Meta-classifiers for exploiting feature dependencies in automatic target recognition

Umamahesh Srinivas

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

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

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

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

ATR System description

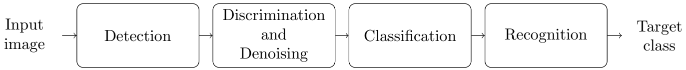


Figure: Schematic of general ATR system.

- 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

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

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

- 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

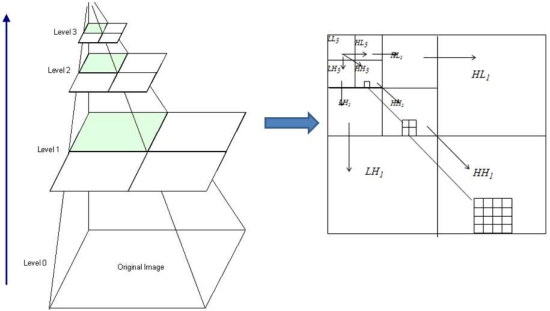
Individual classifier schemes

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[7]
- Spatial domain features
- Training class template: eigen-vector corresponding to largest singular value of training data matrix
- Correlation score decision engine

[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

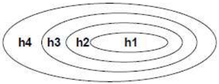
Support vector machines

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

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

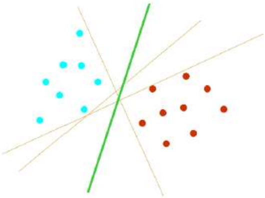
where R is the generalization error, R_{emp} is the empirical error and h is the Vapnik-Chervonenkis dimension.

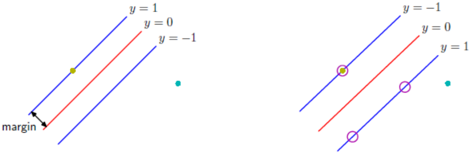
- Structural risk minimization: minimize the upper bound for the generalization error.



h1 < h2 < h3 ...

Margin maximization





Margin maximization

- Determine separating hyperplane $\mathbf{w}.\mathbf{x} + b = 0$ 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:

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

  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

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 $x_i \in S,\ y_i \in \{-1, +1\}$

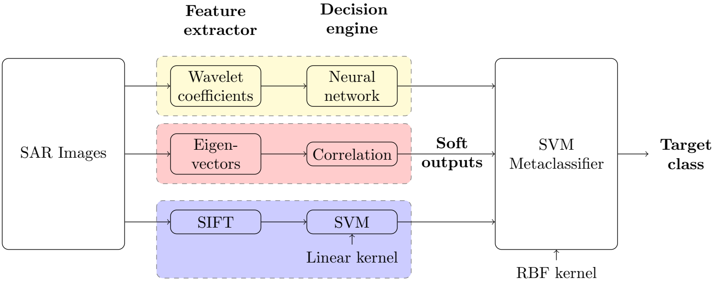
2: Initialize $D_1(i) = \frac{1}{N}, i = 1, 2, \ldots, N$

3: For $t = 1, 2, \ldots, T$:

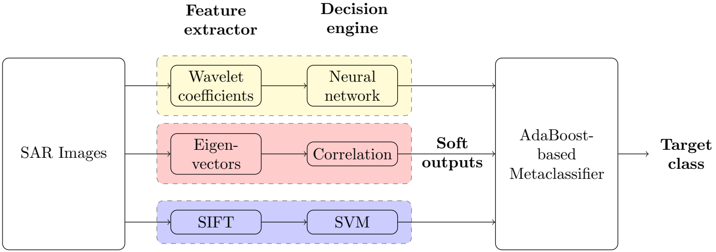
- Train weak learner using distribution D_t
- Determine weak hypothesis $f_t : S \mapsto \{-1, +1\}$ with error ϵ_t
- Choose $\beta_t = \frac{1}{2} \ln \left(\frac{1-\epsilon_t}{\epsilon_t} \right)$
- $D_{t+1}(i) = \frac{D_t(i) \exp(-\beta_t y_i f_t(x_i))}{Z_t}$, where Z_t is a normalization factor

4: Output soft decision $F(x) = \sum_{t=1}^{T} \beta_t f_t(x)$.

SVM-based meta-classification



AdaBoost-based meta-classification



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

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: The target classes used 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**

Results: Classification

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**

Results: Classification

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**

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**

Results: Classification

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**

Results: Recognition

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

	SN_C21	SN_9563	SN_9566
SN_C21	**0.71**	0.16	0.13
SN_9563	0.18	**0.68**	0.14
SN_9566	0.10	0.16	**0.74**

Results: Recognition

Table: BMP-2 Recognition: Confusion matrix for eigen-template matching classifier.

	SN_C21	SN_9563	SN_9566
SN_C21	**0.69**	0.16	0.15
SN_9563	0.19	**0.64**	0.17
SN_9566	0.11	0.18	**0.71**

Results: Recognition

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

	SN_C21	SN_9563	SN_9566
SN_C21	**0.73**	0.15	0.13
SN_9563	0.13	**0.69**	0.18
SN_9566	0.14	0.11	**0.75**

Results: Recognition

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

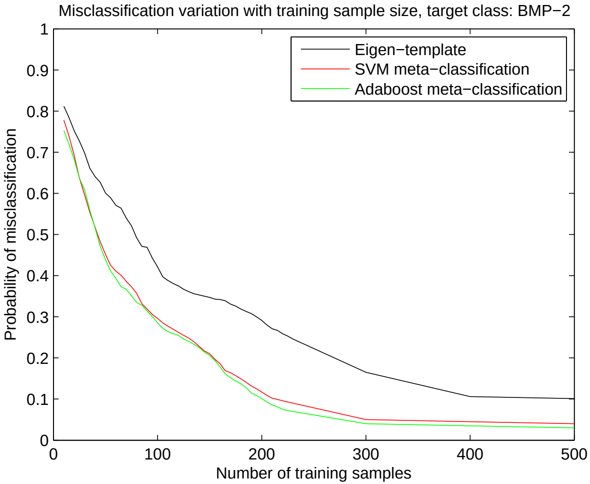
	SN_C21	SN_9563	SN_9566
SN_C21	**0.75**	0.12	0.13
SN_9563	0.13	**0.72**	0.15
SN_9566	0.08	0.13	**0.79**

Results: Recognition

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

	SN_C21	SN_9563	SN_9566
SN_C21	**0.75**	0.13	0.12
SN_9563	0.13	**0.73**	0.14
SN_9566	0.10	0.12	**0.78**

Classification rate versus training size



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