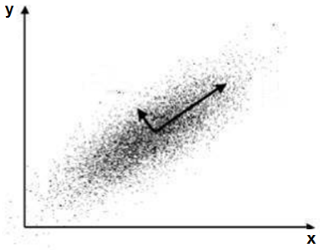


Figure: Proposed target classification framework.

Principal Component Analysis (PCA)

- Statistical tool for dimensionality reduction via change of basis
- Modeling an observation of physical phenomena as a linear combination of basis vectors



- Eigenvectors of data covariance matrix form the projection basis
- Applications in image classification: eigenfaces for face recognition[12], eigen-templates for ATR[13]

[12] Turk and Pentland, IEEE Conf. CVPR, 1991

[13] Bhatnagar et al., IEEE ICASSP, 1998

Singular Value Decomposition (SVD)

- Generalization of PCA
- Data matrix $\boldsymbol{X} \in \mathbb{R}^{m \times N}$ can be factorized as:

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^T = \sum_{i=1}^{r} \lambda_i \boldsymbol{u}_i \boldsymbol{v}_i^T$$

- $\boldsymbol{U} = [\boldsymbol{u}_1 \ \boldsymbol{u}_2 \ \cdots \ \boldsymbol{u}_m] \in \mathbb{R}^{m \times m}$: matrix of eigenvectors of $\boldsymbol{X}\boldsymbol{X}^T$
- $\boldsymbol{V} = [\boldsymbol{v}_1 \ \mathbf{v}_2 \ \cdots \ \boldsymbol{v}_N] \in \mathbb{R}^{N \times N}$: matrix of eigenvectors of $\boldsymbol{X}^T\boldsymbol{X}$
- $\boldsymbol{\Lambda} \in \mathbb{R}^{m \times N}$: diagonal matrix containing singular values

Singular Value Decomposition (SVD)

- Generalization of PCA
- Data matrix $\boldsymbol{X} \in \mathbb{R}^{m \times N}$ can be factorized as:

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^T = \sum_{i=1}^{r} \lambda_i \boldsymbol{u}_i \boldsymbol{v}_i^T$$

- $\boldsymbol{U} = [\boldsymbol{u}_1 \ \boldsymbol{u}_2 \ \cdots \ \boldsymbol{u}_m] \in \mathbb{R}^{m \times m}$: matrix of eigenvectors of $\boldsymbol{X}\boldsymbol{X}^T$
- $\boldsymbol{V} = [\boldsymbol{v}_1 \ \mathbf{v}_2 \ \cdots \ \boldsymbol{v}_N] \in \mathbb{R}^{N \times N}$: matrix of eigenvectors of $\boldsymbol{X}^T\boldsymbol{X}$
- $\boldsymbol{\Lambda} \in \mathbb{R}^{m \times N}$: diagonal matrix containing singular values

Properties:

- r: rank of $\boldsymbol{X}$
- $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_r > 0$
- $\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}_m$, $\boldsymbol{V}^T\boldsymbol{V} = \boldsymbol{I}_N$

Singular Value Decomposition (SVD)

- Low-rank approximation:

$$\boldsymbol{X}_{k} = \sum_{i=1}^{k} \lambda_i \boldsymbol{u}_i \boldsymbol{v}_i^T$$

- Dimensionality reduction when $k \ll r$
- Robustness to noise
- Of all k-rank approximations, $\boldsymbol{X}_k$ is optimal

$$\boldsymbol{X}_k = \arg \min_{\text{rank}(\tilde{\boldsymbol{X}})=k} \|\boldsymbol{X} - \tilde{\boldsymbol{X}}\|_F$$

Singular Value Decomposition (SVD)

- Low-rank approximation:

$$\boldsymbol{X}_{k} = \sum_{i=1}^{k} \lambda_i \boldsymbol{u}_i \boldsymbol{v}_i^T$$

- Dimensionality reduction when $k \ll r$
- Robustness to noise
- Of all k-rank approximations, $\boldsymbol{X}_k$ is optimal

$$\boldsymbol{X}_k = \arg \min_{\text{rank}(\tilde{\boldsymbol{X}})=k} \|\boldsymbol{X} - \tilde{\boldsymbol{X}}\|_F$$

Drawbacks:

- Orthogonality of basis vectors unnatural for ATR problem
- $\boldsymbol{U}$ and $\boldsymbol{V}$ have both positive and negative elements in general $\rightarrow$ interpretation of basis vectors difficult

Non-negative Matrix Approximation (NNMA)

- Follows from non-negative matrix factorization (NMF) technique[14]

$$\boldsymbol{X} = \boldsymbol{W}\boldsymbol{H}; \quad \boldsymbol{W}, \boldsymbol{H} \geq \boldsymbol{0}$$

- SAR ATR: Underlying generative model is a linear combination of basis functions with element-wise non-negative components
- Ready interpretation of $\boldsymbol{W}$ as basis matrix
- Dimensionality reduction: choose $\boldsymbol{W}_k$ (first k columns) instead of $\boldsymbol{W}$

[14] Lee and Seung, Nature, 1999

Non-negative Matrix Approximation (NNMA)

- Follows from non-negative matrix factorization (NMF) technique[14]

$$\boldsymbol{X} = \boldsymbol{W}\boldsymbol{H}; \quad \boldsymbol{W}, \boldsymbol{H} \geq \mathbf{0}$$

- SAR ATR: Underlying generative model is a linear combination of basis functions with element-wise non-negative components
- Ready interpretation of $\boldsymbol{W}$ as basis matrix
- Dimensionality reduction: choose $\boldsymbol{W}_k$ (first k columns) instead of $\boldsymbol{W}$

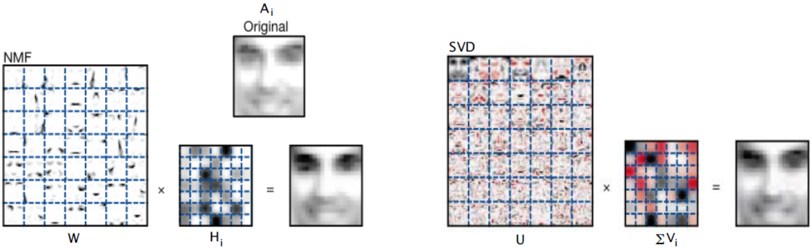


Figure: Illustration: NMF vs. PCA for image representation.

[14] Lee and Seung, Nature, 1999

Non-negative Matrix Approximation (NNMA)

Properties:

- Basis vectors $\boldsymbol{w}_i$ not orthogonal by design
- Sparsity of $\boldsymbol{W}, \boldsymbol{H}$ can be enforced additionally
- $\boldsymbol{W}, \boldsymbol{H}$ not unique

Non-negative Matrix Approximation (NNMA)

Properties:

- Basis vectors $\boldsymbol{w}_i$ not orthogonal by design
- Sparsity of $\boldsymbol{W}, \boldsymbol{H}$ can be enforced additionally
- $\boldsymbol{W}, \boldsymbol{H}$ not unique

Advantages over SVD/PCA for ATR:

- Easy interpretation of basis vectors
- No orthogonality restriction on basis vectors

Non-negative Matrix Approximation (NNMA)

Alternating Least Squares[15]:

$$\min_{\boldsymbol{W},\boldsymbol{H}} \quad \|\boldsymbol{X} - \boldsymbol{W}\boldsymbol{H}\|_F^2$$
$$\text{s.t.} \quad \boldsymbol{W}, \boldsymbol{H} \geq \boldsymbol{0}$$

- Not jointly convex in $\boldsymbol{W}, \boldsymbol{H}$ (separably convex however)

[15] Paatero and Tapper, 1994

[16] Lee and Seung, 2000

Non-negative Matrix Approximation (NNMA)

Alternating Least Squares[15]:

$$\min_{\boldsymbol{W},\boldsymbol{H}} \quad \|\boldsymbol{X} - \boldsymbol{W}\boldsymbol{H}\|_F^2$$
$$\text{s.t.} \quad \boldsymbol{W}, \boldsymbol{H} \geq \mathbf{0}$$

- Not jointly convex in $\boldsymbol{W}, \boldsymbol{H}$ (separably convex however)

Alternate formulation: Divergence update[16]

$$\min_{\boldsymbol{W},\boldsymbol{H}} D(\boldsymbol{X}\|\boldsymbol{W}\boldsymbol{H}) = \sum_{i,j} \left(\boldsymbol{X}_{ij} \log \frac{\boldsymbol{X}_{ij}}{[\boldsymbol{W}\boldsymbol{H}]_{ij}} - \boldsymbol{X}_{ij} + [\boldsymbol{W}\boldsymbol{H}]_{ij} \right)$$
$$\text{s.t.} \quad \boldsymbol{W}, \boldsymbol{H} \geq \mathbf{0}$$

[15] Paatero and Tapper, 1994

[16] Lee and Seung, 2000

Non-negative Matrix Approximation (NNMA)

Alternating Least Squares[15]:

$$\min_{\boldsymbol{W},\boldsymbol{H}} \quad \|\boldsymbol{X} - \boldsymbol{W}\boldsymbol{H}\|_F^2$$
$$\text{s.t.} \quad \boldsymbol{W}, \boldsymbol{H} \geq \boldsymbol{0}$$

- Not jointly convex in $\boldsymbol{W}, \boldsymbol{H}$ (separably convex however)

Alternate formulation: Divergence update[16]

$$\min_{\boldsymbol{W},\boldsymbol{H}} D(\boldsymbol{X}||\boldsymbol{W}\boldsymbol{H}) = \sum_{i,j}\left(\boldsymbol{X}_{ij}\log\frac{\boldsymbol{X}_{ij}}{[\boldsymbol{W}\boldsymbol{H}]_{ij}} - \boldsymbol{X}_{ij} + [\boldsymbol{W}\boldsymbol{H}]_{ij}\right)$$
$$\text{s.t.} \quad \boldsymbol{W}, \boldsymbol{H} \geq \boldsymbol{0}$$

Feature extraction (corresponding to target vector $\boldsymbol{y}$):

$$\boldsymbol{h} = \min_h \|\boldsymbol{y} - \boldsymbol{W}\boldsymbol{h}\|_2, \text{ s.t. } \boldsymbol{h} \geq 0$$

[15] Paatero and Tapper, 1994

[16] Lee and Seung, 2000

Support Vector Machine (SVM)[19]

- Decision function of binary SVM classifier:

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

  where $\boldsymbol{s}_i$ are support vectors, N is the number of support vectors, $\{y_i\}$ are support vector class labels.

- Kernel $K : \mathbb{R}^n \times \mathbb{R}^n \mapsto \mathbb{R}$ maps feature space to higher-dimensional space where separating hyperplane may be more easily determined
- Binary classification decision for $\boldsymbol{x}$ depending on whether $f(\boldsymbol{x}) > 0$ or otherwise
- Multi-class classifiers: one-versus-all approach
- Widely used in ATR problems[17,18]

[17] Zhao and Principe, IEEE Trans. Aerosp. Electron. Syst., 2001

[18] Casasent and Wang, Neural Networks, 2005

[19] Vapnik, The nature of statistical learning theory, 1995

Overall classification framework

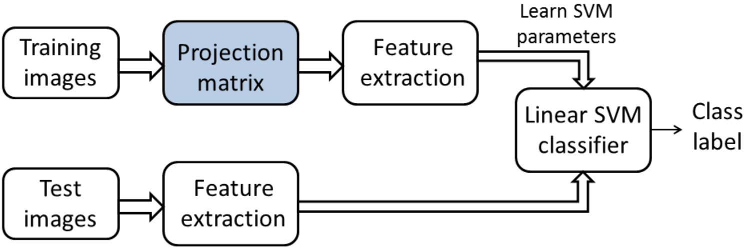


Figure: Proposed target classification framework.

- Projection matrices obtained via PCA and NNMA for feature extraction
- Linear SVM: representative of state-of-the-art classifiers

Experimental set-up

- MSTAR database: one-foot resolution X-band SAR iamges
- Five target classes
  1. T-72 tanks
  2. BMP-2 infantry fighting vehicles
  3. BTR-70 armored personnel carriers
  4. ZIL131 trucks
  5. D7 tractors

Target class	Serial number	# Training images	# Test images
BMP-2	SN_C21	233	196
	SN_9563	233	195
	SN_9566	232	196
BTR-70	SN_C71	233	196
T-72	SN_132	232	196
	SN_812	231	195
	SN_S7	228	191
ZIL131	-	299	274
D7	-	299	274

Table: Target classes in the experiment.

Experimental set-up

- Training images: 17° depression angle
- Test images: 15° depression angle
- Images cropped to 64×64 pixels (i.e. vectorized data in $\mathbb{R}^{4096}$)
- Number of basis vectors: 750 (both PCA and NNMA)

Results: Classification performance

Table: Confusion matrix: PCA basis.

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: Confusion matrix: NNMA basis.

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**

Conclusions

- Non-negative matrix approximation is a suitable choice for feature projection in ATR problems
  - Non-negativity motivated by underlying image physics
  - Achieves dimensionality reduction and captures inter-class variations
  - Better classification performance compared to traditional PCA features

Conclusions

- Non-negative matrix approximation is a suitable choice for feature projection in ATR problems
  - Non-negativity motivated by underlying image physics
  - Achieves dimensionality reduction and captures inter-class variations
  - Better classification performance compared to traditional PCA features
- Future work:
  - NNMA features for meta-classification
  - Class-specific dictionary design

Thank You

Questions?