

LETTER

A Modified AdaBoost Algorithm with New Discrimination Features for High-Resolution SAR Targets Recognition

Kun CHEN^{†a)}, *Nonmember*, Yuehua LI^{†b)}, *Member*, Xingjian XU[†], and Yuanjiang LI^{††}, *Nonmembers*

SUMMARY In this paper, we first propose ten new discrimination features of SAR images in the moving and stationary target acquisition and recognition (MSTAR) database. The Ada_MCBBoost algorithm is then proposed to classify multiclass SAR targets. In the new algorithm, we introduce a novel large-margin loss function to design a multiclass classifier directly instead of decomposing the multiclass problem into a set of binary ones through the error-correcting output codes (ECOC) method. Finally, experiments show that the new features are helpful for SAR targets discrimination; the new algorithm had better recognition performance than three other contrast methods.

Key words: synthetic aperture radar (SAR), automatic target recognition (ATR), adaptive boosting, high-resolution

1. Introduction

The ability to detect targets, discriminate targets and recognize targets on day/night has long made radar systems a key sensor in many military and civilian applications. As an important aspect of SAR application, SAR automatic target recognition (ATR) has gained increasing attention over the last two decades by the radar automatic target recognition (RATR) community [1]. The first step in a typical SAR ATR system is detection, with the purpose of selecting the potential region of interest (ROI). Then, in the discrimination phase, the ROI is processed to remove the clutter false alarms (CFA) and output more accurate target clips. Feature extraction, a key step, can reduce the dimensionality of image chips greatly, extract the effective discrimination features and improve the recognition efficiency. The extracted features are expected to have the properties of effectiveness, robustness and feasibility with tolerable computational complexity. Two approaches are generally employed: select the features from the existing features and extract new features. Finally, targets are recognized by the classifiers according to the features' combination.

Generally, feature extraction methods are categorized as either linear and nonlinear. Principal component analysis (PCA) and linear discrimination analysis (LDA) [2] are two linear methods. Nonlinear methods included the ker-

nel method and the manifold learning method, such as support vector machine (SVM) [3] and locally linear embedding (LLE) [4]. Concerning Boost, the literature reported many success of Boost algorithm for pattern recognition, including Ada_Boost, Logit_Boost, Grad_Boost and Taylor_Boost [5]. All of these are effective techniques for combining multiple weak classifiers to produce a highly accurate ensemble classifier.

In this paper, we use a novel loss function for the Ada_Boost algorithm to accomplish the multiclass recognition problem directly, instead of decomposing the multiclass into a set of binary ones by the error correcting output codes (ECOC) method [6]. The result of the method will converge to a global optimum and has an exponential decrease of the training error upper bound with the increase of the iteration number. Moreover, we extract ten new features that reflect the contrast difference between the target area and CFA for target discrimination. Extensive experiments on the MSTAR database show that the performance of our method outperforms the other methods in SAR target recognition, when utilizing the new features in combination with the existing features.

2. Features for Target Discrimination

In our approach, the images in MSTAR database are represented by two types of features; we named them as the new features and the classical features. Before the extraction of new features, the classical features should be selected first to obtain the useful discriminatory features of target. The targets in the MSTAR database have the randomly distributed poses; eliminating variations of the target pose can significantly reduce the classification error. Therefore, the first feature we selected is the pose of the targets through pose estimation method used in [7]. For majority of the SAR images, the pose estimation error is within $\pm 5^\circ$. The other existing features are selected, 10 features from [8] and 12 features from [9].

To make the feature set as complete as possible, we proposed some new features to help the target discrimination. The SAR images with randomly distributed pose, equivalently, there exist rotation of target in the images. Although we have the pose estimation, the error still exists. In image processing, the geometric invariant moments (GIM) can be represented as important characters of the object, they have the invariant properties of rotational, translational and scale, we can use these features as the target discrimi-

Manuscript received April 16, 2015.

Manuscript revised June 25, 2015.

Manuscript publicized July 21, 2015.

[†]The authors are with School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing, China.

^{††}The author is with Institute of Electronic and Information, Jiangsu University of Science and Technology, Zhenjiang, China.

a) E-mail: kitchen2016@sina.cn

b) E-mail: nlglyh2013@sina.cn (Corresponding author)

DOI: 10.1587/transinf.2015EDL8090

nation features. From the pixels in ROI, the following new features were extracted.

Assume that in the ROI area D , the grayscale distribution is $f(x, y), (x, y) \in D$, and the grayscale out of the area D is zero. Respectively, the $p + q$ order origin moments and central moments are defined as:

$$m_{pq} = \sum_{y=1}^N \sum_{x=1}^M x^p y^q f(x, y), p, q = 0, 1, 2 \dots \quad (1)$$

$$\mu_{pq} = \sum_{y=1}^N \sum_{x=1}^M (x - \bar{x})^p (y - \bar{y})^q f(x, y), p, q = 0, 1, 2 \dots \quad (2)$$

Where N and M are the height and width of ROI and $\bar{x} = \frac{m_{10}}{m_{00}}, \bar{y} = \frac{m_{01}}{m_{00}}$. So the normalized center moment is defined as

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^r} \quad (3)$$

Where $r = \frac{p+q+2}{2}, p + q = 2, 3, \dots$. Then, use the above define, we can build the following seven GIM features.

$$f_1 = \eta_{20} + \eta_{02} \quad (4)$$

$$f_2 = (\eta_{20} + \eta_{02})^2 + 4\eta_{11}^2 \quad (5)$$

$$f_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \quad (6)$$

$$f_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \quad (7)$$

$$f_5 = (\eta_{30} - \eta_{12})(\eta_{30} + \eta_{12})((\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2) + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03}) * (3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2) \quad (8)$$

$$f_6 = (\eta_{20} - \eta_{02})((\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2) + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \quad (9)$$

$$f_7 = (3\eta_{21} + \eta_{03})(\eta_{30} + \eta_{12})((\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2) + (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03}) * (3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2) \quad (10)$$

Seven GIM features can keep the translation, scaling and rotation invariance if the ROI is continuous, they are the famous Hu moments in the image processing field. The last three features are the affine invariant moments (AIM), which are the rotational inertia of ROI. They are defined as follows.

$$f_8 = (\eta_{20}\eta_{02} - \eta_{11}^2)/\eta_{00}^2 \quad (11)$$

$$f_9 = (\eta_{30}^2\eta_{03}^2 - 6\eta_{30}\eta_{21}\eta_{12}\eta_{03} + 4\eta_{30}\eta_{12}^3 + 4\eta_{03}\eta_{21}^3 - 3\eta_{12}^2\eta_{21}^2)/\eta_{00}^{10} \quad (12)$$

$$f_{10} = (\eta_{20}(\eta_{21}\eta_{03} - \eta_{12}^2) - \eta_{11}(\eta_{30}\eta_{03} - \eta_{21}\eta_{12}) + \eta_{02}(\eta_{30}\eta_{12} - \eta_{21}^2))/\eta_{00}^7 \quad (13)$$

All the ten new features and the selected 23 classical features are used in experiments in Sect. 4.

3. Modified Adaptive Boost Algorithm

Existing boost algorithms for multiclass classification

mainly focus on linear combination of weak learners, which may be insufficient to produce an accurate classifier. In this section, an adaptive multiclass boost algorithm that can learn a more complicated combination of weak learners is presented in detail.

Let assume that the labeled dataset be denoted as $(X, C) = \{(x_1, c_1), \dots, (x_i, c_i), \dots, (x_n, c_n)\}, x_i \in R^d, c_i \in \{1, \dots, k\}$ denotes the class label, the data x_i are independently and identically distributed (IID), so the objective function is to learn an optimal mapping $f(x) : X \rightarrow \{1, \dots, k\}$ from the training dataset and a class label c can be assigned to a new input x . Note, in the binary classification, the class labels are ± 1 . However, in the multiclass classification, we need to recode the class label c into a vector y , usually, a set of k distinct unit words $Y = \{y_1, \dots, y_k\}$ were built and each class label k can be mapped into a code-word $y_k \in R^{k-1}$ to identify the class label. Let $f(x) \in R^{k-1}$ be a classifier, the margin of $f(x)$ with respect to class k can be defined as following like in [5], [10].

$$F(x) = \arg \max_k m(f(x), y_k) = (\langle f(x), y_k \rangle - \max_{l \neq k} \langle f(x), y_l \rangle) / 2 \quad (14)$$

Where $\langle a, b \rangle$ denotes the inner product of a and b . therefore, $F(x)$ can find a class which has the largest margin for the classifier $f(x)$, and then we just need to find a optimal classifier $f(x)$ with minimizes the classification risk below.

$$R_L(f(x)) = E_{X,Y}\{L(y, f(x))\} \approx \sum_{i=1}^N L(y^i, f(x_i)) \quad (15)$$

Where $L(a, b)$ is a multiclass loss function, we define it as $L(y, f(x)) = \sum_{i=1}^K \log[1 + \exp(-\langle f(x), y - y_i \rangle)]$. In the general boost algorithm the optimal classifier $f(x)$ is approximated as a linear combination of weak learners, that is to say, $f(x)$ is one kind of the linear combinations of weak learners $q_j(x) : X \rightarrow R^{k-1}$, but in this paper, we use a more complicated combination: the sum of Schur product $\Omega_Q = \{p(x) | p(x) = \sum_j q_{j,1}(x) \otimes \dots \otimes q_{j,m}(x), q \in Q\}$, where the \otimes denotes the Schur product, $Q = \{q_1(x), \dots, q_m(x)\}$ is the set of all multiclass weak learners $q(x) : X \rightarrow R^{k-1}$. It easy to prove that the functional space Ω_Q is a convex set, so the following optimization problem is a convex optimization problem and the risk can achieve the global minimum.

$$\begin{cases} \min_{f(x)} R_L(f(x)) \\ s.t \quad f(x) \in \Omega_Q \end{cases} \quad (16)$$

After t iterations, the classifier $f(x)$ is assumed to be the form of $f^t(x) = \sum_{j=1}^S p_j^t(x)$, S is the number of Schur product of $q(x)$, we define $p_j^t(x)$ as the form in Eq. (17).

$$p_j^t(x) = q_{j,1}(x) \otimes \dots \otimes q_{j,m_j}(x), m_j \in N \quad (17)$$

In the iteration process of the boost algorithm, each term can be updated by a new weak learner, $p_j^{t+1}(x) = p_j^t(x) \otimes q(x)$, so

the updated classifier can be achieved.

$$\begin{aligned} f^{t+1}(x) &= \sum_{i \neq j} p_i^t(x) + p_j^t(x) \otimes q(x) \\ &= \Theta_j^t(x) + p_j^t(x) \otimes q(x) \end{aligned} \quad (18)$$

Where $\Theta_j^t(x) = f^t(x) - p_j^t(x)$, around the point $\Theta_j^t(x)$, the first and second order functional derivatives of the risk $R_L(f^{t+1}(x))$ with respect to the above update in $f^t(x)$ are

$$\begin{aligned} \delta R_L(f^t; g, j) &= \frac{\partial R_L(\Theta_j^t(x) + \varepsilon p_j^t(x) \otimes q(x))}{\partial \varepsilon} \\ &= \frac{\partial}{\partial \varepsilon} \sum_{i=1}^N L(y_i, Q_j^t(x_i) + \varepsilon p_j^t(x) \otimes q(x)) \\ &= - \sum_{i=1}^N \langle q(x_i), O_i \rangle \end{aligned} \quad (19)$$

$$O_i = \sum_{k=1}^K (p_j^t(x) \odot (y_i - y^k) \frac{\exp(-\langle Q_j^t(x_i), y_i - y^k \rangle)}{1 + \exp(-\langle Q_j^t(x_i), y_i - y^k \rangle)}) \quad (20)$$

$$\begin{aligned} \delta^2 R_L(f^t; g, j) &= \frac{\partial^2 R_L(\Theta_j^t(x) + \varepsilon p_j^t(x) \otimes q(x))}{\partial \varepsilon^2} \\ &= \frac{\partial}{\partial \varepsilon^2} \sum_{i=1}^N L(y_i, Q_j^t(x_i) + \varepsilon p_j^t(x) \otimes q(x)) \\ &= \sum_{i=1}^N \sum_{k=1}^K \langle q(x_i), p_j^t(x) \otimes (y_i - y^k) \rangle^2 \Phi_{i,k} \end{aligned} \quad (21)$$

$$\Phi_{i,k} = \frac{\exp(-\langle Q_j^t(x_i), y_i - y^k \rangle)}{(1 + \exp(-\langle Q_j^t(x_i), y_i - y^k \rangle))^2} \quad (22)$$

To each j , using the gradient descent method, there also can use the Newton method as the optimization strategy; in the algorithm we use the gradient descent method. The former brings the greatest decrease of the risk, we can obtain the best weak learner with Eq. (19) and Eq. (21). Moreover, we can obtain the optimal step size as Eq. (24).

$$q_j^* = \arg \min_{q \in Q} \delta R_L(f^t(x); q(x)) \quad (23)$$

$$\alpha_j^* = \arg \min_{\alpha \in R} R_L(Q_j^t + \alpha p_j^t \odot q_j^*) \quad (24)$$

Table 1 Ada_MCBBoost algorithm

Input: dataset (X, C) , the number of classes K , a set of K distinct unit codewords Y , multiclass loss function $L(a, b)$, and the number of iterations T .

Output: $F(x) = \arg \max_k \frac{\delta(\langle f^T(x), y^k \rangle)}{\sum_{k=1}^K \delta(\langle f^T(x), y^k \rangle)}$

Algorithm:

Initialization: set $t = 0$, $S = 0$, and $f^t(x) = 0$.

Do

For $j = 1$ to S

Find the greatest decrease direction of the risk q_j^* by using the method used in [10] through the Eq. (19) and Eq. (21), Then use the Eq. (24) to get the optimal step size α_j^* . Compute the update risk $R_L^j(f^{t+1}(x))$ via Eq. (25).

end

Set $\hat{j} = \arg \min_j R_L^j(f^{t+1})$, $j \in \{0, \dots, S\}$ and then calculate

$p_{\hat{j}}^{t+1}(x) = \alpha_{\hat{j}}^* p_{\hat{j}}^t(x) \odot q_{\hat{j}}^*(x)$.

If $j \neq \hat{j}$, update $p_j^{t+1}(x) \leftarrow p_j^t(x)$

Update $f^{t+1}(x) \leftarrow \sum_{j=0}^S p_j^{t+1}(x)$ and $t = t + 1$.

While $t < T$

Hence, the updated classifier has the following risk

$$R_L^j(f^{t+1}) = R_L(Q_j^t + \alpha_j^* p_j^t \otimes q_j^*) \quad (25)$$

During each of the iteration, we calculate the optimal multiclass weak classifier, the risk and the direction which brings the greatest decrease of the classification risk [5]. The algorithm is summarized in Table 1 in brevity, we named it Ada_MCBBoost.

4. Experimental Results

In this paper, we use the SAR images in the MSTAR public release database, with 128×128 pixels and 1×1 foot resolution, to evaluate the performance of the proposed algorithm. Here, the task is to classify three distinct types of ground vehicles: BMP2s (sn-9596, sn-9566, and sn-c21), BTR70 (sn-c71) and T72s (sn-132, sn-821, and sn-s7). Every image has a different poses, which covers from the 0° to 360° aspect range randomly. The depression angles of the images are 15° and 17° . We put the SAR images at the depression angle of 17° in the training dataset and the depression angle of 15° in the testing dataset. Table 2 list the types and the size included in training and testing datasets. All the original SAR images have been preprocessed as following steps. 1) Eliminate the interference of background clutters and target shadow, obtain the ROI. 2) Use the ROI to define the binary mask matrices of the images and extract the target of SAR images by masking the binary matrices and re-center the location of the target. 3) Normalize the energy of images in the same range and execute the gray enhancement of the SAR images based on the power function. At last, we extract the classical features mentioned in Sect. 2 and the new features we proposed to compose the training and testing datasets. Before the feature extraction, each SAR image's size is cut to 64×64 pixels. In order to demonstrate the effectiveness of our proposed algorithm, we compare our algorithm with three other methods (KPCA, KLDA and NGCSE) [11]. The kernel function of KPCA and KLDA is the radial basis function (RBF), in NGCSE, we set the parameters of $k1 = 10$, $k2 = 20$, in this situation the algorithm can gets the best performance as discussed in the literature [11], set the maximum number of iterations of Ada_MCBBoost, conservatively, to 50, from our experience, approximately 20 iteration steps are enough to yield a sufficiently accurate classifier. The confuse matrix is shown in Table 3 and the best accuracy (Ba1 with the new features,

Table 2 Summary of MSTAR database

	Training Set		Testing Set	
	Serial Number	Size	Serial Number	Size
BTR70	sn-c71	233	sn-c71	196
	sn-9563	233	sn-9563	195
BMP2	sn-9566	232	sn-9566	196
	sn-c21	233	sn-c21	196
	sn-132	232	sn-132	196
T72	sn-812	231	sn-812	195
	sn-s7	228	sn-s7	191

Table 3 The confusion matrix.

Method	KPCA			KLDA			NGSCE			Ada_MCBBoost		
	BTR70	T72	BMP2	BTR70	T72	BMP2	BTR70	T72	BMP2	BTR70	T72	BMP2
BTR70	188	5	3	187	6	3	192	2	2	195	1	0
T72	14	557	11	12	558	14	5	570	7	2	576	4
BMP2	12	14	561	9	17	561	4	8	574	3	3	581

Table 4 Best accuracy, stand deviation and the time consumption of each testing sample by various algorithms

Method	Ba1(%)	Ba2(%)	Sd(%)	Tc(s)
KPCA	95.68	94.03	2.11	0.041
KLDA	95.33	94.11	1.93	0.061
NGCSE	97.88	95.12	1.15	0.085
Ada_MCBBoost	99.05	96.39	1.14	0.103

Ba2 without the new features), stand deviation (Sd) and the average time consumption (Tc) of recognizing one test samples are presented in Table 4.

As can be observed from Table 3 and Table 4, the Ada_MCBBoost achieves the highest recognition rate. However, the time consumption presented in Table 4 indicates that Ada_MCBBoost needs more time than three other algorithms. The result of KPCA and KLDA are similar, both of them use the kernel trick, which solve the linearly inseparable problem by transforming the samples into a higher or even infinite dimensional space. However, the recognition performance depends on the selection of kernel functions, if the kernel function is selected perfectly, the performance can approach to 100%. NGCSE, based on manifold learning theory, outperforms both KPCA and KLDA in recognition accuracy. A manifold structure is better than the global linear structure to the spatial distribution of the high-dimensional SAR image data. Therefore, the performance of NGCSE is superior to that of KPCA and KLDA. To the weakness of weak learners can limit the recognition performance in the high-dimensional dataset, we solve the problem through the combination of weak learners in a non-linear rather than linear way effectively. Besides, the weak learners are continually combined into complex combinations, maybe such combinations lead to the better performance. As the new algorithm depends on the iterative approach, which increases computational burden, but we still can ensure real-time performance of the algorithm, seeing in the column of Tc in Table 4, this is the main drawback of our method. Compare the second column with the third column in Table 4, we come to the conclusion that the new features are useful features to target discrimination, the best accuracy raises about 1.5% to the KPCA and KLDA, 2.5% to the NGCSE and Ada_MCBBoost, after using the new discrimination features.

5. Conclusions

In this paper, ten new discrimination features of the SAR images were given. Based on the new and classical features, an algorithm to recognize the multiclass SAR target in the MSTAR dataset was used. In the algorithm, we use a new

large-margin loss function to solve the convex optimization problem and design the multiclass classifier directly, the Schur product of the weak learners replaces the linear combination of weak learners perfectly. Experiments on the MSTAR dataset demonstrate the effectiveness of our proposed method.

Acknowledgements

This work is supported partly by Postdoctoral Research Fund Plan of Jiangsu province under Grants 1302027C and 2014 innovation project of Jiangsu province KYLX_0369. In addition, the authors wish to thank the anonymous reviewers and editors for their valuable comments and suggestions.

References

- [1] D.E. Dudgeon and R.T. Lacos, "An overview of automatic target recognition," *Lincoln Lab. J.*, vol.6, no.1, pp.3–10, 1993.
- [2] A.K. Mishra, "Validation of PCA and LDA for SAR ATR," in *TENCON-IEEE Region 10 conference*, Hyderabad, India, 2008, 1–6.
- [3] Q. Zhao and J.C. Principe, "Support vector machines for SAR automatic target recognition," *IEEE Trans. Aerosp. Electron. Syst.*, vol.37, no.2, pp.643–654, 2001.
- [4] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol.290, no.5500, pp.2323–2326, 2000.
- [5] S. Wang, P. Pan, and Y. Lu, "An adaptive multiclass boosting algorithm for classification" in: *Proceeding of International joint conference on Neural Networks (IJCNN)*, pp.1159–1166, Beijing, China, 2014.
- [6] T.G. Dietterich and G. Bakiri, "Solving multiclass learning problems via error-correcting output codes," *J. Artificial Intelligence Research*, vol.2, pp.263–286, 1995.
- [7] Y. Sun, Z. Liu, S. Todorovic, and J. Li, "Adaptive boosting for SAR automatic target recognition," *IEEE Trans. Aerosp. Electron. Syst.*, vol.43, no.1, pp.112–125, 2007.
- [8] L.M. Novak, S.D. Halversen, G.J. Owirka, and M. Hiatt, "Effects of polarization and resolution on the performance of a SAR automatic target recognition system," *Lincoln Lab. J.*, vol.8, no.1, pp.49–68, 1995.
- [9] B. Bhanu and Y. Lin, "Genetic algorithm based feature selection for target detection in SAR images," *Image Vis. Comput.*, vol.21, no.7, pp.591–608, 2003.
- [10] M.J. Saberian and N. Vasconcelos, "Multiclass boosting: Theory and algorithms," *Advances in Neural Information Processing Systems (NIPS)*, Granada, Spain, Dec. 2011.
- [11] Y.L. Huang, J.F. Pei, J.Y. Yang, B. Wang, and X. Liu, "Neighborhood geometric center scaling embedding for SAR ATR," *IEEE Trans. Aerosp. Electron. Syst.*, vol.50, no.1, pp.180–192, 2014.