Meta-classifiers for Exploiting Feature Dependencies in Automatic Target Recognition

Umamahesh Srinivas[†] Vishal Monga[†] Raghu G. Raj[‡]

[†]Pennsylvania State University
University Park, USA

[‡]U.S. Naval Research Laboratory
Washington DC, USA

2011 IEEE Radar Conference

May 26, 2011

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: Sample targets and their SAR images. Courtesy: Gomes et al.

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

ATR system description

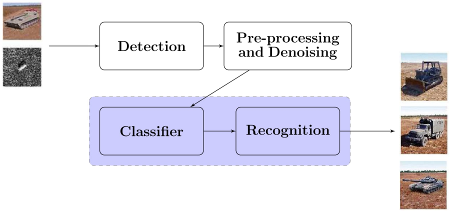


Figure: Schematic of ATR framework. The classification and recognition stages assign an input image/ feature to one of many target classes.

ATR system description

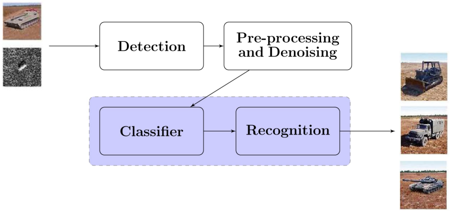


Figure: Schematic of ATR framework. The classification and recognition stages assign an input image/ feature to one of many target classes.

- Detection and discrimination: Identification of target signatures in the presence of clutter

ATR system description

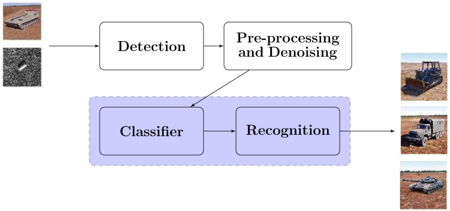


Figure: Schematic of ATR framework. The classification and recognition stages assign an input image/ feature to one of many target classes.

- Detection and discrimination: Identification of target signatures in the presence of clutter
- Denoising: Pre-processing (e.g. removing speckle in SAR imagery)

ATR system description

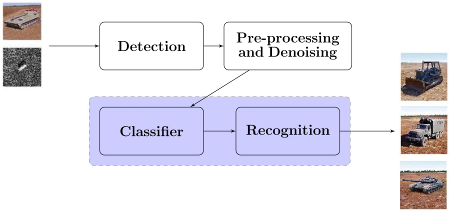


Figure: Schematic of ATR framework. The classification and recognition stages assign an input image/ feature to one of many target classes.

- Detection and discrimination: Identification of target signatures in the presence of clutter
- Denoising: Pre-processing (e.g. removing speckle in SAR imagery)
- Classification: Separation of targets into different classes

ATR system description

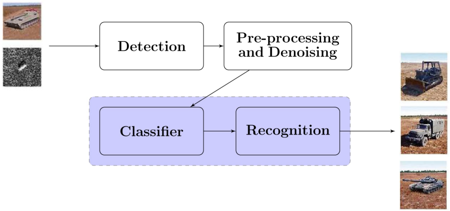


Figure: Schematic of ATR framework. The classification and recognition stages assign an input image/ feature to one of many target classes.

- Detection and discrimination: Identification of target signatures in the presence of clutter
- Denoising: Pre-processing (e.g. removing speckle in SAR imagery)
- Classification: Separation of targets into different classes
- Recognition: Distinguishing between sub-classes within a target class; harder problem than classification

Target classification

- Rich family of algorithmic tools developed over two decades
- Two-stage framework

[2] Olson et al., IEEE Trans. Image Process., 1997
[3] Casasent et al., Neural Networks, 2005
[4] Gomes et al., IEEE Radar Conf., 2008
[5] Bhatnagar et al., IEEE ICASSP, 1998
[6] Grenander et al., IEEE Trans. PAMI, 1998

Target classification

- Rich family of algorithmic tools developed over two decades
- Two-stage framework
- Feature extraction from sensed imagery
    - Geometric feature-point descriptors[2]
    - Transform domain coefficients - wavelets[3,4]
    - Eigen-templates[5]
    - Estimation-theoretic templates[6]

[2] Olson et al., IEEE Trans. Image Process., 1997
[3] Casasent et al., Neural Networks, 2005
[4] Gomes et al., IEEE Radar Conf., 2008
[5] Bhatnagar et al., IEEE ICASSP, 1998
[6] Grenander et al., IEEE Trans. PAMI, 1998

Target classification

- Decision engine which performs class assignment
  - Linear and quadratic discriminant analysis
  - Neural networks[7]
  - Support vector machines (SVM)[8]
  - Hierarchical SVM[9]

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

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

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

Recent research trends

- Search for 'best possible' features from a classification standpoint
- Limited understanding of inter-relationships among different sets of features

Recent research trends

- Search for ‘best possible’ features from a classification standpoint
- Limited understanding of inter-relationships among different sets of features
- No single optimal feature set-decision engine combination
- Exploit complementary yet correlated information offered by different sets of classifiers
    - Classifier fusion

Classifier fusion

- Same set of features with different decision engines; (mostly) educated heuristic schemes
- Combination of outputs from four decision engines using FLIR data[10]
- Product of individual classification probabilities[11]
- Voting strategy[12]
- Boosting[13]

[10] Rizvi et al., Applied Imagery Pattern Recognition Workshop, 2003

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

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

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

[14] Nasrabadi, IEEE Int. Conf. Image Process., 2008

Classifier fusion

- Same set of features with different decision engines; (mostly) educated heuristic schemes
- Combination of outputs from four decision engines using FLIR data[10]
- Product of individual classification probabilities[11]
- Voting strategy[12]
- Boosting[13]
- Related research problem: multi-sensor ATR[14]

[10] Rizvi et al., Applied Imagery Pattern Recognition Workshop, 2003
[11] Paul et al., IEEE ICASSP, 2003
[12] Gomes et al., IEEE Radar Conf., 2008
[13] Sun et al., IEEE Trans. Aerosp. Electron. Syst., 2007
[14] Nasrabadi, IEEE Int. Conf. Image Process., 2008

Contribution of our work

- Meta-classification: Principled strategy to combine complementary benefits[15]

[15] Lin et al., Int. Workshop Knowledge Discovery in Multimedia, 2002

Contribution of our work

- Meta-classification: Principled strategy to combine complementary benefits[15]
- Meta-classifier: Combines classifier decisions from individual classifiers to improve overall classification performance
- Two-stage approach:
  1. Obtain different feature sets via multiple projections (to suitable basis)
  2. Combine "soft" outputs from individual classifiers into composite meta-feature vector for classification

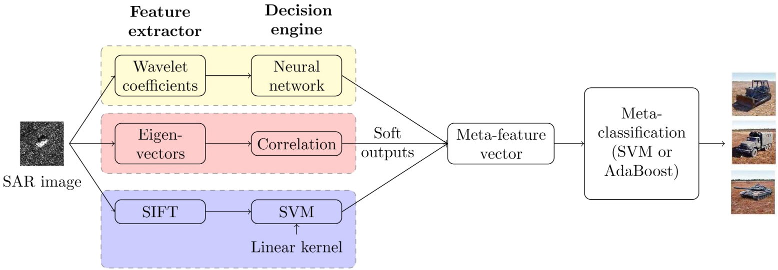
[15] Lin et al., Int. Workshop Knowledge Discovery in Multimedia, 2002

Contribution of our work

- Meta-classification: Principled strategy to combine complementary benefits[15]
- Meta-classifier: Combines classifier decisions from individual classifiers to improve overall classification performance
- Two-stage approach:
  1. Obtain different feature sets via multiple projections (to suitable basis)
  2. Combine "soft" outputs from individual classifiers into composite meta-feature vector for classification
- Two intuitively-motivated schemes proposed for SAR imagery:
  - Meta-classification using SVMs
  - Meta-classification using boosting

[15] Lin et al., Int. Workshop Knowledge Discovery in Multimedia, 2002

Meta-classification framework



- Complementary merits of different sets of features exploited
- Meta-classification: Creates common ground for combination of diverse types of features

Different feature extractors

1. Wavelet features + neural network[16]
   - Transform domain features (in $\mathbb{R}^{256}$)
   - LL sub-band coefficients from two-level decomposition using reverse biorthogonal wavelets
   - Multilayer perceptron neural network

[16] Sandirasegaram, Tech. Memo. DRDC Ottawa, 2005

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

[18] Grauman et al., Int. Conf. Comp. Vision, 2005

Different feature extractors

1. Wavelet features + neural network[16]
   - Transform domain features (in $\mathbb{R}^{256}$)
   - LL sub-band coefficients from two-level decomposition using reverse biorthogonal wavelets
   - Multilayer perceptron neural network
2. Eigen-templates + correlation[17]
   - Spatial domain features (in $\mathbb{R}^{4096}$)
   - Training class template: eigen-vector corresponding to largest singular value of training data matrix
   - Decision engine: correlation score

[16] Sandirasegaram, Tech. Memo. DRDC Ottawa, 2005
[17] Bhatnagar et al., IEEE ICASSP, 1998
[18] Grauman et al., Int. Conf. Comp. Vision, 2005

Different feature extractors

1. Wavelet features + neural network[16]
   - Transform domain features (in $\mathbb{R}^{256}$)
   - LL sub-band coefficients from two-level decomposition using reverse biorthogonal wavelets
   - Multilayer perceptron neural network
2. Eigen-templates + correlation[17]
   - Spatial domain features (in $\mathbb{R}^{4096}$)
   - Training class template: eigen-vector corresponding to largest singular value of training data matrix
   - Decision engine: correlation score
3. Scale invariant feature transform (SIFT) + SVM
   - Computer vision-based features (in $\mathbb{R}^{128}$)
   - SIFT: robustness to change in image scale, illumination, local geometric transformations and noise
   - SVM decision engine[18]

---

[16] Sandirasegaram, Tech. Memo. DRDC Ottawa, 2005

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

[18] Grauman et al., Int. Conf. Comp. Vision, 2005

Support Vector Machine[19]

- Decision function of binary SVM classifier:

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

  where $\mathbf{s}_i$ are support vectors, N is the number of support vectors

- 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 $\mathbf{x}$ depending on whether $f(\mathbf{x}) > 0$ or otherwise
- Multi-class classifiers: one-versus-all approach

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

Boosting[20]

- 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

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 ϵ_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 $F(x) = \sum_{t=1}^{T} \beta_t f_t(x)$.

[20] Freund et al., Journal of Japanese Society for Artificial Intelligence, 1999

Image pre-processing

- SAR imagery: low spatial resolution and contrast, clutter, noise
- Speckle noise: Interference between radar waves reflected off target; signal-dependent and multiplicative

$$y[\mathbf{m}] = x[\mathbf{m}] + \sqrt{x[\mathbf{m}]}\; n[\mathbf{m}]$$

[21] Frost et al., IEEE Trans. PAMI, 1982

[22] Yu et al., IEEE Trans. Image Process., 2002

Image pre-processing

- SAR imagery: low spatial resolution and contrast, clutter, noise
- Speckle noise: Interference between radar waves reflected off target; signal-dependent and multiplicative

$$y[\mathbf{m}] = x[\mathbf{m}] + \sqrt{x[\mathbf{m}]}\, n[\mathbf{m}]$$

- Speckle denoising: important inverse problem[21]
- Denoising using anisotropic diffusion[22]
  - Better mean preservation
  - Variance reduction
  - Edge localization

[21] Frost et al., IEEE Trans. PAMI, 1982

[22] Yu et al., IEEE Trans. Image Process., 2002

Image pre-processing

- SAR imagery: low spatial resolution and contrast, clutter, noise
- Speckle noise: Interference between radar waves reflected off target; signal-dependent and multiplicative

$$y[\mathbf{m}] = x[\mathbf{m}] + \sqrt{x[\mathbf{m}]}\; n[\mathbf{m}]$$

- Speckle denoising: important inverse problem[21]
- Denoising using anisotropic diffusion[22]
    - Better mean preservation
    - Variance reduction
    - Edge localization
- Registration of image templates - frame centering
- Energy normalization

[21] Frost et al., IEEE Trans. PAMI, 1982

[22] Yu et al., IEEE Trans. Image Process., 2002

Experiments

- Moving and Stationary Target Acquisition and Recognition (MSTAR) database for 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
- SLICY confusers to test rejection performance

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.

Results: Classification

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

	BMP-2	BTR-70	T-72	ZIL131	D7	Other
BMP-2	**0.80**	0.06	0.09	0.01	0.04	0
BTR-70	0.03	**0.93**	0.02	0	0.02	0
T-72	0.08	0	**0.77**	0.10	0.04	0.01
ZIL131	0.08	0	0.05	**0.84**	0.03	0
D7	0	0.03	0.06	0.05	**0.86**	0
Confuser	0	0	0.01	0	0	**0.99**

(a)

Table: Confusion matrix for eigen-template matching classifier.

	BMP-2	BTR-70	T-72	ZIL131	D7	Other
BMP-2	**0.76**	0.09	0.05	0.03	0.05	0.02
BTR-70	0.04	**0.88**	0.05	0	0.03	0
T-72	0.06	0.06	**0.73**	0.10	0.04	0.01
ZIL131	0.02	0.04	0.07	**0.79**	0.08	0
D7	0	0.03	0.06	0.04	**0.87**	0
Confuser	0.01	0	0	0	0	**0.99**

(b)

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

	BMP-2	BTR-70	T-72	ZIL131	D7	Other
BMP-2	**0.85**	0.07	0.03	0	0.03	0.02
BTR-70	0.02	**0.91**	0.05	0	0.02	0
T-72	0.03	0.04	**0.82**	0.06	0.04	0.01
ZIL131	0	0.04	0.03	**0.86**	0.07	0
D7	0	0	0.06	0.05	**0.89**	0
Confuser	0.01	0	0.02	0	0	**0.97**

(c)

Results: Classification

Table: Confusion matrix for SVM meta-classifier.

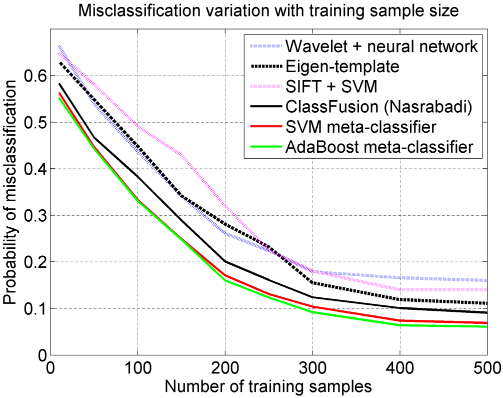
	BMP-2	BTR-70	T-72	ZIL131	D7	Other
BMP-2	**0.91**	0.03	0.02	0.02	0.03	0
BTR-70	0.01	**0.94**	0.02	0.01	0.02	0
T-72	0.03	0.02	**0.89**	0.03	0.03	0
ZIL131	0.01	0.04	0.03	**0.89**	0.03	0
D7	0	0.01	0.05	0.04	**0.90**	0
Confuser	0	0	0	0	0	**1.00**

Table: Confusion matrix for AdaBoost meta-classifier.

	BMP-2	BTR-70	T-72	ZIL131	D7	Other
BMP-2	**0.93**	0.02	0.03	0.01	0.01	0
BTR-70	0.02	**0.95**	0.02	0	0.01	0
T-72	0.04	0.02	**0.89**	0.04	0.02	0
ZIL131	0.01	0.03	0.02	**0.90**	0.04	0
D7	0	0.03	0.03	0.03	**0.91**	0
Confuser	0	0	0	0	0	**1.00**

Experiment: Performance as function of training size

- Practical concern for ATR: limited training resources
- Binary classification problem: T-72 and BMP-2 classes
- Probability of misclassification → average of false-alarm and miss.



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
- Robustness in limited training paradigm; superior asymptotic performance

[23] Srinivas et al., to appear in IEEE Int. Conf. Image Process., Sept. 2011

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
- Robustness in limited training paradigm; superior asymptotic performance
- Extension of current work: graphical-model based classification framework to exploit feature dependencies[23]

[23] Srinivas et al., to appear in IEEE Int. Conf. Image Process., Sept. 2011