

Article

Robust Radar Emitter Recognition Based on the Three-Dimensional Distribution Feature and Transfer Learning

Zhutian Yang ^{1,*}, Wei Qiu ^{2,†}, Hongjian Sun ^{3,†} and Arumugam Nallanathan ^{4,†}

¹ School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150001, China

² State Key Laboratory of Urban Water Resources and Environment, Harbin Institute of Technology, Harbin 150001, China; qiuweihit@126.com

³ School of Engineering and Computer Science, Durham University, Durham DH1 3LE, UK; hongjian.sun@durham.ac.uk

⁴ Department of Informatics, King's college London, London WC2R 2LS, UK; arumugam.nallanathan@kcl.ac.uk

* Correspondence: yangzhutian@hit.edu.cn; Tel./Fax: +86-451-8641-8284

† These authors contributed equally to this work.

Academic Editor: Leonhard M. Reindl

Received: 11 December 2015; Accepted: 19 February 2016; Published: 25 February 2016

Abstract: Due to the increasing complexity of electromagnetic signals, there exists a significant challenge for radar emitter signal recognition. To address this challenge, multi-component radar emitter recognition under a complicated noise environment is studied in this paper. A novel radar emitter recognition approach based on the three-dimensional distribution feature and transfer learning is proposed. The cubic feature for the time-frequency-energy distribution is proposed to describe the intra-pulse modulation information of radar emitters. Furthermore, the feature is reconstructed by using transfer learning in order to obtain the robust feature against signal noise rate (SNR) variation. Last, but not the least, the relevance vector machine is used to classify radar emitter signals. Simulations demonstrate that the approach proposed in this paper has better performances in accuracy and robustness than existing approaches.

Keywords: radar emitter recognition; Wigner–Ville distribution; three-dimensional distribution feature; transfer learning; relevance vector machine

1. Introduction

Radar emitter recognition (RER) based on a collection of received radar signals is a specific type of identification termed specific emitter identification (SEI), which is used to distinguish different copies of the same type of radar emitter [1–4]. RER is a topic of wide interest in both military and civil applications [5,6]. In military applications, radar emitter recognition is a critical issue in radar reconnaissance systems [7]. Specifically, radar emitter recognition provides an important means of detecting hostile radar targets. In civilian applications, radar emitter recognition technology is used to suppress criminal activities by identifying navigation radars deployed on cars and ships [5]. However, the growing complexity of electromagnetic signals encountered in modern environments is a challenge for radar emitter signal recognition [7]. Traditional approaches are becoming inefficient against this emerging issue, especially when several radar emitter signals are captured [8,9]. Therefore, multi-component radar emitter recognition under a complicated noise environment is a challenge.

Many research works have been done, e.g., stochastic context-free grammar analysis [7,10], symbolic time series analysis [11–13] and time-frequency analysis [14–19]. Thereinto, time-frequency

representation (TFR) has been frequently used for the analysis, the detection and the classification of nonstationary signals [20–23]. In [24], Swiercz proposed automatic classification of linear frequency modulation (LFM) signals for radar emitter recognition using wavelet decomposition and learning vector quantization (LVQ) classifier. In [25], a radar emitter recognition approach is proposed by using a classifier combining a minimum Mahalanobis distance classifier and SVM, which can recognize the numbers of intra-pulse modulation radar emitter signals. In [26], a hybrid radar emitter recognition is proposed. In [27], a robust radar emitter recognition based on fuzzy theory is proposed, which can recognize radar emitters robustly to some extent. However, these approaches have certain limitations for recognizing radar emitters in a complicated noise environment, which necessitates further investigation of robust radar emitter recognition.

Against this background, this research aims at proposing a robust scheme for multi-component radar emitter recognition under a complicated noise environment. The remainder of this paper is organized as follows. Section 2 gives the novel radar emitter recognition approach followed by the simulation and numerical analysis in Section 3. This paper is concluded in Section 4.

2. Radar Emitter Recognition System

2.1. System Model

The multi-component radar emitter recognition under complicated noise environment is studied in this paper. A radar emitter recognition approach based on the three-dimensional distribution feature and transfer learning are proposed. The process of radar emitter recognition is shown in Figure 1.

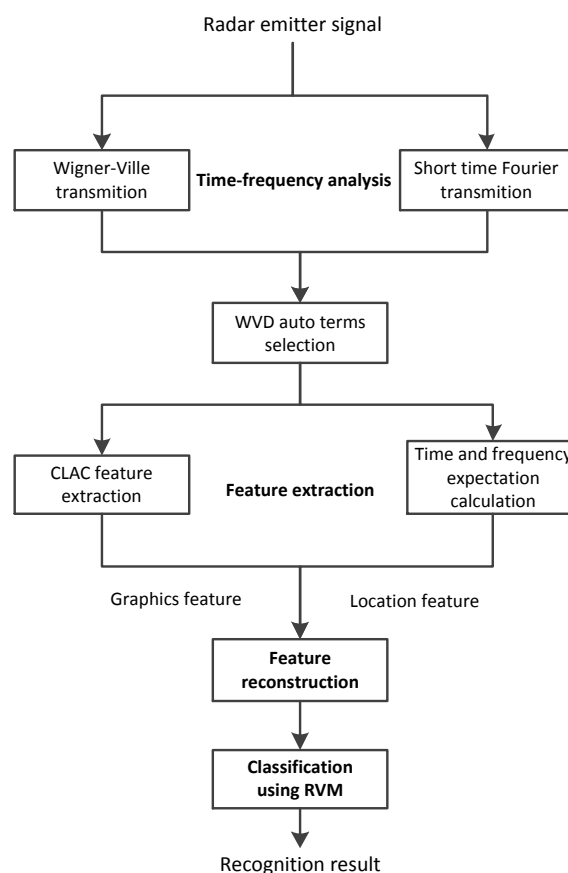


Figure 1. Model of the radar emitter recognition approach proposed in this paper.

Three functional modules are included in this research:

1. time-frequency analysis,
2. feature extraction,
3. classification.

The time-frequency analysis is a preprocess of feature extraction and has been well known as a well-developed technique [28]. In this paper, the Wigner–Ville distribution (WVD) is used to represent the characteristics of radar emitters. However, the WVD cross terms are interferences of feature extraction. Therefore, the WVD cross terms are suppressed by using a superimposition of multiple spectrograms method [29]. The auto WVDs of radar signals are detected.

In the feature extraction block, a hybrid feature is extracted, which consists of a graphics feature and a location feature. Since redundancy features are contained in the graphics features, these features would be reconstructed. Feature reconstruction is based on transfer learning.

After sorting and feature extraction, radar emitter signals are described by vectors. These describing vectors are used for classification by using RVM.

2.2. Time-Frequency Analysis of Radar Emitter Signals

WVD is a famous time-frequency distribution method, which is widely used in signal processing for its good information-preserving property. The WVD of a given signal $x(t)$ is defined by:

$$W_x(t, f) = \int_{-\infty}^{+\infty} R_x(t, \tau) \exp(-2\pi j f \tau) d\tau \quad (1)$$

$$R_x(t, \tau) = x\left(t + \frac{\tau}{2}\right) x^*\left(t - \frac{\tau}{2}\right) \quad (2)$$

where τ denotes the dummy variable; $*$ and $W_x(t, f)$ denote the complex conjugate and the energy distribution of $x(t)$ in frequency f at time t , respectively.

However, a criticism against WVD in application is the cross term, which is caused by the presence of several signal components. In order to suppress the cross terms caused by multi-component signals and noise, a suppression method by using the superimposition of multiple spectrograms is adopted. The spectrogram of the input signal can be obtained by the square of the short time Fourier transformation. Let $f(t)$ be the input signal sampling value, namely $f(t) = x(n\Delta t)$, where $0 \leq n < N$. Therefore, in the case of the discrete signals, the short time Fourier transformation is given by:

$$S_{f, L_t, \beta}(n, k) = \sum_{m=0}^{N-1} f(m) g_{L_t, \beta}^*(m-n) \exp\left(-\frac{2\pi m k}{N}\right) \quad (3)$$

$$g_{L_t, \beta}(n) = \frac{1}{\sqrt[4]{\pi L_t^2}} \exp\left\{-\frac{n^2}{2L_t^2} + j\frac{\beta}{2}n^2\right\} \quad (4)$$

where $g_{L_t, \beta}(n)$ is a discrete window function, $L_t = \sigma_t / \Delta t$, $\beta = \eta \Delta t^2$.

The discrete signal spectrogram is given by:

$$D_f(n, k) = \frac{1}{I+1} \sum_{i=0}^I |S_{f, L_t, \beta}(n, k)|^2 \quad (5)$$

Auto term regions of the spectrogram can be determined by judging $D_f(n, k)$ using a threshold, *i.e.*,

$$\Omega = \{(n, k) : D_f(n, k) \geq \eta\} \quad (6)$$

where η is the threshold, which is given by:

$$\eta = \max \left\{ \mu : \sum_{D_f(n,k) \geq \mu} D_f(n,k) \geq \gamma \sum_{n=0}^{N-1} |f(n)|^2 \right\} \quad (7)$$

where γ is a factor, which means the ratio of the energy of the auto-term region against the whole WVD. It is proven by experiments that γ is between 0.7 and 0.98 [29,30], and we shall use 0.8 in this paper.

The auto-terms of WVD are obtained by determining regions, which is given by:

$$W_{auto}(n,k) = S_{\Omega}(n,k) W_f(n,k) \quad (8)$$

$$S_{\Omega}(n,k) = \begin{cases} 1, & (n,k) \in \Omega \\ 0, & (n,k) \notin \Omega \end{cases} \quad (9)$$

Then, the auto-terms of WVD are obtained. The excellent mathematical properties of WVD are inherited. The expression of the interaction among the time, frequency and energy of a radar emitter is given in the form of WVD auto-terms. The auto-terms of WVD are normalized in each dimension and then used as the input of feature extraction. The approach to obtain the WVD auto-terms is stated with more details in [30].

2.3. Feature Extraction

The information of the distribution in dimensions of time, frequency and energy is an effective characteristic to distinguish different radar emitter signals. To extract the effective information in radar emitter signal pulses, a three-dimensional distribution feature is proposed, which consists of two parts, namely, the graphics feature and the location feature. This feature is used to extract information on WVDs of radar emitter targets.

The graphics feature is based on the cubic local auto-correlation (CLAC) function. The cubic local auto-correlation function is used for feature extraction and was proposed by Otsu in 2009 [31], which is an efficient way to describe three-dimensional distributions. The WVD-based CLAC feature is proposed and adopted as a graphics feature in this paper. A brief introduction of WVD-based CLAC feature is as follows.

Let $g(\mathbf{r})$ be three-way (cubic) data defined on the region $D : T \times F \times E$ with $\mathbf{r} = (t, f, e)^T$, where F and E denote the frequency and energy of the radar emitter signal, respectively, and T is the time length of the time window. Then, the N -th order auto-correlation function is defined by:

$$R_N(\mathbf{a}_1, \dots, \mathbf{a}_N) = \int_{D_S} g(\mathbf{r})g(\mathbf{r} + \mathbf{a}_1) \cdots g(\mathbf{r} + \mathbf{a}_N) d\mathbf{r} \quad (10)$$

$$D_S = \{\mathbf{r} | \mathbf{r} + \mathbf{a}_i \in D \forall i\}$$

where $\mathbf{a}_i (i = 1, \dots, N)$ denote displacement vectors from the reference point \mathbf{r} . We limit $N \leq 2$ and \mathbf{a}_i to a local region of a cube centering on \mathbf{r} . A CLAC feature (vector) is made up of $R_N(\mathbf{a}_1, \dots, \mathbf{a}_N)$ with various $\mathbf{a}_1, \dots, \mathbf{a}_N$ in the local region. The computation of CLAC features is described as follows.

Firstly, Equation (10) is translated to a corresponding discrete version as Equation (11) shows,

$$R_N(\mathbf{a}_1, \dots, \mathbf{a}_N) = \sum_{t,f,e \in D_S} g(t,f,e)g(t+a_{1t},f+a_{1f},e+a_{1e}) \cdots g(t+a_{Nt},f+a_{Nf},e+a_{Ne}) \quad (11)$$

where $a_{ie}, a_{if}, a_{it} \in \{\pm \Delta r, 0\}$ and $N \in \{0, 1, 2\}$, namely the same shifting distances are exploited in t , f and e dimensions. Then, the N -th-order autocorrelation function of binary images can be regarded as counting the number of points satisfying some logical condition, namely,

$$g(\mathbf{r}) \wedge g(\mathbf{r} + \mathbf{a}_1) \wedge \cdots \wedge g(\mathbf{r} + \mathbf{a}_N) = 1 \quad (12)$$

Therefore, the N -th-order autocorrelation function can be transformed into counting the patterns characterized by the above logical statement over g . Obviously, the scan by the reference point (t, f, e) can be restricted to the “1” points, viz., $g(r) = 1$, in g . The configuration $(\mathbf{r}, \mathbf{r} + \mathbf{a}_1, \dots, \mathbf{r} + \mathbf{a}_N)$ is represented by a local mask pattern.

Let M be the reference space insisting of 27 space units, as shown in Figure 2. In M , there are three layers, namely M_{-1}, M_0, M_{+1} , and positions in each layer are termed $a, b, c, d, e, f, g, h, i$.

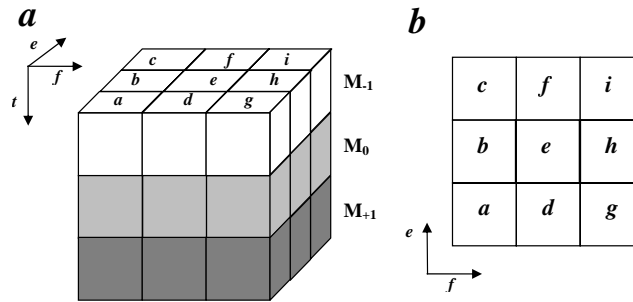


Figure 2. Model of the reference space.

Each space unit in M is given a weight, which is 2^n ($n \in [0, 1, \dots, 26]$). The convolution between M and the hyperplane of a WVD auto-term in the (t, f, e) space is carried out, which is given by:

$$C(t, f, e) = \sum_t \sum_f \sum_e M(t', f', e') W(t - t', f - f', e - e') \tag{13}$$

where $W(t, f, e)$ denotes the discrete WVD auto-term. Then, the number of each mask pattern in WVD is counted by comparing the convolution result matrix and flag values of mask patterns. The numbers of mask patterns are used as features.

Obviously, no more than three points can denote the N -th-order autocorrelation ($N = 0, 1, 2$). There are many possible mask patterns, including duplicated patterns in terms of point configurations, and these mask patterns are reduced. In the case of binary image ($f(\mathbf{r}) = 0$ or 1), $f(\mathbf{r})^k = f(\mathbf{r}), \forall k > 0$. The number of mask patterns is reduced to 251. Since the dimension of CLAC features corresponds to the number of mask patterns, we shall use the 251 dimensions features as the graphics features.

However, due to the shift-invariance of the autocorrelation, in the case that the point configuration of $(\mathbf{r}^{(1)}, \mathbf{r}^{(1)} + \mathbf{a}_1^{(1)}, \dots, \mathbf{r}^{(1)} + \mathbf{a}_N^{(1)})$ matches that of $(\mathbf{r}^{(2)}, \mathbf{r}^{(2)} + \mathbf{a}_1^{(2)}, \dots, \mathbf{r}^{(2)} + \mathbf{a}_N^{(2)})$ by shifting, $R_N(a_1^{(1)}, \dots, a_N^{(1)})$ will have the same value as $R_N(a_1^{(2)}, \dots, a_N^{(2)})$, which may lead to the confusion about the difference among targets with a similar WVD distribution. Therefore, positional information is adopted, so as to address the confusion caused by shifting in time and frequency. We locate the target WVD auto-term by computing the gravity center, which is represented by the time and frequency expectations. The time and frequency expectations of a WVD are given by:

$$\hat{t}_\Omega(t, f) = \frac{\int_{-\infty}^{+\infty} t' \Omega(t' - t, f' - f) W_x(t', f') dt' df'}{E_\Omega(t, f)} \tag{14}$$

$$\hat{f}_\Omega(t, f) = \frac{\int_{-\infty}^{+\infty} f' \Omega(t' - t, f' - f) W_x(t', f') dt' df'}{E_\Omega(t, f)} \tag{15}$$

where the E_Ω denotes the energy in region Ω .

In this paper, the gravity center position of the WVD auto-term is termed as the location feature.

2.4. Feature Learning Based on Transfer Learning

In order to characterize different radar emitter targets, the 3D feature is extracted. However, this feature has a very large number of dimensions, *i.e.*, 251 dimensions, which results in much redundant information for recognition. Therefore, the dimension reduction should be adopted. On the other hand, the variation of the signal noise rate (SNR) is an adverse effect on radar emitter recognition. Hence, a feature reconstruction approach based on transfer learning is adopted in this paper. By feature reconstruction, the robust feature against SNR variation is obtained, and the dimension of the feature is reduced, simultaneously.

Transfer learning is proposed to deal with the problem of how to reuse the knowledge learned previously from other data or features [32,33]. The idea behind transfer learning is to exploit the common knowledge between different learning tasks in order to share statistical strength and transfer knowledge across tasks [33–35]. In this paper, the learning task is to transfer common knowledge among recognitions in different noise environments.

The shared-hidden-layer autoencoder (SHLA) is an efficient approach for transfer learning [36]. It is utilized to obtain the compendious feature set with common knowledge from the original feature set. The structure of the SHLA is shown in Figure 3.

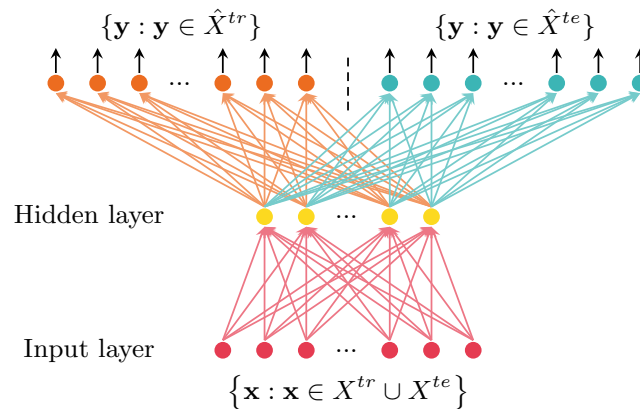


Figure 3. Illustration of the shared-hidden-layer autoencoder (SHLA) on the training set and test set.

As shown, the target values are set to be equal to the input, and the hidden representation $h(x)$ is:

$$h(x) = f(W_1x + b_1) \quad (16)$$

where $f(z)$ is a non-linear activation function, $f(z) = 1/(1 + \exp(z))$, W_1 is a weight matrix and b_1 is a bias vector.

The network output maps the hidden representation h back for the reconstruction of \tilde{x} :

$$\tilde{x} = f(W_2h(x) + b_2) \quad (17)$$

where W_2 is a weight matrix and b_2 is a bias vector.

In SHLA, same parameters for the mapping are shared between the input layer and the hidden layer. However, independent parameters are used in the reconstruction process. Let X_{tr} be the training set and X_{te} be the test set. Two objective functions are given by:

$$J_{tr}(\theta_{tr}) = \sum_{x \in \mathcal{X}^{tr}} \|x - \tilde{x}\|^2 \quad (18)$$

$$J_{te}(\theta_{te}) = \sum_{x \in \mathcal{X}^{te}} \|x - \tilde{x}\|^2 \quad (19)$$

where the parameters $\theta_{tr} = \{W_1, W_2^{tr}, b_1, b_2^{tr}\}$ and $\theta_{te} = \{W_1, W_2^{te}, b_1, b_2^{te}\}$ share the same parameters $\{W_1, b_1\}$.

Besides, the overall objective function is obtained by joining the two parts, given by:

$$J_{SA}(\theta_{SA}) = J_{tr}(\theta_{tr}) + \gamma J_{te}(\theta_{te}) \quad (20)$$

where $\theta_{SA} = \{W_1, W_2^{tr}, W_2^{te}, b_1, b_2^{tr}, b_2^{te}\}$, which will be optimized during training; γ is the hyper-parameter to control the strength of regularization.

In training, the back-propagation algorithm is adopted. The shared hidden layer makes the distribution induced by the training set similar to the distribution induced by the target set. Autoencoders find the useful features hidden in data automatically. This approach is wildly used in building a deep feature hierarchy, under the modelings of unsupervised, supervised and semi-supervised [37,38]. In this research, the reconstructed feature can be obtained from the hidden layer in SHLA for radar emitter robust recognition against SNR variation.

2.5. Classification Using RVM

The relevance vector machine (RVM), which is a sparse Bayesian modeling approach, is proposed by in [39,40]. RVM provides an approach for sparse classification. A small number of fixed basis functions are selected from a large potential candidate dictionary and weighted linearly. Therefore, a significant advantage of RVM is free from the satisfying Mercer's condition of kernel functions.

For binary classification, the likelihood function is given by:

$$P(\mathbf{t}|\boldsymbol{\omega}) = \prod_{n=1}^N \sigma\{y(x_n, \boldsymbol{\omega})\}^{t_n} [1 - \sigma\{y(x_n, \boldsymbol{\omega})\}]^{1-t_n} \quad (21)$$

where $t_n \in (0, 1)$ denotes the target value; $\boldsymbol{\omega}$ is a Gauss conditional probability, which has zero expectation and variance α_i^{-1} .

The maximum posterior probability estimation can be found by seeking the mode point of the Gaussian function, *i.e.*, μ_{MP} .

Due to:

$$P(\boldsymbol{\omega}|\mathbf{t}, \boldsymbol{\alpha}) = \frac{P(\mathbf{t}|\boldsymbol{\omega})P(\boldsymbol{\omega}|\boldsymbol{\alpha})}{P(\mathbf{t}|\boldsymbol{\alpha})} \quad (22)$$

the posterior probability according to $\boldsymbol{\omega}$ can be maximized by maximizing:

$$\begin{aligned} \log\{P(\boldsymbol{\omega}|\mathbf{t}, \boldsymbol{\alpha})\} &= \log\{P(\mathbf{t}|\boldsymbol{\omega})\} + \log\{P(\boldsymbol{\omega}|\boldsymbol{\alpha})\} - \log\{P(\mathbf{t}|\boldsymbol{\alpha})\} \\ &= \sum_{n=1}^N [t_n \log y_n + (1 - t_n) \log(-y_n)] - \frac{1}{2} \boldsymbol{\omega}^T \mathbf{A} \boldsymbol{\omega} + c \end{aligned} \quad (23)$$

where $y_n = \sigma\{y(x_n, \boldsymbol{\omega})\}$ and c is a constant.

The marginal likelihood function is given by:

$$P(\boldsymbol{\omega}|\mathbf{t}, \boldsymbol{\alpha}) = \int P(\mathbf{t}|\boldsymbol{\omega})P(\boldsymbol{\omega}|\boldsymbol{\alpha}) d\boldsymbol{\omega} \simeq P(\mathbf{t}|\boldsymbol{\omega}_{MP})P(\boldsymbol{\omega}_{MP}|\boldsymbol{\alpha}) (2\pi)^{M/2} |\boldsymbol{\Sigma}|^{1/2} \quad (24)$$

Suppose $\hat{\mathbf{t}}$ is the approximation of the Gaussian posterior distribution, with $\mu_{MP} = \boldsymbol{\Sigma} \boldsymbol{\Phi}^T \mathbf{B} \hat{\mathbf{t}}$ and $\boldsymbol{\Sigma} = (\boldsymbol{\Phi}^T \mathbf{B} \boldsymbol{\Phi} + \mathbf{A})^{-1}$. Sparse Bayesian learning can then be formulated as a *type-II maximum likelihood* procedure. The logarithm of approximate marginal likelihood function is given by:

$$\log P(\mathbf{t}|\boldsymbol{\alpha}) = -\frac{1}{2} \left\{ N \log(2\pi) + \log|\mathbf{C}| + \hat{\mathbf{t}}^T \mathbf{C}^{-1} \hat{\mathbf{t}} \right\} \quad (25)$$

where $\mathbf{C} = \mathbf{B} + \boldsymbol{\Phi} \mathbf{A}^{-1} \boldsymbol{\Phi}^T$.

Fast marginal likelihood maximization for sparse Bayesian models is exploited in this paper. More details of the fast approach can be found in [40].

3. Results and Discussion

The validity and robustness of the proposed approach will be evaluated by simulations. We investigate the performance of the proposed framework on an emulation dataset, which consists of six types of radar emitter targets generated emulationally based on the information of radars in the National Missile Defense (NMD) system [41]. In this dataset, mono-pulse radar and pulse compression radar are involved. The intra-pulse modulation types consist of linear frequency modulation (LFM), binary phase shift keying (BPSK) and quadri-phase shift keying (QPSK). The parameters of these radar emitter targets are given in detail by Table 1, where RF denotes radio frequency, PW denotes pulse width, FTR denotes frequency-time rate and CS coding scheme. The approaches proposed by Wang, L. *et al.* [27], Zhang, G. *et al.* [25] and Wu, Z. *et al.* [26] are for comparison. We note that the CLAC feature is adopted for these approaches in comparison.

Table 1. Information of radar emitter targets. PW, pulse width; FTR, frequency-time rate; CS, coding scheme; LFM, linear frequency modulation.

No.	Modulation	RF (MHz)	PW (μ s)	FTR (MHz/ μ s)/CS
1	LFM	[4890, 5050], [5240, 5370], [5510, 5630]	[0.6, 1.2]	7.8
2	Mono-pulse	[5010, 5220], [5350, 5510]	[0.2, 0.5]	-
3	BPSK	[5260, 5550]	[0.3, 0.7]	Barker (7)
4	QPSK	[5410, 5510], [5630, 5680]	[0.6, 1.1]	Frank (16)
5	LFM	[5290, 5580]	[0.3, 0.6]	0.1
6	QPSK	[5500, 5620], [5660, 5730]	[1.0, 1.4]	Frank (16)

Based on these signals, the training and test signals are generated respectively. Two hundred samples of each type radar emitter, 1200 in total, are used in training. In testing, the recognition accuracy is calculated averaged over 500 random generations of each type radar emitter. Both training and testing samples are mixed with additive white Gaussian noise. In order to evaluate the generalization performance of the recognition method, the training samples cover only samples of a 30-dB SNR, while the test samples cover samples in the range from -5 dB to 30 dB. In addition, the simulations were carried out on a personal computer with a 2.9-GHz dual-processor and 4 GB of memory.

We evaluate the results of the radar emitter signal preprocess in Figure 4. Examples for radar emitter target distributions in the time-frequency-energy space are shown. We can see that WVD cross terms have been removed by using the superimposition of multiple spectrograms and threshold judgment. Each radar target is represented by its WVD auto-term. Furthermore, the relevance among time, frequency and energy is reflected exactly by the silhouette of the WVD auto-term, which will be the basis of the three-dimensional distribution feature extraction.

However, the silhouette of a radar emitter target WVD auto-term varies as the value of Δr varies. Different values of Δr will even lead to a different feature set. Therefore, we evaluate the training accuracy of the proposed approach over different Δr . The result is shown in Figure 5. To simplify forthcoming expressions, we use the parameter m to indicate the value of Δr . It is defined by $m = \log_2 \Delta r$. Figure 5 indicates that the training accuracy reaches the peak when $m = 4$. When $m < 4$, the training accuracy increases as Δr increases. When $m > 4$, the training accuracy decreases quickly as Δr increases. This is because sampling to WVD auto-terms is too sparse to express the details of the intra-pulse characteristics of radar emitter targets.

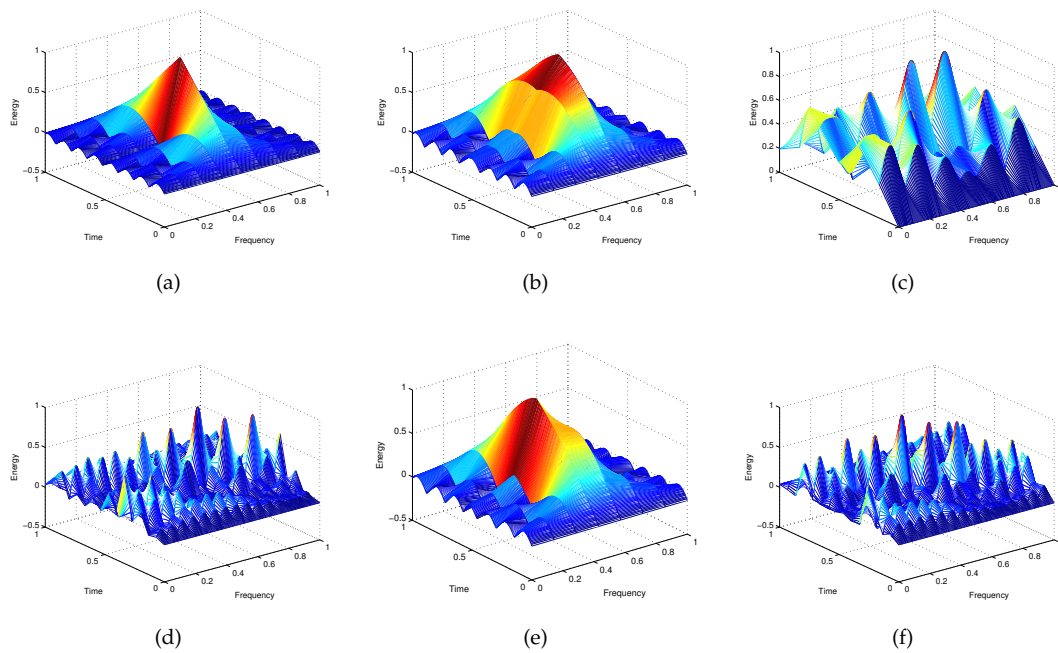


Figure 4. Normalized Wigner–Ville distribution (WVD) auto-terms of known radar emitter signals. (a) WVD auto-term of type 1 radar emitter signal; (b) WVD auto-term of type 2 radar emitter signal; (c) WVD auto-term of type 3 radar emitter signal; (d) WVD auto-term of type 4 radar emitter signal; (e) WVD auto-term of type 5 radar emitter signal; (f) WVD auto-term of type 6 radar emitter signal.

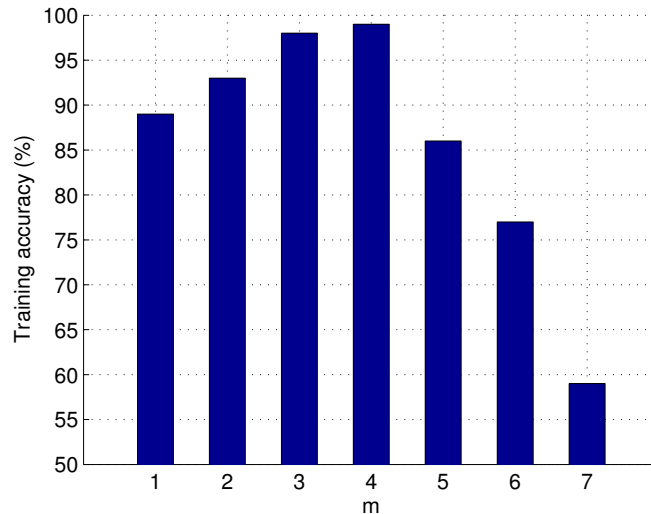


Figure 5. Training accuracy vs. m .

Next, we evaluate the performance of proposed scheme in term of the *average recognition rate* (ARR) over different sizes of the feature sets reconstructed, namely m . Actually, m is also the number of the hidden nodes of SHLA, which can influence the generation in different noise environments. In SHLA, the attempted hyper-parameter and weight decay values were the following: $\gamma \in \{0.1, 0.3, 0.5, 1, 2, 3\}$, $\lambda \in \{0.0001, 0.001, 0.01, 0.1\}$. As shown in Figure 6, the ARR increases as the size of the reconstructed feature set increases. When the size of the reconstructed feature set is 25, the training accuracy is more than 92%. Therefore, we can know that the proposed approach can get a satisfactory result to recognize the radar emitter targets with a small number of features. This result can prove the proposed approach’s validity well.

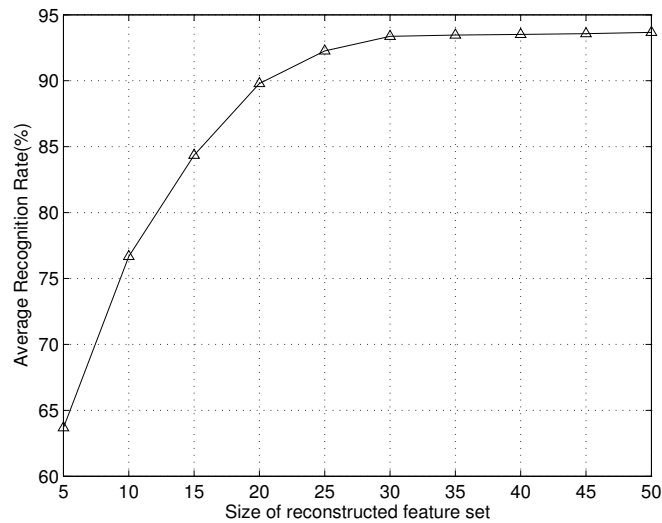


Figure 6. Recognition rate *vs.* size of reconstructed feature set m .

In Figure 7, we evaluate the ARR performance of the proposed scheme over different training set sizes. We can know that the ARRs of all approaches in comparison increase as the training size increases. When the training sample number is more than 200, all approaches obtain more than 90% average recognition rates. In the comparison, the scheme proposed in this paper has the best performance. When only 100 samples are used for training, the ARR is more than 86%. When 300 samples are used for training, the ARR reaches more than 98%. This is because the WVD-based CLAC feature proposed in this paper has a high dimension and is efficient in representing the regularities among energy, time and frequency. The high dimensional feature has a good ability to reflect the details of samples, which is good for the following feature reconstruction. In addition, RVM is efficient at learning with a small size sample set. Therefore, the proposed scheme has a good performance in ARR over the training set size.

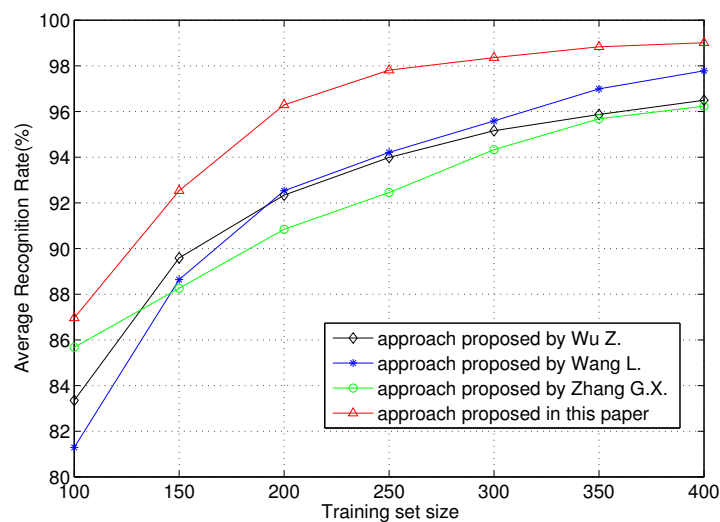


Figure 7. Average recognition rates *vs.* training set size.

In Figure 8, we evaluate the robustness performance of the proposed scheme over different SNRs. The proposed approach has a stable performance in different SNR conditions. In a low SNR environment, the proposed approach can recognize radar emitter targets effectively. When the SNR is -5 dB, the ARR is more than 80%, while those of other approaches are less than 75%. In a high SNR environment, the proposed approach has a high ARR. When SNR is 10 dB, the

ARR approach proposed in this paper reaches more than 90%. This is because, during the feature reconstruction, the common knowledge among different SNR regions is found by using transfer learning. This knowledge makes the reconstructed feature robust against SNR variation. Moreover, the ARR of the approach proposed in this paper decreases gently as SNR decreases, unlike those of other approaches. Therefore, the result indicates that the approach proposed in this paper has robustness against SNR variation.

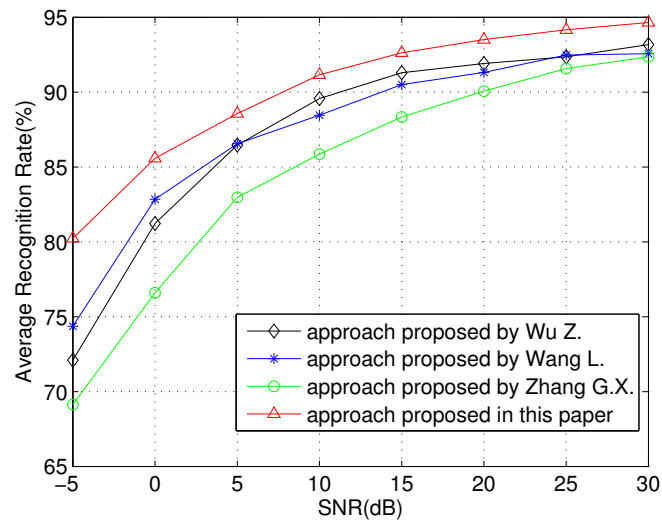


Figure 8. Recognition *vs.* signal to noise ratio.

Last, we evaluate the training time performance of the proposed scheme over different training set sizes. Because the same feature is used in all approaches, the time of feature reconstruction (feature selection) and classifier training is evaluated. As shown in Figure 9, approaches proposed by Wang L. and Zhang G.X. have good performances in training time. The approach proposed in this paper needs the most time for training. The training time increases rapidly as training set size increases. This is because SHLA used in this paper is a neural network-based transfer learning method. The high dimension of the CLAC feature goes against neural network training. Therefore, the training time performance of the proposed scheme is worse than other approaches.

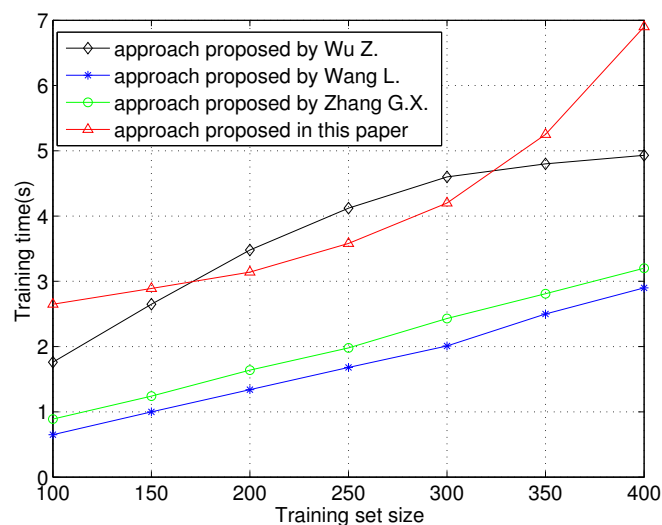


Figure 9. Training time *vs.* training set size.

4. Conclusions

In this paper, multi-component radar emitter recognition under a complicated noise environment is studied in this paper. A radar emitter recognition approach based on the three-dimensional distribution feature and transfer learning has been proposed, which can recognize the radar emitter robustly. Three functional modules are included in this approach, *i.e.*, time-frequency analysis, feature extraction and classification.

In preprocessing, the approach exploits the superimposition of multiple spectrograms represses the cross terms of WVD. In feature extraction, the three-dimensional distribution feature is proposed, which consists of two parts, *i.e.*, the graphics feature and the location feature. The three-dimensional distribution feature can represent the intra-pulse modulation information of radar emitter signals. In order to improve the robustness, the feature is reconstructed by using transfer learning, which can find the common knowledge in radar emitter signals in different noise environments. Therefore, the reconstructed feature is robust against SNR variations. RVM is applied to classify the radar emitter samples. It is proven by simulation that the proposed approach can recognize the radar emitter more accurately and robustly than existing methods. It is a new scheme for radar emitter recognition in complicated electromagnetic environments.

We admit that radar emitter recognition is subsequent processing of radar signal detection in an electronic reconnaissance system. The detection of the radar signal is not our focus in this paper. Improving training efficiency is the focus of our future works.

Acknowledgments: This work was supported by the Fundamental Research Funds for the Central Universities (Grant No. HIT.NSRIF.2015023), the Aerospace Support Technology Fund of China (2014-HT-HGD-13) and the National Natural Science Foundation of China (Grant No. 61102084).

Author Contributions: Zhutian Yang conceived the approach; Zhutian Yang and Wei Qiu designed the experiments; Hongjian Sun performed the experiments; Arumugam Nallanathan analyzed the data.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Dudczyk, J.; Kawalec, A. Specific emitter identification based on graphical representation of the distribution of radar signal parameters. *Jokull* **2013**, *63*, 408–416.
2. Dudczyk, J.; Kawalec, A. Identification of emitter sources in the aspect of their fractal features. *Bull. Pol. Acad. Sci. Tech. Sci.* **2013**, *61*, 623–628.
3. Dudczyk, J.; Kawalec, A. Fractal Features of Specific Emitter Identification. *Acta Phys. Pol.* **2013**, *124*, 406–409.
4. Kawalec, A.; Owczarek, R. Radar emitter recognition using intrapulse data. In Proceedings of the NordSec 2005-The 10th Nordic Workshop on Secure IT-Systems, Warsaw, Poland, 17–19 May 2004; pp. 444–457.
5. Ren, M.; Cai, J.; Zhu, Y.; He, M. Radar Emitter Signal Classification Based on Mutual Information and Fuzzy Support Vector Machines. In Proceedings of the International Conference on Software Process 2008, Beijing, China, 26–29 October 2008; pp. 1641–1646.
6. Davy, J. Data Modeling and Simulation Applied to Radar Signal Recognition. *Prov. Med. Surg. J.* **2005**, *26*, 165–173.
7. Latombe, G.; Granger, E.; Dilkes, F. Fast Learning of Grammar Production Probabilities in Radar Electronic Support. *IEEE Trans. Aerosp. Electron. Syst.* **2010**, *46*, 1262–1290.
8. Bezousek, P.; Schejbal, V. Radar Technology in the Czech Republic. *IEEE Aerosp. Electron. Syst. Mag.* **2004**, *19*, 27–34.
9. Zhang, L.; Cui, N.; Liu, M.; Zhao, Y. Asynchronous filtering of discrete-time switched linear systems with average dwell time. *IEEE Trans. Circ. Syst. I Regul. Pap.* **2011**, *58*, 1109–1118.
10. Zhang, L.; Shi, P.; Boukas, E.K.; Wang, C. H ∞ model reduction for uncertain switched linear discrete-time systems. *Automatica* **2008**, *44*, 2944–2949.

11. Chen, T.W.; Jin, W.D. Feature extraction of radar emitter signals based on symbolic time series analysis. In Proceedings of the 2007 International Conference on Wavelet Analysis and Pattern Recognition, Beijing, China, 2–4 November 2007; pp. 1277–1282.
12. Ma, X.; Djouadi, S.M.; Li, H. State estimation over a semi-Markov model based cognitive radio system. *IEEE Trans. Wirel. Commun.* **2012**, *11*, 2391–2401.
13. Zhang, L.; Gao, H.; Kaynak, O. Network-induced constraints in networked control systems—A survey. *IEEE Trans. Ind. Inform.* **2013**, *9*, 403–416.
14. Li, X.; Bi, G.; Ju, Y. Quantitative SNR Analysis for ISAR Imaging using LPFT. *IEEE Trans. Aerosp. Electron. Syst.* **2009**, *45*, 1241–1248.
15. Ma, X.; Olama, M.M.; Djouadi, S.M.; Charalambous, C.D. Estimation and identification of time-varying long-term fading channels via the particle filter and the EM algorithm. In Proceedings of the 2011 IEEE Radio and Wireless Symposium (RWS), Phoenix, AZ, USA, 16–19 January 2011; pp. 13–16.
16. Ma, X.; Djouadi, S.M.; Kuruganti, T.P.; Nutaro, J.J.; Li, H. Control and estimation through cognitive radio with distributed and dynamic spectral activity. In Proceedings of the 2010 American Control Conference (ACC), Baltimore, MD, USA, 30 June–2 July 2010; pp. 289–294.
17. Xing, M.; Wu, R.; Li, Y.; Bao, Z. New ISAR imaging algorithm based on modified Wigner-Ville distribution. *IET Radar Sonar Navig.* **2009**, *3*, 70–80.
18. Zhang, L.; Shi, P. Model reduction for switched LPV systems with average dwell time. *IEEE Trans. Autom. Control* **2008**, *53*, 2443–2448.
19. Jianxun, L.; Qiang, L.V.; Jianming, G.; Chunhua, X. An intelligent signal processing method of radar anti deceptive jamming. In Proceedings of the 7th International Symposium on Test Measurement, Beijing, China, 5–8 August 2007.
20. Bin, G.F.; Liao, C.J.; Li, X.J. The method of fault feature extraction from acoustic emission signals using Wigner-Ville distribution. *Adv. Mater. Res.* **2011**, *216*, 732–737.
21. Feng, L.; Peng, S.D.; Ran, T.; Yue, W. Multi-component LFM Signal Feature Extraction Based on Improved Wigner-Hough Transform. In Proceedings of the 4th International Conference on Wireless Communications, Networking and Mobile Computing, Dalian, China, 12–17 October 2008.
22. Zhang, L.; Shi, P. Stability, gain and asynchronous control of discrete-time switched systems with average dwell time. *IEEE Trans. Autom. Control* **2009**, *54*, 2192–2199.
23. Li, L.; Ji, H.B.; Jiang, L. Quadratic time-frequency analysis and sequential recognition for specific emitter identification. *IET Signal Process.* **2011**, *5*, 568–574.
24. Swiercz, E. Automatic Classification of LFM Signals for Radar Emitter Recognition Using Wavelet Decomposition and LVQ Classifier. *Acta Phys. Pol. A* **2011**, *119*, 488–494.
25. Zhang, G.; Li, X. A new recognition system for radar emitter signals. *Kybernetes* **2012**, *41*, 1351–1360.
26. Wu, Z.; Yang, Z.; Sun, H.; Yin, Z.; Nallanathan, A. Hybrid radar emitter recognition based on rough *k*-means classifier and support vector machine. *EURASIP J. Adv. Signal Process.* **2012**, *2012*, 1–9.
27. Wang, L.; Ji, H.B.; Jin, Y. Fuzzy Passive-Aggressive classification: A robust and efficient algorithm for online classification problems. *Inf. Sci.* **2013**, *220*, 46–63.
28. Haykin, S.; Bhattacharya, T. Modular learning strategy for signal detection in a nonstationary environment. *IEEE Trans. Signal Process.* **1997**, *45*, 1619–1637.
29. Li, Q.; Shui, P.; Lin, Y. A New Method to Suppress Cross-Terms of WVD via Thresholding Superimposition of Multiple Spectrograms. *J. Electron. Inf. Technol.* **2006**, *28*, 1435–1438.
30. Wu, Z.; Yang, Z.; Yin, Z.; Quan, T. A Novel Radar Detection Approach Based on Hybrid Time-Frequency Analysis and Adaptive Threshold Selection. In *Knowledge Engineering and Management*; Springer: Berlin Heidelberg, Germany 2014; pp. 79–89.
31. Kobayashi, T.; Otsu, N. Three-way auto-correlation approach to motion recognition. *Pattern Recognit. Lett.* **2009**, *30*, 212–221.
32. Kanamori, T.; Hido, S.; Sugiyama, M. Efficient direct density ratio estimation for non-stationarity adaptation and outlier detection. In Proceedings of the Twenty-Second Annual Conference on Neural Information Processing Systems, Vancouver, Vancouver, British Columbia, BC, Canada, 8–11 December 2008; pp. 809–816.

33. Deng, J.; Zhang, Z.; Marchi, E.; Schuller, B. Sparse autoencoder-based feature transfer learning for speech emotion recognition. In Proceedings of the 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII), Geneva, Switzerland, 2–5 September 2013; pp. 511–516.
34. Bengio, Y.; Courville, A.; Vincent, P. Representation learning: A review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1798–1828.
35. Bengio, Y. Deep learning of representations for unsupervised and transfer learning. *Unsuperv. Transf. Learn. Chall. Mach. Learn.* **2012**, *7*, 17–36.
36. Deng, J.; Xia, R.; Zhang, Z.; Liu, Y.; Schuller, B. Introducing shared-hidden-layer autoencoders for transfer learning and their application in acoustic emotion recognition. In Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, 4–9 May 2014; pp. 4818–4822.
37. Bengio, Y.; Lamblin, P.; Popovici, D.; Larochelle, H. Greedy layer-wise training of deep networks. *Adv. Neural Inf. Process. Syst.* **2007**, *19*, 153–160.
38. Hinton, G.E.; Salakhutdinov, R.R. Reducing the dimensionality of data with neural networks. *Science* **2006**, *313*, 504–507.
39. Tipping, M. Sparse Bayesian learning and the relevance vector machine. *J. Mach. Learn. Res.* **2001**, *1*, 211–244.
40. Tipping, M.E.; Faul, A.C. Feature extraction of radar emitter signals based on symbolic time series analysis. In Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics, Beijing, China, 2–4 November 2003.
41. Ma, J. The measurement capability and performance of the NMD-GBR radar. *Aerosp. Electron. Warf.* **2002**, *5*, 1–8.



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).