SAR Automatic Target Recognition via Non-negative Matrix Approximations

Vahid Riasati†       Umamahesh Srinivas‡       Vishal Monga‡

†MacAulay-Brown Inc.       ‡Pennsylvania State University
Dayton, OH       University Park, PA


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
   - Eigen-templates
   - Transform domain coefficients - wavelets

---

4. Casasent et al., Neural Networks, 2005
5. Daniell et al., Optical Engineering, 1992

04/24/2012
Target classification

Two-stage framework:

1. **Feature extraction** from sensed imagery
   - Geometric feature-point descriptors
   - Eigen-templates
   - Transform domain coefficients - wavelets

2. **Decision engine** which performs class assignment
   - Linear and quadratic discriminant analysis
   - Neural networks
   - Support vector machines (SVM)

---

4. Casasent et al., Neural Networks, 2005
5. Daniell et al., Optical Engineering, 1992

04/24/2012
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
  - Voting strategy
  - Boosting
  - Meta-classification

---

7 Paul et al., IEEE ICASSP, 2003
8 Gomes et al., IEEE Radar Conf., 2008
10 Srinivas et al., IEEE Radar Conf., 2011

04/24/2012
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
  - Inherent low-dimensional space that captures image information with minimal redundancy\(^{11}\)
  - Computational benefits for real-time applications

Optimization problem:

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

- \(y\): target image in \(\mathbb{R}^m\)
- \(x\): corresponding feature vector in \(\mathbb{R}^n, n < m\)
- \(A\): 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
  - Inherent low-dimensional space that captures image information with minimal redundancy\(^\text{11}\)
  - Computational benefits for real-time applications

- Optimization problem:
  \[ \hat{x} = \arg \min_{\hat{x}} \|y - A\hat{x}\|_2 \]

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

How to choose \(A\)?

\(^{11}\) Jolliffe, Principal Component Analysis, Springer, 1986

04/24/2012
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 \textit{linear} combination of basis vectors

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

\textsuperscript{12} Turk and Pentland, IEEE Conf. CVPR, 1991
\textsuperscript{13} Bhatnagar et al., IEEE ICASSP, 1998
Singular Value Decomposition (SVD)

- Generalization of PCA
- Data matrix $X \in \mathbb{R}^{m \times N}$ can be factorized as:
  $$X = U \Lambda V^T = \sum_{i=1}^{r} \lambda_i u_i v_i^T$$
- $U = [u_1 \ u_2 \ \cdots \ u_m] \in \mathbb{R}^{m \times m}$: matrix of eigenvectors of $XX^T$
- $V = [v_1 \ v_2 \ \cdots \ v_N] \in \mathbb{R}^{N \times N}$: matrix of eigenvectors of $X^TX$
- $\Lambda \in \mathbb{R}^{m \times N}$: diagonal matrix containing singular values
Singular Value Decomposition (SVD)

- Generalization of PCA

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

$$X = U \Lambda V^T = \sum_{i=1}^{r} \lambda_i u_i v_i^T$$

- $U = [u_1 \ u_2 \ \cdots \ u_m] \in \mathbb{R}^{m \times m}$: matrix of eigenvectors of $XX^T$
- $V = [v_1 \ v_2 \ \cdots \ v_N] \in \mathbb{R}^{N \times N}$: matrix of eigenvectors of $X^TX$
- $\Lambda \in \mathbb{R}^{m \times N}$: diagonal matrix containing singular values

Properties:

- $r$: rank of $X$
- $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_r > 0$
- $U^TU = I_m$, $V^TV = I_N$

04/24/2012
Singular Value Decomposition (SVD)

- Low-rank approximation:
  \[ X_k = \sum_{i=1}^{k} \lambda_i u_i v_i^T \]

- Dimensionality reduction when \( k \ll r \)

- Robustness to noise

- Of all \( k \)-rank approximations, \( X_k \) is optimal
  \[ X_k = \arg \min_{\text{rank}(\tilde{X})=k} \| X - \tilde{X} \|_F \]
Singular Value Decomposition (SVD)

- Low-rank approximation:

  \[ X_k = \sum_{i=1}^{k} \lambda_i u_i v_i^T \]

- Dimensionality reduction when \( k \ll r \)

- Robustness to noise

- Of all \( k \)-rank approximations, \( X_k \) is optimal

  \[
  X_k = \arg \min \operatorname{rank}(\tilde{X})=k \|X - \tilde{X}\|_F
  \]

Drawbacks:

- Orthogonality of basis vectors unnatural for ATR problem

- \( U \) and \( V \) 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\(^\text{14}\)
  \[ X = WH; \quad W, H \geq 0 \]
- SAR ATR: Underlying generative model is a linear combination of basis functions with element-wise non-negative components
- Ready interpretation of \( W \) as basis matrix
- Dimensionality reduction: choose \( W_k \) (first \( k \) columns) instead of \( W \)

---

\(^\text{14}\) Lee and Seung, Nature, 1999
Non-negative Matrix Approximation (NNMA)

- Follows from non-negative matrix factorization (NMF) technique:\[ X = WH; \quad W, H \geq 0 \]
- SAR ATR: Underlying generative model is a linear combination of basis functions with element-wise non-negative components
- Ready interpretation of $W$ as basis matrix
- Dimensionality reduction: choose $W_k$ (first $k$ columns) instead of $W$

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

---

14 Lee and Seung, Nature, 1999

04/24/2012
Non-negative Matrix Approximation (NNMA)

Properties:

- Basis vectors $w_i$ not orthogonal by design
- Sparsity of $W, H$ can be enforced additionally
- $W, H$ not unique
Non-negative Matrix Approximation (NNMA)

Properties:
- Basis vectors $w_i$ not orthogonal by design
- Sparsity of $W, H$ can be enforced additionally
- $W, H$ 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\textsuperscript{15}:  

\[
\begin{align*}
\min_{W, H} & \quad \| X - WH \|_F^2 \\
\text{s.t.} & \quad W, H \geq 0
\end{align*}
\]

- Not jointly convex in $W, H$ (separably convex however)
Non-negative Matrix Approximation (NNMA)

Alternating Least Squares\textsuperscript{15}:

$$\min_{W,H} \|X - WH\|_F^2$$

s.t. \( W, H \geq 0 \)

- Not jointly convex in \( W, H \) (separably convex however)

Alternate formulation: Divergence update\textsuperscript{16}

$$\min_{W,H} \mathcal{D}(X\|WH) = \sum_{i,j} \left( X_{ij} \log \frac{X_{ij}}{[WH]_{ij}} - X_{ij} + [WH]_{ij} \right)$$

s.t. \( W, H \geq 0 \)

\textsuperscript{15} Paatero and Tapper, 1994

\textsuperscript{16} Lee and Seung, 2000
Non-negative Matrix Approximation (NNMA)

Alternating Least Squares\(^{15}\):

\[
\min_{W,H} \| X - WH \|_F^2 \\
\text{s.t. } W, H \geq 0
\]

- Not jointly convex in \( W, H \) (separably convex however)

Alternate formulation: Divergence update\(^{16}\)

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

Feature extraction (corresponding to target vector \( y \)):

\[
h = \min_h \| y - Wh \|_2, \text{ s.t. } h \geq 0
\]

---

\(^{15}\) Paatero and Tapper, 1994

\(^{16}\) Lee and Seung, 2000
Support Vector Machine (SVM)\textsuperscript{19}

- Decision function of binary SVM classifier:

\[
  f(x) = \sum_{i=1}^{N} \alpha_i y_i K(s_i, x) + b,
\]

where \(s_i\) 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 \(x\) depending on whether \(f(x) > 0\) or otherwise.

- Multi-class classifiers: one-versus-all approach.

- Widely used in ATR problems\textsuperscript{17,18}

\textsuperscript{18} Casasent and Wang, Neural Networks, 2005
\textsuperscript{19} Vapnik, The nature of statistical learning theory, 1995
Overall classification framework

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

Figure: Proposed target classification framework.
Experimental set-up

- MSTAR database: one-foot resolution X-band SAR images
- 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

<table>
<thead>
<tr>
<th>Target class</th>
<th>Serial number</th>
<th># Training images</th>
<th># Test images</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMP-2</td>
<td>SN_C21</td>
<td>233</td>
<td>196</td>
</tr>
<tr>
<td></td>
<td>SN_9563</td>
<td>233</td>
<td>195</td>
</tr>
<tr>
<td></td>
<td>SN_9566</td>
<td>232</td>
<td>196</td>
</tr>
<tr>
<td>BTR-70</td>
<td>SN_C71</td>
<td>233</td>
<td>196</td>
</tr>
<tr>
<td>T-72</td>
<td>SN_132</td>
<td>232</td>
<td>196</td>
</tr>
<tr>
<td></td>
<td>SN_812</td>
<td>231</td>
<td>195</td>
</tr>
<tr>
<td></td>
<td>SN_S7</td>
<td>228</td>
<td>191</td>
</tr>
<tr>
<td>ZIL131</td>
<td>-</td>
<td>299</td>
<td>274</td>
</tr>
<tr>
<td>D7</td>
<td>-</td>
<td>299</td>
<td>274</td>
</tr>
</tbody>
</table>

Table: Target classes in the experiment.
Experimental set-up

- Training images: $17^\circ$ depression angle
- Test images: $15^\circ$ depression angle
- Images cropped to $64 \times 64$ 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.

<table>
<thead>
<tr>
<th>Class</th>
<th>BMP-2</th>
<th>BTR-70</th>
<th>T-72</th>
<th>ZIL131</th>
<th>D7</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMP-2</td>
<td>0.84</td>
<td>0.06</td>
<td>0.04</td>
<td>0.02</td>
<td>0.04</td>
</tr>
<tr>
<td>BTR-70</td>
<td>0.05</td>
<td>0.87</td>
<td>0.03</td>
<td>0.02</td>
<td>0.03</td>
</tr>
<tr>
<td>T-72</td>
<td>0.03</td>
<td>0.07</td>
<td>0.83</td>
<td>0.03</td>
<td>0.04</td>
</tr>
<tr>
<td>ZIL131</td>
<td>0.05</td>
<td>0.03</td>
<td>0.02</td>
<td>0.84</td>
<td>0.06</td>
</tr>
<tr>
<td>D7</td>
<td>0.06</td>
<td>0.02</td>
<td>0.04</td>
<td>0.06</td>
<td>0.82</td>
</tr>
</tbody>
</table>

Table: Confusion matrix: NNMA basis.

<table>
<thead>
<tr>
<th>Class</th>
<th>BMP-2</th>
<th>BTR-70</th>
<th>T-72</th>
<th>ZIL131</th>
<th>D7</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMP-2</td>
<td>0.86</td>
<td>0.05</td>
<td>0.02</td>
<td>0.05</td>
<td>0.02</td>
</tr>
<tr>
<td>BTR-70</td>
<td>0.07</td>
<td>0.88</td>
<td>0.04</td>
<td>0.01</td>
<td>0.0</td>
</tr>
<tr>
<td>T-72</td>
<td>0.03</td>
<td>0.04</td>
<td>0.86</td>
<td>0.02</td>
<td>0.05</td>
</tr>
<tr>
<td>ZIL131</td>
<td>0.01</td>
<td>0.06</td>
<td>0.05</td>
<td>0.87</td>
<td>0.01</td>
</tr>
<tr>
<td>D7</td>
<td>0.04</td>
<td>0.02</td>
<td>0.06</td>
<td>0.04</td>
<td>0.84</td>
</tr>
</tbody>
</table>
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?