Meta-classifiers for exploiting feature dependencies in automatic target recognition

iPAL Group Meeting

September 03, 2010
(Work being submitted to IEEE Radar Conference 2011)
Outline

- Automatic Target Recognition
- Meta-classification
- Image Pre-processing
- Individual classification schemes
- Support Vector Machines
- Boosting
- Experiments
- Results
- Conclusions

09/03/2010 iPAL Group Meeting
Automatic Target Recognition (ATR)

- Automatic (or aided) identification and recognition of targets
- Highly important capability for defense weapon systems\(^1\)
- Data acquired by a variety of sensors: SAR, ISAR, FLIR, LADAR, hyperspectral.
- Diverse scenarios: air-to-ground, air-to-air, surface-to-surface

\[\text{Figure: Sample targets and their SAR images. Courtesy: Gomes et al.}\]

\(^{1}\text{Bhanu et al., IEEE AES Systems Magazine, 1993}\]
ATR System description

- **Detection and discrimination**: Identification of target signatures in the presence of clutter
- **Denoising**: Useful pre-processing step, especially for synthetic aperture radar (SAR) imagery, known to suffer from speckle noise
- **Classification**: Separation of targets into different classes
- **Recognition**: Distinguishing between sub-classes within a target class; harder problem than classification

**Figure**: Schematic of general ATR system.
Target classification

Two main components:

- **Feature extraction**: Image dimensionality-reduction operation
  - Geometric feature-point descriptors (Olson et al., 1997)
  - Transform domain coefficients (Casasent et al., 2005)
  - Eigen-templates (Bhatnagar et al., 1998)

- **Decision engine**: Makes classification decisions
  - Linear and quadratic discriminant analysis
  - Neural networks (Daniell et al., 1992)
  - Support vector machines (SVM) (Zhao et al., 2001)
Motivation for current work

- Search for ‘best possible’ identification features
- Limited understanding of inter-relationships among different sets of features
- No single feature extractor and decision engine optimal from a classification standpoint

---

2 Paul et al., ICASSP 2003  
3 Gomes et al., IEEE Radar Conf., 2008
Motivation for current work

- Search for ‘best possible’ identification features
- Limited understanding of inter-relationships among different sets of features
- No single feature extractor and decision engine optimal from a classification standpoint
- Exploit complementary benefits offered by different sets of features

\(^2\)Paul et al., ICASSP 2003
\(^3\)Gomes et al., IEEE Radar Conf., 2008
Motivation for current work

- Search for ‘best possible’ identification features
- Limited understanding of inter-relationships among different sets of features
- No single feature extractor and decision engine optimal from a classification standpoint
- Exploit complementary benefits offered by different sets of features
- Prior attempts at ATR composite classifiers: same set of features with different decision engines\(^2,^3\)

\(^2\)Paul et al., ICASSP 2003
\(^3\)Gomes et al., IEEE Radar Conf., 2008
Meta-classification

- Principled strategy to exploit complementary benefits (compared to heuristic fusion techniques so far)

- Inspired by recent work in multimodal document classification\(^4\)

- **Meta-classifier**: Combines classifier decisions from individual classifiers to improve overall classification performance

- **Two-stage approach**:
  - Soft outputs from individual classifiers
  - Classification using composite meta-feature vector

- Two intuitively-motivated schemes proposed for SAR imagery:
  - Meta-classification using **SVMs**
  - Meta-classification using **boosting**

---

\(^4\)Chen et al., MMSP 2009
Image pre-processing

- SAR images degraded due to low spatial resolution and contrast, clutter, noise

- **Speckle noise**: Interference between radar waves reflected off target; signal-dependent and multiplicative

\[ y[m] = x[m] + \sqrt{x[m]} \cdot n[m] \]

- Speckle denoising: important inverse problem\(^5\); not explored so far as pre-processing step in SAR ATR

- Denoising using anisotropic diffusion\(^6\): better mean preservation, variance reduction and edge localization

- Registration of image templates

---

\(^5\)Frost et al., IEEE PAMI 1982
\(^6\)Yu et al., IEEE TIP 2002
Individual classifier schemes

Three different feature extractor-decision engine combinations:

- Wavelet features + neural network
- Eigen-templates + correlation
- Scale invariant feature transform (SIFT) + SVM
Classifier 1

- Transform domain features
- LL sub-band coefficients from two-level decomposition using reverse biorthogonal mother wavelets
- Multilayer perceptron neural network (Gomes et al.)
  - One hidden layer
  - Sigmoid logistic activation function
  - Back-propagation to update weights
Classifier 2

- Eigen-templates as feature vectors\textsuperscript{7}
- Spatial domain features
- Training class template: eigen-vector corresponding to largest singular value of training data matrix
- Correlation score decision engine

\textsuperscript{7}Bhatnagar et al., IEEE 1998
Classifier 3

- Computer vision-based features
- SIFT: robustness to change in image scale, illumination, local geometric transformations and noise
- SVM decision engine\(^8\)

\(^8\)Grauman et al., ICCV 2005
Support vector machines

**Problem**: Given \( m \) i.i.d. observations \((x_i, y_i), x_i \in \mathbb{R}^n, y_i \in \{-1, +1\}\), \( i = 1, 2, \ldots, m \) drawn from a distribution \( P(x, y) \), learn the mapping \( x_i \mapsto y_i \).

\[
R \leq R_{emp} + \sqrt{\left( \frac{h(\log(2m/h) + 1) - \log(\eta/4)}{m} \right)},
\]

where \( R \) is the generalization error, \( R_{emp} \) is the empirical error and \( h \) is the Vapnik-Chervonenkis dimension.

- **Structural risk minimization**: minimize the upper bound for the generalization error.
Margin maximization
Margin maximization

- Determine separating hyperplane $\mathbf{w} \cdot \mathbf{x} + b = 0$ with largest margin
- Maximize $\frac{2}{\|\mathbf{w}\|}$ subject to $y_i(\mathbf{w} \cdot \mathbf{x}_i + b - 1) \geq 0 \ \forall \ i$
- Equivalently, minimize $\|\mathbf{w}\|^2$ subject to $y_i(\mathbf{w} \cdot \mathbf{x}_i + b - 1) \geq 0 \ \forall \ i$
- Minimize $L_P = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^{m} \alpha_i y_i(\mathbf{w} \cdot \mathbf{x}_i + b) + \sum_{i=1}^{m} \alpha_i$
- Convex quadratic programming problem $\Rightarrow$ solve the dual problem
- Maximize $L_D = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j$
- KKT conditions
SVM classifier

- 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.

- Kernel \(K : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}\) 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.
Boosting

- Boost the performance of weak learners into a classification algorithm with arbitrarily accurate performance
- Maintain a distribution of weights over the training set
- Weights on incorrectly classified examples are increased iteratively
- Slow learners are penalized for harder examples
AdaBoost algorithm

Algorithm 1 Adaboost learning algorithm
1: Input data $(x_i, y_i), i = 1, 2, \ldots, N$, where $x_i \in S, y_i \in \{-1, +1\}$
2: Initialize $D_1(i) = \frac{1}{N}, i = 1, 2, \ldots, N$
3: For $t = 1, 2, \ldots, T$
   - Train weak learner using distribution $D_t$
   - Determine weak hypothesis $f_t : S \mapsto \{-1, +1\}$ with error $\epsilon_t$
   - Choose $\beta_t = \frac{1}{2} \ln \left( \frac{1 - \epsilon_t}{\epsilon_t} \right)$
   - $D_{t+1}(i) = \frac{D_t(i) \exp(\beta_t y_i f_t(x_i))}{Z_t}$, where $Z_t$ is a normalization factor
4: Output soft decision $F(x) = \sum_{t=1}^{T} \beta_t f_t(x)$. 

09/03/2010 iPAL Group Meeting
SVM-based meta-classification

- SAR Images
  - Feature extractor
    - Wavelet coefficients
    - Eigen-vectors
    - SIFT
  - Decision engine
    - Neural network
    - Correlation
    - SVM
      - Linear kernel
      - RBF kernel
    - Soft outputs
  - SVM Meta-classifier
    - Target class
AdaBoost-based meta-classification

- SAR Images
- Feature extractor: Wavelet coefficients, Eigen-vectors
  - Neural network
  - Correlation
  - SIFT, SVM
- Decision engine: Soft outputs
- AdaBoost-based Metaclassifier
- Target class
Experiments

- Moving and Stationary Target Acquisition and Recognition (MSTAR) database for SAR images

- Advantages of SAR: reduced sensitivity to weather conditions, day-night operation, penetration capability through obstacles

- Two sets of experiments to bring out differences between classification and recognition

- Five target classes: T-72 tanks, BMP-2 infantry fighting vehicles, BTR-70 armored personnel carriers, ZIL trucks and D7 tractors

- SLICY confusers to test rejection performance

- Confusion matrix gives classification rates
## Datasets

<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:** The target classes used in the experiment.
Results: Classification

Table: Confusion matrix for wavelet features + neural network classifier.

<table>
<thead>
<tr>
<th></th>
<th>BMP-2</th>
<th>BTR-70</th>
<th>T-72</th>
<th>ZIL131</th>
<th>D7</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMP-2</td>
<td>0.80</td>
<td>0.06</td>
<td>0.09</td>
<td>0.01</td>
<td>0.04</td>
<td>0</td>
</tr>
<tr>
<td>BTR-70</td>
<td>0.03</td>
<td>0.93</td>
<td>0.02</td>
<td>0</td>
<td>0.02</td>
<td>0</td>
</tr>
<tr>
<td>T-72</td>
<td>0.08</td>
<td>0</td>
<td>0.77</td>
<td>0.10</td>
<td>0.04</td>
<td>0.01</td>
</tr>
<tr>
<td>ZIL131</td>
<td>0.08</td>
<td>0</td>
<td>0.05</td>
<td>0.84</td>
<td>0.03</td>
<td>0</td>
</tr>
<tr>
<td>D7</td>
<td>0</td>
<td>0.03</td>
<td>0.06</td>
<td>0.05</td>
<td>0.86</td>
<td>0</td>
</tr>
<tr>
<td>Confuser</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0.99</td>
</tr>
</tbody>
</table>
## Results: Classification

**Table:** Confusion matrix for eigen-template matching classifier.

<table>
<thead>
<tr>
<th></th>
<th>BMP-2</th>
<th>BTR-70</th>
<th>T-72</th>
<th>ZIL131</th>
<th>D7</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMP-2</td>
<td>0.76</td>
<td>0.09</td>
<td>0.05</td>
<td>0.03</td>
<td>0.05</td>
<td>0.02</td>
</tr>
<tr>
<td>BTR-70</td>
<td>0.04</td>
<td>0.88</td>
<td>0.05</td>
<td>0</td>
<td>0.03</td>
<td>0</td>
</tr>
<tr>
<td>T-72</td>
<td>0.06</td>
<td>0.06</td>
<td>0.73</td>
<td>0.10</td>
<td>0.04</td>
<td>0.01</td>
</tr>
<tr>
<td>ZIL131</td>
<td>0.02</td>
<td>0.04</td>
<td>0.07</td>
<td>0.79</td>
<td>0.08</td>
<td>0</td>
</tr>
<tr>
<td>D7</td>
<td>0</td>
<td>0.03</td>
<td>0.06</td>
<td>0.04</td>
<td>0.87</td>
<td>0</td>
</tr>
<tr>
<td>Confuser</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.99</td>
</tr>
</tbody>
</table>
Results: Classification

**Table**: Confusion matrix for SIFT features + linear SVM classifier.

<table>
<thead>
<tr>
<th></th>
<th>BMP-2</th>
<th>BTR-70</th>
<th>T-72</th>
<th>ZIL131</th>
<th>D7</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMP-2</td>
<td><strong>0.85</strong></td>
<td>0.07</td>
<td>0.03</td>
<td>0</td>
<td>0.03</td>
<td>0.02</td>
</tr>
<tr>
<td>BTR-70</td>
<td>0.02</td>
<td><strong>0.91</strong></td>
<td>0.05</td>
<td>0</td>
<td>0.02</td>
<td>0</td>
</tr>
<tr>
<td>T-72</td>
<td>0.03</td>
<td>0.04</td>
<td><strong>0.82</strong></td>
<td>0.06</td>
<td>0.04</td>
<td>0.01</td>
</tr>
<tr>
<td>ZIL131</td>
<td>0</td>
<td>0.04</td>
<td>0.03</td>
<td><strong>0.86</strong></td>
<td>0.07</td>
<td>0</td>
</tr>
<tr>
<td>D7</td>
<td>0</td>
<td>0</td>
<td>0.06</td>
<td>0.05</td>
<td><strong>0.89</strong></td>
<td>0</td>
</tr>
<tr>
<td>Confuser</td>
<td>0.01</td>
<td>0</td>
<td>0.02</td>
<td>0</td>
<td>0</td>
<td><strong>0.97</strong></td>
</tr>
</tbody>
</table>
Results: Classification

Table: Confusion matrix for SVM meta-classifier.

<table>
<thead>
<tr>
<th></th>
<th>BMP-2</th>
<th>BTR-70</th>
<th>T-72</th>
<th>ZIL131</th>
<th>D7</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMP-2</td>
<td>0.91</td>
<td>0.03</td>
<td>0.02</td>
<td>0.02</td>
<td>0.03</td>
<td>0</td>
</tr>
<tr>
<td>BTR-70</td>
<td>0.01</td>
<td>0.94</td>
<td>0.02</td>
<td>0.01</td>
<td>0.02</td>
<td>0</td>
</tr>
<tr>
<td>T-72</td>
<td>0.03</td>
<td>0.02</td>
<td>0.89</td>
<td>0.03</td>
<td>0.03</td>
<td>0</td>
</tr>
<tr>
<td>ZIL131</td>
<td>0.01</td>
<td>0.04</td>
<td>0.03</td>
<td>0.89</td>
<td>0.03</td>
<td>0</td>
</tr>
<tr>
<td>D7</td>
<td>0</td>
<td>0.01</td>
<td>0.05</td>
<td>0.04</td>
<td>0.90</td>
<td>0</td>
</tr>
<tr>
<td>Confuser</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Results: Classification

Table: Confusion matrix for Adaboost meta-classifier.

<table>
<thead>
<tr>
<th></th>
<th>BMP-2</th>
<th>BTR-70</th>
<th>T-72</th>
<th>ZIL131</th>
<th>D7</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMP-2</td>
<td>0.93</td>
<td>0.02</td>
<td>0.03</td>
<td>0.01</td>
<td>0.01</td>
<td>0</td>
</tr>
<tr>
<td>BTR-70</td>
<td>0.02</td>
<td>0.95</td>
<td>0.02</td>
<td>0</td>
<td>0.01</td>
<td>0</td>
</tr>
<tr>
<td>T-72</td>
<td>0.04</td>
<td>0.02</td>
<td>0.89</td>
<td>0.04</td>
<td>0.02</td>
<td>0</td>
</tr>
<tr>
<td>ZIL131</td>
<td>0.01</td>
<td>0.03</td>
<td>0.02</td>
<td>0.90</td>
<td>0.04</td>
<td>0</td>
</tr>
<tr>
<td>D7</td>
<td>0</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.91</td>
<td>0</td>
</tr>
<tr>
<td>Confuser</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.00</td>
</tr>
</tbody>
</table>
**Results: Recognition**

**Table:** BMP-2 Recognition: Confusion matrix for wavelet features + neural network classifier.

<table>
<thead>
<tr>
<th></th>
<th>SN_C21</th>
<th>SN_9563</th>
<th>SN_9566</th>
</tr>
</thead>
<tbody>
<tr>
<td>SN_C21</td>
<td>0.71</td>
<td>0.16</td>
<td>0.13</td>
</tr>
<tr>
<td>SN_9563</td>
<td>0.18</td>
<td>0.68</td>
<td>0.14</td>
</tr>
<tr>
<td>SN_9566</td>
<td>0.10</td>
<td>0.16</td>
<td>0.74</td>
</tr>
</tbody>
</table>
Table: BMP-2 Recognition: Confusion matrix for eigen-template matching classifier.

<table>
<thead>
<tr>
<th></th>
<th>SN_C21</th>
<th>SN_9563</th>
<th>SN_9566</th>
</tr>
</thead>
<tbody>
<tr>
<td>SN_C21</td>
<td>0.69</td>
<td>0.16</td>
<td>0.15</td>
</tr>
<tr>
<td>SN_9563</td>
<td>0.19</td>
<td>0.64</td>
<td>0.17</td>
</tr>
<tr>
<td>SN_9566</td>
<td>0.11</td>
<td>0.18</td>
<td>0.71</td>
</tr>
</tbody>
</table>
Results: Recognition

Table: BMP-2 Recognition: Confusion matrix for SIFT features + linear SVM classifier.

<table>
<thead>
<tr>
<th></th>
<th>SN_C21</th>
<th>SN_9563</th>
<th>SN_9566</th>
</tr>
</thead>
<tbody>
<tr>
<td>SN_C21</td>
<td>0.73</td>
<td>0.15</td>
<td>0.13</td>
</tr>
<tr>
<td>SN_9563</td>
<td>0.13</td>
<td>0.69</td>
<td>0.18</td>
</tr>
<tr>
<td>SN_9566</td>
<td>0.14</td>
<td>0.11</td>
<td>0.75</td>
</tr>
</tbody>
</table>
Table: BMP-2 Recognition: Confusion matrix for SVM meta-classifier.

<table>
<thead>
<tr>
<th></th>
<th>SN_C21</th>
<th>SN_9563</th>
<th>SN_9566</th>
</tr>
</thead>
<tbody>
<tr>
<td>SN_C21</td>
<td>0.75</td>
<td>0.12</td>
<td>0.13</td>
</tr>
<tr>
<td>SN_9563</td>
<td>0.13</td>
<td>0.72</td>
<td>0.15</td>
</tr>
<tr>
<td>SN_9566</td>
<td>0.08</td>
<td>0.13</td>
<td>0.79</td>
</tr>
</tbody>
</table>
# Results: Recognition

## Table: BMP-2 Recognition: Confusion matrix for Adaboost meta-classifier.

<table>
<thead>
<tr>
<th></th>
<th>SN_C21</th>
<th>SN_9563</th>
<th>SN_9566</th>
</tr>
</thead>
<tbody>
<tr>
<td>SN_C21</td>
<td>0.75</td>
<td>0.13</td>
<td>0.12</td>
</tr>
<tr>
<td>SN_9563</td>
<td>0.13</td>
<td>0.73</td>
<td>0.14</td>
</tr>
<tr>
<td>SN_9566</td>
<td>0.10</td>
<td>0.12</td>
<td>0.78</td>
</tr>
</tbody>
</table>
Classification rate versus training size

Misclassification variation with training sample size, target class: BMP–2

- Eigen-template
- SVM meta-classification
- Adaboost meta-classification
Conclusions

- Virtues of different feature extractors and decision engines combined in a principled manner
- Two meta-classification schemes proposed, based on SVM and AdaBoost
- Test on benchmark SAR datasets show improvements in classification performance
- Pre-processing improves classification performance
Acknowledgments

- Prof. Vishal Monga, Penn State
- Dr. Raghu G. Raj, Naval Research Laboratory