SAR target recognition using multiple views decision fusion

HUAN Ruohong¹, YANG Ruliang²

1. College of Computer Science and Technology, Zhejiang University of Technology, Zhejiang Hangzhou 310023, China; 2. Institute of Electronics, Chinese Academy of Sciences, Beijing 100190, China

Abstract: In this paper, a new synthetic aperture radar (SAR) image target recognition approach based on multiple views decision fusion is presented. Image chips are represented as feature vectors by 2-D wavelet transformation and principal component analysis algorithm. The feature vectors are classified using algorithms of support vector machine (SVM). After multiple views of the same vehicle collected at different aspects classified by SVM, the outputs are then fused using Bayesian approach and the final classification decision is generated. Experiments are implemented with three class targets in Moving and Stationary Target Acquisition and Recognition (MSTAR) Program database. Experimental results indicate that there are significant target recognition performance benefits in the probability of correct classification when three or more views are used for decision fusion. Therefore, the approach proposed is an effective method for SAR image target recognition.

Key words: synthetic aperture radar (SAR), target recognition, multiple views, decision fusion, Bayesian

CLC number: TP751.1/TP722.6  Document code: A


1 INTRODUCTION

Synthetic aperture radar (SAR) image target recognition is essential in SAR image interpretation and analysis, which is a hot issue in SAR image processing and pattern recognition. Assuming that a target is detected and its position is known in the SAR image implies that a target chip can be extracted from the SAR image of a scene for recognition. Generally, target recognition consists of two processes: feature extraction and classification. Principal component analysis (PCA) is a very effective feature extraction algorithm. First, orthonormal vector basis is obtained through singular value decomposition and feature vector analysis. Feature vectors for representing chips are gained by mapping the chips to the orthonormal vector basis, which reduces the feature vectors in dimension significantly and reduces the processing time greatly (Safari et al., 2004; Puyati et al., 2006). In the classification step, multi-class support vector machine (SVM) is often used as classifier. SVM establishes the optimal classification surface in feature classes, therefore, has excellent classification performance (Zhao & Principe, 2001; Lee et al., 2003; Safari et al., 2004).

SAR images are highly sensitive to target aspect, due to shadowing effects, interaction of the signature with the environment, projection of a three dimensional scene onto a slant plane and other reasons due to the aspect dependence of radar cross-sections (O’Sullivan et al., 2001; Brown, 2003). The ability to discriminate between targets in SAR imagery also varies greatly with target aspect. Therefore, we consider the exploitation of multiple views of a target may provide more robust classification performance than only using single view and the number of images needed to significantly improve performance. To solve these questions, Ettinger and Snyder (2002) have proposed a multi-look fusion method on hypothesis layer. Brown (2003) has proposed robust classifiers based on Bayesian approach.

This paper presents a new SAR image target recognition approach based on multiple views decision fusion. Image chips are represented as feature vectors by 2-D wavelet transformation and principal component analysis algorithm. The feature vectors are classified using algorithms of support vector machine. After multiple views of the same vehicle collected at different aspects classified by SVM, the outputs are then fused using Bayesian approach and the final classification decision is generated. Experiments are implemented for verification and analysis with three class targets in MSTAR database.

2 SAR IMAGE TARGET RECOGNITION APPROACH

The SAR image target recognition approach proposed in this paper consists of four steps as shown in Fig. 1, which are pre-processing, feature extraction, SVM classification and Bayesian decision fusion.
2.1 Data preparation

In this paper, SAR chips included in Moving and Stationary Target Acquisition and Recognition (MSTAR) Program database are used. The publicly released portion of the MSTAR database contains SAR images of 10 class vehicles. In this paper, we use three types of vehicles which are BMP2, BTR70, and T72. Each of the targets has views at 15° and 17° depression angles. The data in depression 17° are used for training and the other for testing. There are about 190—300 different aspect versions of each target at each depression angle. Table 1 lists type and sample number of training and testing set. Fig. 2 depicts original SAR target images at different aspects. Fig. 3 depicts multiple views of a target.

<table>
<thead>
<tr>
<th>Training set</th>
<th>Sample number</th>
<th>Testing set</th>
<th>Sample number</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMP2_c21</td>
<td>233</td>
<td>BMP2_c21</td>
<td>196</td>
</tr>
<tr>
<td>BTR70_c71</td>
<td>233</td>
<td>BMP2_9563</td>
<td>195</td>
</tr>
<tr>
<td>T72_132</td>
<td>232</td>
<td>BMP2_9566</td>
<td>196</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BTR70_c71</td>
<td>196</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T72_132</td>
<td>196</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T72_812</td>
<td>195</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T72_s7</td>
<td>191</td>
</tr>
</tbody>
</table>

2.2 Preprocessing

Some feature extraction algorithms and classification algorithms are sensitive to location shift, rotation, and non-uniform illumination (Sandirasegaram & Englisth, 2005). So, preprocessing is necessary. In this paper, we first rotate each target to a vertical orientation using ground truth information to bring the targets into a standardized target orientation. Then, highest energy reflecting point of the target chip is found and located to the centre of a new chip, the size of which is 64 pixels by 64 pixels. The final preprocessing step is to normalize the target chips. Normalization alters the pixel values such that, the mean intensity is zero and the standard deviation value is one for each chip. Fig. 4 (a) and Fig. 4 (b) respectively depict the chip of target T72 before and after preprocessing. Comparing Fig. 4 (a) and Fig. 4 (b), the chip after preprocessing is clearer than before and the details are enhanced.

2.3 Feature extraction

Feature extraction is an important step in the target recognition process. Feature extraction algorithms extract unique target information or signature from each chip. 2-D wavelet transformation is used here to perform 3 levels decomposition. LL3, which contains low frequency components, is picked for future extraction. PCA is then employed. LL3, of which the size is 8×8, is represented by a 64-dimension vector. Data matrix \( X_{mn} \) is composed of those vectors from all training set, where \( m = 64 \) and \( n \) is the sample number in training set. We calculate the correlation matrix \( C = E[X_{mn}X_{mn}^T] \). The eigenvectors \( \mathbf{v}_i \) and eigenvalues \( \lambda_i \), \( i = 1,2,...,m \), are computed from the correlation matrix. The eigenvectors with the largest \( p \) eigenvalues are selected for \( p \) vector as the orthonormal vector basis of the chip database. \( p \) is decided by \( \lambda_1 + \lambda_2 + \cdots + \lambda_p \geq 0.9 \), where \( \lambda_1, \lambda_2, \cdots, \lambda_p \) are unitary largest \( p \) eigenvalues. The
transformation matrix is formed from these eigenvectors in the column manner, which is 

\[ W = [v_1 v_2 \cdots v_p]^T \]

The extracted \( p \) dimension feature vectors \( y \) of the input data \( x \), can be calculated by following equation: 

\[ y = W \cdot x \]  

Finally, we get a 24-dimension feature vector for each chip. Fig. 5 shows feature extraction process. Left is the preprocessed image, middle is LL3 after 3 levels 2-D wavelet decomposition and right is feature vector gained by PCA.

2.4 Classification

Using the vector of extracted features, the classifier must be able to correctly decide which class the target belongs to. In this paper, we use support vector machine (SVM) as classifier. SVM, as a method of learning and separating binary classes, is superior in classification performance, and has been in the spotlight for pattern recognition. The basic principle of SVM can be generalized as follow (Vapnik, 1999): mapping the data to a high-dimension Euclidean space (feature space) using a kernel function for nonlinear map-

\[ \phi: \mathbb{R}^n \rightarrow E \]

, finding the decision surface in the new feature space, using kernel function for nonlinear mapping. Therefore, arbitrary test data \( x \) can be classified by

\[ f(x) = \text{sgn} \left\{ \sum_{i=1}^{n} \alpha_i^* y_i K(x_i, x) + b^* \right\} \]  

where \( x_i \) is support vector, \( y_i \in \{ -1, 1 \} \) is class label corresponded to \( x_i \), \( K(x_i, x) \) is kernel function, \( \alpha_i^* \) is Lagrange multiplier corresponded to \( x_i \), \( b^* \) is classification threshold value and \( \text{sgn} \) is symbol function.

SVM is a binary classifier in basic. Since our goal is to identify three types of targets in MSTAR dataset, we need to extend it to multi-class classifier. We first decompose the multi-class problem into several binary problems with one-against-one scheme, and use voting rule for decision making. Gaussian kernel function

\[ K(x_i, x) = \exp \left( -\frac{|x - x_i|^2}{\sigma^2} \right) \]

is applied as the kernel function for SVM.

2.5 Bayesian decision fusion

At last step, SVM is used as a single view classifier to classify the target. Here, we apply Bayesian decision fusion method to extend single view classifier to multiple views classifier. Assuming that we have a set of \( K \) images of a target, each of which is classified into one of \( Q \) distinct classes. The output of our multi-class SVM classifier is in terms of score. In order to express the output of the classifier as the estimated posterior probability, 

\[ y_{k,q} = \left( \frac{S_{k,q}}{\sum_j S_{k,j}} \right)^n, n=3 \]

is performed for nonlinear transformation (Rizvi & Nasrabadi, 2003), where \( S_{k,q}, q = 1, 2, \ldots, Q \), \( k = 1, 2, \ldots, K \) is the \( q \) th unconstrained output (score) of \( k \) th view from the output of the SVM classifier. \( y_{k,q} \) represents the estimated posterior probability that \( k \) th image \( x_k \) belongs to the class \( q \), estimated by SVM classifier.

\[ y_k = \{ y_{k,q}; q = 1, 2, \cdots, Q \} \]

Using Bayesian rule, Eq. (2) can be expressed as:

\[ P(q \mid x_k) = \frac{P(x_k \mid q)P(q)}{P(x_k)} \]  

As the priori knowledge \( P(q) \) is unknown, suppose the possibility of belonging to various classes is equal. So,

\[ P(q) = \frac{1}{Q}, 1 \leq q \leq Q \]  

For a certain target \( x_k \), \( P(x_k) \) is a fixed constant for all the classes. So, \( y_{k,q} \) is equivalent to the likelihood probability \( P(x_k \mid q) \).

The classification decision of \( k \) th view from SVM classifier is

\[ \theta_k = \arg \max _{1 \leq q \leq Q} a_{k,q} \]  

The joint probability of all views is defined as:

\[ y_q = \prod_{k=1}^{K} P(x_k \mid q) \]  

Substituting the log-likelihood of the probabilities, the Eq. (7) can be written as

\[ a_q = \sum_{k=1}^{K} (x_k \mid q) \]  

The classification decision of \( K \) views is

\[ \theta = \arg \max _{1 \leq q \leq Q} a_q \]

Bayesian decision fusion process is shown in Fig. 6. Left are three SAR images of target BMP2_9563, of which the aspects are 85°, 115° and 145° respectively. Middle is \( y_{k,q} \) of three images after SVM classification and the class decisions for single view obtained from Eq. (6), which are class three, class one and class one respectively. We know from Fig. 6, the first view is misclassified as class three, if three views are recognized separately. Right in Fig. 6 gives \( a_q \) after three images Bayesian decision fusion and the final class decision obtained from Eq. (9), which is class one. We know from Fig. 6, the correct class of those three images can be obtained after Bayesian decision fusion. That is to say, Bayesian decision fusion can correct the error, which occurs when the first view is recognized alone.
3 EXPERIMENTAL RESULTS AND ANALYSIS

SAR signatures vary greatly with aspect, as shown previously in Fig. 3. Thus, recognition performance may also be expected to vary with target aspect. Given that this is the case, exploitation of multi-aspect images of a target should provide more robust recognition performance than only using single image. The number of multi-aspect images, the angular interval between them and the approaches used for exploitation, are all issues that affect the final recognition performance of a target.

Applying BMP2, BTR70 and T72 three class images in MSTAR, our experiments are to examine the recognition performance sensitivity of the proposed approach to the number of views and the aspect intervals. Probability of correct classification (PCC) is calculated via correct classification sample number dividing by total sample number, which is the most important measurement for recognition performance. Fig. 7 shows PCC in our approach using $2^5$ multi-aspect views with the aspect interval ranging from $1^\circ$ to $60^\circ$. Fig. 7 shows the more the views are used, the higher the PCC is, and it arrives at 100% when five views are used in some aspect intervals. When three or more views are used, the PCC advances significantly compared with that when two views are used. That is because more information in multi-aspect views are exploited by algorithms when more views are used, which results in higher PCC. When only two views are used, aspect interval has few effects on the PCC. When three or more views are used, in a small aspect interval, which may be $20^\circ$ approximately, the PCC increases with aspect interval increasing. Out of that aspect interval, the PCC has little relationship with aspect interval.

Table 2 lists the highest PCC obtained by our approach using $1^5$ views respectively and compares them with PCC gotten by some typical recognition methods using single view. Applying our approach with single view means we make class decision just after image preprocessing, feature extraction and SVM classification. Ross et al. (1998) presented a template matching method, which formed templates by averaging training image samples in every $10^\circ$ aspect unit and used minimum distance rule for matching. Nilubol and Pham (1998) took Radon transformation to images in a number of discrete angles, constructed feature vectors by statistical variables and used hidden Markov models for recognition. Zhao and Principe (2001) presented a method using support vector machine for recognition, which implemented without feature extraction, and used SVM to image samples for classification in every $30^\circ$ aspect unit. From Table 2, we conclude the average PCC obtained from multiple views decision fusion is not only significantly higher than that obtained from several other methods using single view, but also higher than that obtained from our approach using single view.

<table>
<thead>
<tr>
<th>Class</th>
<th>BMP2</th>
<th>BTR70</th>
<th>T72</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single</td>
<td>96.25</td>
<td>99.49</td>
<td>96.22</td>
<td>96.70</td>
</tr>
<tr>
<td>Two views</td>
<td>97.44</td>
<td>100.00</td>
<td>96.91</td>
<td>97.58</td>
</tr>
<tr>
<td>Three views</td>
<td>99.83</td>
<td>100.00</td>
<td>99.66</td>
<td>99.78</td>
</tr>
<tr>
<td>Four views</td>
<td>100.00</td>
<td>100.00</td>
<td>99.66</td>
<td>99.85</td>
</tr>
<tr>
<td>Five views</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>Template matching (Ross et al.)</td>
<td>82.79</td>
<td>93.37</td>
<td>94.50</td>
<td>89.30</td>
</tr>
<tr>
<td>HMM (Nilubol and Pham)</td>
<td>90.80</td>
<td>92.30</td>
<td>100.00</td>
<td>94.90</td>
</tr>
<tr>
<td>SVM (Zhao and Principe)</td>
<td>90.97</td>
<td>99.49</td>
<td>88.14</td>
<td>90.99</td>
</tr>
</tbody>
</table>

Fig. 7 Probability of correct classification for multiple views decision fusion
4 CONCLUSIONS

A SAR image target recognition approach based on multiple views Bayesian decision fusion was proposed in this paper. The recognition performance sensitivity of the proposed approach to the number of views and the aspect intervals was analyzed. Experimental results indicated that there were significant target recognition performance benefits in the probability of correct classification compared with some other methods using single view for recognition, when three or more views were used for decision fusion in some certain aspect intervals. Therefore, the approach proposed is an effective method for SAR image target recognition.

REFERENCES


Lee C, Park S, Chang W and Park J. 2003. Improving the performance of multi-class SVMs in face recognition with nearest neighbor rule. Proceedings of the 15th IEEE International Conference on Tools with Artificial Intelligence (ICTAI’03), Sacramento, California, USA


Sandirasegaram N and English R. 2005. Comparative analysis of feature extraction (2D FFT and wavelet) and classification (Lp metric distances, MLP NN, and HNeT) algorithms for SAR imagery. Proc. SPIE, 5808: 314—325


宦若虹 1, 杨汝良 2

1. 310023; 2. 100190

摘要: Synthetic Aperture Radar, SAR

关键词: SAR, MSTAR, PCA, SVM

1 SAR (Synthetic Aperture Radar, SAR)

2 SAR (Principal Component Analysis, PCA)

3 (Safari, 2004; Puyati, 2006)

4 SAR (Support Vector Machine, SVM)

5 SAR, MSTAR (Zhao & Principe, 2001; Lee, 2003; Safari, 2004)

6, 7, 8 (O'Sullivan, 2001; Brown, 2003)

收稿日期: 2008-12-02; 修订日期: 2009-04-01

E-mail: huanrh@gmail.com
2.1 SAR

MSTAR SAR

SAR BMP2, BTR70 T72

15° 17° 15° 17° 15° 17° 190—300

BMP2, BTR70 T72

表 1 训练样本、测试样本种类及样本数

<table>
<thead>
<tr>
<th>定量值</th>
<th>定量值</th>
<th>定量值</th>
<th>定量值</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMP2_c21</td>
<td>233</td>
<td>BMP2_c21</td>
<td>196</td>
</tr>
<tr>
<td>BTR70_c71</td>
<td>233</td>
<td>BMP2_9563</td>
<td>195</td>
</tr>
<tr>
<td>T72_132</td>
<td>232</td>
<td>BMP2_9566</td>
<td>196</td>
</tr>
<tr>
<td>BTR70_c71</td>
<td>196</td>
<td>T72_132</td>
<td>196</td>
</tr>
<tr>
<td>T72_812</td>
<td>196</td>
<td>T72_s7</td>
<td>191</td>
</tr>
</tbody>
</table>

2.2

(a) BMP2, 320°; (b) BTR70, 62°; (c) T72, 90°

2.3

(a) BMP2, 320°; (b) BTR70, 62°; (c) T72, 90°

W = [v_1, v_2, ..., v_p]^T, x \cdot p \cdot y \cdot y = W \cdot x \cdot p \cdot y \cdot y

(a) BMP2, 320°; (b) BTR70, 62°; (c) T72, 90°
\( y_k = \{ y_{k,q} : q = 1, 2, \ldots, Q \} \)  
\[ (2) \]

\[ P(q | x_k) = \frac{P(x_k | q) P(q)}{P(x_k)} \]  
\[ (3) \]

\[ P(q) = \frac{1}{Q}, 1 \leq q \leq Q \]  
\[ (4) \]

\[ \theta_k = \arg \max_{1 \leq q \leq Q} a_{k,q} \]  
\[ (6) \]

\( \theta = \arg \max_{1 \leq q \leq Q} a_q \)  
\[ (9) \]
### Table 2: Identification Rate Comparison

<table>
<thead>
<tr>
<th></th>
<th>BMP2</th>
<th>BTR70</th>
<th>T72</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>96.25</td>
<td>99.49</td>
<td>96.22</td>
<td>96.70</td>
</tr>
<tr>
<td>2</td>
<td>97.44</td>
<td>100.00</td>
<td>96.91</td>
<td>97.58</td>
</tr>
<tr>
<td>3</td>
<td>99.83</td>
<td>100.00</td>
<td>99.66</td>
<td>99.78</td>
</tr>
<tr>
<td>4</td>
<td>100.00</td>
<td>100.00</td>
<td>99.66</td>
<td>99.85</td>
</tr>
<tr>
<td>5</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
</tr>
</tbody>
</table>

- Ross (1998): 82.79, 93.37, 94.50, 89.30
- Nilubol, Pham (1998): 90.80, 92.30, 100.00, 94.90
- Zhao, Principe: 90.97, 99.49, 88.14, 90.99

*Note:* The identification rates are given in percentage (%).
REFERENCES


Lee C, Park S, Chang W and Park J. 2003. Improving the performance of multi-class SVMs in face recognition with nearest neighbor rule. Proceedings of the 15th IEEE International Conference on Tools with Artificial Intelligence (ICTAI’03), Sacramento, California, USA


Sandirasegaram N and English R. 2005. Comparative analysis of feature extraction (2D FFT and wavelet) and classification (Lp metric distances, MLP NN, and HNet) algorithms for SAR imagery. Proc. SPIE, 5808: 314—325
