Moving Target Recognition Based on Transfer Learning and Three-Dimensional Over-Complete Dictionary

Zhutian Yang, *Member, IEEE*, Jun Deng, *Student Member, IEEE*, and Arumugam Nallanathan, *Senior Member, IEEE*

***Abstract*— In radar target recognition using high-resolution range profile, moving target recognition is a challenging issue, due to the target-aspect angle variation. To address the problem, two key issues need to be solved. First, we need to reflect the target moving status. Next, we need to find the common knowledge among different target-aspect angles. Accordingly, a novel moving target recognition based on three distribution over-complete dictionary in conjunction with transfer learning is proposed. Specifically, we propose a three distribution over-complete dictionary to represent the target and extract its moving status by dictionary learning. Moreover, we structure the feature set with generation among target-aspect angles by using a transfer learning method. This framework can be trained by using a small number of samples from limited target-aspect angles to recognize the targets of other target-aspect angles. Another advantage of this method is that it is robust against signal noise rate variation. Simulation results are presented to demonstrate the effectiveness of the proposed scheme.**

***Index Terms*— Moving target recognition, three-dimensional over-complete dictionary, transfer learning, noise robust.**

I. Introduction

RADAR automatic target recognition (RATR) is a state-of-the-art application of radars, which is a subject of wide interest in both civil and military applications [1]–[9]. Generally, three kinds of sensor information can be used for target recognition, viz. high-resolution range profile (HRRP), synthetic aperture radar (SAR) image and inverse synthetic aperture radar (ISAR) image. Due to the advent of wide band radar, it is easy to obtain HRRP of a target [2], [3]. Therefore, radar auto target recognition based on HRRP is most attractive. Since the echo returning from target scattering centers is in a form of complex on the radar line-of-sight (LOS), HRRP is a real vector composed of the amplitude of the complex of target scatterers in each range cell. Therefore, several geometric structure information can be obtained, such as scatterer distribution and target size. These information can be used to recognize the target. Therefore, radar target recognition based on HRRP has received intensive attention from the radar automatic target recognition community [3]–[5], [7], [9].

The moving target recognition is a critical issue in RATR. As discussed in [3], [7], and [9], most moving targets we concern are noncooperative. The target motion information, especially the target-aspect angle, is difficult to measure precisely, which brings great difficulties to motion compensation. At the same time, another approach for moving target recognition is to make the classification in the ISAR image domain. An obvious advantage of HRRP based RATR is that data need not be processed to form an image in training or classification phases [2], [9]. However, there also exist challenges in HRRP based RATR, due to target-aspect angle variation.

To address the problem, many researchers have made efforts recently. In [3], a scheme using dynamic system and principal component analysis (PCA) was proposed, which can recognize targets with different target-aspects effectively. In [8] and [9], the temporal dependence of multiple HRRPs in a sequence and hidden Markov models are utilized for RATR. Du *et al.* [9] proposed a moving target recognition approach by using multitask learning framework, which presents a promising result. More recently, an approach for radar HRRP target recognition was presented in [7], which combines the empirical mode decomposition (EMD) method with the nonnegative gradient projection for sparse reconstruction (NGPSR). These approaches made some achievements, but do not eliminate effects to recognition caused by aspect angle variation.

Against this background, we present a novel moving target recognition approach, which proposes a space-time-energy *three-dimensional over-complete dictionary* to describe moving targets and utilizes *transfer learning* to find the common knowledge among data of targets with different aspect angles, such that moving targets can be recognized and refrain from the influence of motion. The distinct features of this work are outlined as follows.

- In order to leverage the spatio-temporal information of HRRP sequence, a cubic space of time, distance and energy is built where the moving target can be represented in the cubic distribution of time, distance and energy. The changing of scatters' location and energy can be obtained.

Manuscript received September 14, 2015; revised May 9, 2016; accepted May 11, 2016. Date of publication May 13, 2016; date of current version June 16, 2016. This work was supported in part by the China Fundamental Research Funds for the Central Universities under Grant HIT. NSRIF. 2015023 and in part by the Chinese Defence Advance Research Program of Science and Technology under Grant 2014-HT-HGD-13. The associate editor coordinating the review of this paper and approving it for publication was Dr. Amitava Chatterjee.

Z. Yang is with the School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150001, China (e-mail: yangzt@ieee.org).

J. Deng is with the Machine Intelligence and Signal Processing Group, Technische Universität München, Munich 80333, Germany (e-mail: jun.deng@tum.de).

A. Nallanathan is with the Department of Informatics, King's College London, London WC2R 2LS, U.K. (e-mail: arumugam.nallanathan@kcl.ac.uk).

Digital Object Identifier 10.1109/JSEN.2016.2568462

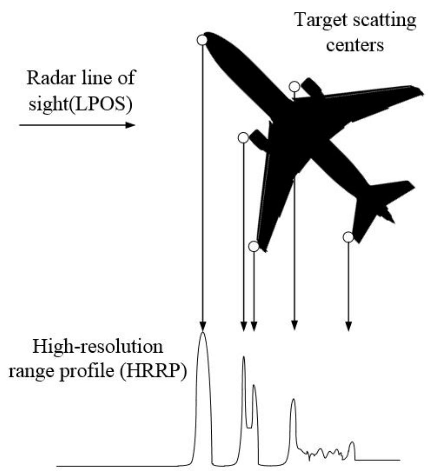

Fig. 1. An example of a plane target HRRP sample and scatters. This picture is cited from [9].

The moving and structure characteristic can be studied through the distribution.

- In this paper, sparse representation is utilized for moving target recognition. An over-complete dictionary for three-dimensional distribution sparse representation is proposed. By dictionary learning, the spatio-temporal distribution of moving targets can be can be represented exactly and decomposed sparsely.
- In order to address pose sensitivity of moving targets, the feature set is learned from the updated dictionary by using transfer learning, which can find the common knowledge among different target-aspect angles. Therefore, the features are robust against the influence by target's motion.

The remainder of the paper is organized as follows. First, the radar automatic target recognition system model with spatio-temporal property and sparse representation are presented in Section II. In Section III, next, the sparse representation based on spatio-temporal property is proposed, then the recognition based on sparse representation and transfer learning is presented. Section IV then illustrates the simulation results to demonstrate the performances of the proposed scheme. Finally, we conclude this study in Section V.

II. System Description

HRRP, a real vector, which is exploited wildly in radar target recognition, is the amplitude of the coherent summations of the complex time return from target scatterers in each range cell. The gain process of HRRP is shown in Fig. 1.

Unfortunately, target recognition on HRRP is incapable in moving target recognition. The variation of target-aspect caused by target moving will change the target statistical characteristics substantially. In this paper, we study the approach of moving target recognition based on three distribution (3D) over-complete dictionary and transfer learning. The model of moving target system is shown in Fig. 2.

The moving target recognition system can be separated into two parts, namely, training and recognition. In the former, the training targets are represented sparsely, which consists of three stages. The first stage is the mapping between the moving target information and the spatio-temporal-energy 3D distribution based on known target HRRP sequences. In the second stage, the 3D over-complete dictionary is updated based on known target to adapt the moving target recognition. The last stage is to find the key knowledge with generalization among different target-aspect angles, by using transfer learning. In the recognition, target samples with unknown target-aspect angles are decomposed based on the sparse representation model and recognized.

In this paper, the vectors of the same target due to different target-aspect angles is defined as different domains. Therefore, a domain D consists of two components: a feature space X and a marginal probability distribution $P(x)$, where $X = \{x_1, x_2, \ldots, x_n\}$. In this framework, the learning task is radar target classification, and each target is taken as a vector. X is the space of all target vectors, x_i is the ith target vector corresponding to some targets, and X is a particular learning sample. For samples in two different domains, they have the same feature spaces, but different marginal probability distributions. Given a specific domain, $D = \{X; P(X)\}$, a task consists of two components: an objective predictive function $f(\cdot)$ and a label space Y. The task can be denoted by $\tau = (Y, f(\cdot))$ and learned from the training data, which consist of pairs $\{x_i; y_i\}$, where $x_i \in X$ and $y_i \in Y$. The function $f(\cdot)$ can be used to predict the corresponding label of a new instance x. From the viewpoint of probability, $f(x)$ can be written as $P(y|x)$ [10].

We study the case where there exists one source domain D_S, and several target domains, D_T, as this is by far a bottleneck of the radar target recognition research works. More specifically, we denote the source domain data as $D_S = \left\{\left(x_{S_1}, y_{S_1}\right), \ldots, \left(x_{S_n}, y_{S_n}\right)\right\}$, where $x_{S_i} \in X_S$ is the data instance and $y_{S_i} \in Y_S$ is the corresponding class label. In our paper, D_S can be a set of term vectors from HRRP, together with their associated class labels. Similarly, we denote the target-domain data as $D_T = \left\{\left(x_{T_1}, y_{T_1}\right), \ldots, \left(x_{T_n}, y_{T_n}\right)\right\}$, where the input x_{T_i} is in X_T and $y_{T_i} \in Y_T$ is the corresponding output.

What we need to do is to find a learning task T_S, to improve the learning of the target predictive function $f_T(\cdot)$ in D_T using the knowledge in D_S and T_S, where $D_S \neq D_T$. The task is defined as a pair $\tau = \{y, P(Y|X)\}$. The condition $\tau_S \neq \tau_T$ implies that $P(Y_S|X_S) \neq P(Y_T|X_T)$. The domains are different, namely, the feature spaces between the domains are the same but the marginal probability distributions between domain data are different; i.e., $P(X_S) \neq P(X_T)$, where $X_{S_i} \in \chi_S$ and $X_{T_i} \in \chi_T$.

The details of our approach will be introduced in the following section.

III. Feature Learning and Recognition

A. Three-Dimensional Over-Complete Dictionary Based on Higher Order Autocorrelation Function

In this paper, a 3D distribution of HRRP sequence is used to describe a moving target, which will be decomposed upon a 3D over-complete dictionary. Then, the feature with generation against target-aspect angle variation will be extracted from the 3D over-complete dictionary. Therefore, the 3D over-complete dictionary will be introduced first.

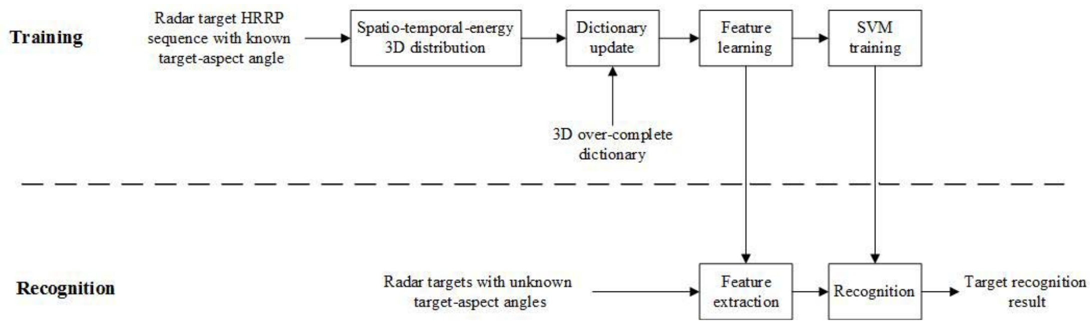

Fig. 2. The model of the proposed moving target recognition system.

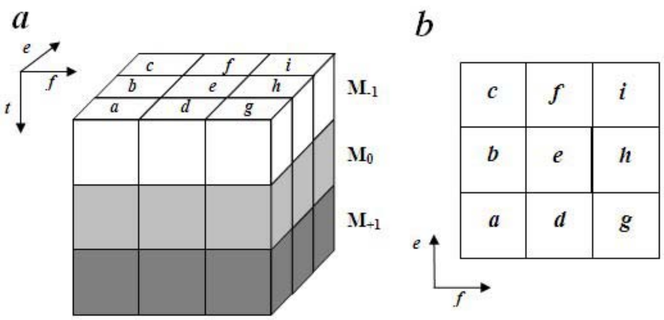

Fig. 3. The reference space model.

Recent years have witnessed a growing interest in the research for sparse representations of signals [11], [12], using an over-complete dictionary matrix $\mathbf{D} = \mathbb{R}^{d\times m}$ with m prototype signal-atoms for columns, which is donated by $\{d_j\}_{j=1}^{m}$. The representation of a signal $y \in \mathbb{R}^d$ may either be exact $y = \mathbf{D}x$, where the vector $x \in \mathbb{R}^m$ contains the representation coefficients of the signal y. In this paper, a 3D over-complete dictionary is obtained by using higher order local auto-correlation function.

Let $g(\mathbf{r})$ be three-way (cubic) data defined on the region $D_T : T \times F \times E$ with $\mathbf{r} = (t, f, e)^{\mathrm{T}}$, where F and E denote the frequency and energy of a target sample, respectively, and T is the time length of the time window. Let a_{ie}, a_{if}, a_{it} be the shifts in energy, frequency and time, respectively. Particularly, we make the restriction that $a_{ie}, a_{if}, a_{it} \in \{\pm\Delta r, 0\}$, so that the range of higher order autocorrelation function is restricted in a cube, namely, $N \in \{0, 1, \ldots, 26\}$, e.g., as shown in Fig. 3.

Then the Nth order auto-correlation function is defined by

$$R_N(\mathbf{a}_1, \ldots, \mathbf{a}_N) = \int_{D_S} g(\mathbf{r})g(\mathbf{r}+\mathbf{a}_1)\cdots g(\mathbf{r}+\mathbf{a}_N)d\mathbf{r} \quad (1)$$

$$D_S = \{\mathbf{r}|\mathbf{r}+\mathbf{a}_i \in D_T \forall i\} \quad (2)$$

where $\mathbf{a}_i (i = 1, \ldots, N)$ denote shift vectors from the reference point $\mathbf{r}$. A higher order local auto-correlation (HLAC) vector is made up of $R_N(\mathbf{a}_1, \ldots, \mathbf{a}_N)$ with various $\mathbf{a}_1, \ldots, \mathbf{a}_N$ in the local region.

Specifically, Eq. (1) is translated to a corresponding discrete version:

$$R_N(\mathbf{a}_1, ..., \mathbf{a}_N) = \sum_{t,f,e} g(t, f, e)g(t + a_{1t}, f + a_{1f}, e + a_{1e}) \cdots g(t + a_{Nt}, f + a_{Nf}, e + a_{Ne}) \quad (3)$$

Algorithm 1. Three-Dimensional Over-Complete Dictionary Based on Higher-Order Autocorrelation

Define the cubic local higher-order autocorrelation function, as (1).
for $i = 1 : d$ **do**
  Obtain the logical conditions of the i th-order cubic local autocorrelation function.
end
for $j = 1 : N_c$ **do**
  for $i = 1 : m$ **do**
    if $X(r) * A_i \neq 0$, **then**
      $N_i = N_i + 1$
    end
  end
end
Obtain the coefficient matrix **c** and **X** is decomposed as $\mathbf{X} = \mathbf{Dc}$.

In the case of binary distribution ($f(\mathbf{r})$ = 0 or 1), the scan by the reference point (t, f, e) can be restricted to the "1" points, viz., $g(r) = 1$, in g. The configuration $(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_N)$ is represented by a local 3D distributions in D_T, namely, atoms. Then, the Nth-order autocorrelation function of a binary distribution can be regarded as counting the number of points satisfying some logical condition, namely,

$$g(r) \wedge g(r + a_1) \wedge \cdots \wedge g(r + a_N) = 1 \quad (4)$$

The higher order autocorrelation function can be transformed into counting the patterns characterized by the above logical statement over g. When scanned by the reference point $\mathbf{r}$, $\mathbf{X}$ can be decomposed by using atoms in the CHLAC dictionary. The algorithm of three-dimensional over-complete dictionary based on higher order autocorrelation is shown as Algorithm 1, where $\mathbf{D} \in \mathbf{R}^{d\times m}$, $d = 26$ and m= $2^{27} - 1$. $D(i)$, N_c and N_i donate the i atom, the number of points in region $\mathbf{D}$ and the ith atom in dictionary, respectively.

B. Dictionary Learning

Dictionary learning is to find an sub-dictionary which represents the training signals best. Therefore, during dictionary learning, the dictionaries are trained to adapt to the training data. More precisely, let $\mathbf{Y}$ be the training data, where each column of $\mathbf{Y}$ corresponds to one training sample. For a given dictionary $\mathbf{D}$, the excellently learned sub-dictionary $\mathbf{D}^* \in \mathbf{R}^{m \times d}$ is the one that minimizes

$$\|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 \tag{5}$$

where is $\|\cdot\|$ the Frobenius norm.

The learned dictionary will have the potential to offer an improved performance compared with the predefined dictionary, since the salient information directly from the training data is captured by the atoms derived [12]. However, dictionary learning is a joint optimization problem and this process usually involves higher computational complexity. To decrease the computational complexity, a succinct learning approach for 3D dictionary is proposed.

In this approach, dictionary learning is achieved by an analytical way, instead of directly solving the joint optimization problem. We analyze contents of the 3D distribution of training data such that the atoms existing in training data can be found. We delete the uncorrelated atoms from the initial dictionary and obtain the dictionary learned.

First, we build a reference cube $\boldsymbol{M}$, e.g., shown in Fig. 3. In $\boldsymbol{M}$, there are three layers (i.e., M_{-1}, M_0 and M_{+1}) and positions are labeled by $a, b, c, d, e, f, g, h, i$ in each layer. Therefore, 27 units are in $\boldsymbol{M}$ in total. The unit e in M_0 layer is termed as reference point.

We give each unit (from a in M_{-1} to i in M_{-1}) a weight, which is 2^n $(n \in [0, 1, \cdots, 26])$. Therefore, each atom composing of units can be represented by a weight, which comes from the accumulation of unit weight. Execute the discrete convolution between $\boldsymbol{M}$ and training samples, namely

$$C(t, f, e) = \sum_t \sum_f \sum_e M(t', f', e') X_T \times (t - t', f - f', e - e') \tag{6}$$

where $X_T(t, f, e)$ denotes the 3D distribution of the training sample.

Then, as the reference point traverses the 3D distribution of training data, convolution results will be recorded in the training data distribution. Therefore, the frequency of each atom appearance in the training data can be calculated. The training data can be decomposed into atoms and uncorrelated atoms can be found. The size of the dictionary is reduced.

C. Feature Learning Based on Transfer Learning

In order to characterize different targets, the time-frequency-energy distribution are decomposed by using atoms in $\mathbf{D}^*$. Due to the target moving, the difference among target-aspect angles of the same target can confuse the recognition. Therefore, transfer learning method is used here is to enhance the generalization of recognition against targets with different target-aspect angles.

Transfer learning has been proposed to deal with the problem of how to reuse the knowledge learned previously from other data or features [13], [14]. The idea behind transfer learning is to exploit commonalities between different learning tasks in order to share statistical strength, and transfer knowledge across tasks [14]–[16].

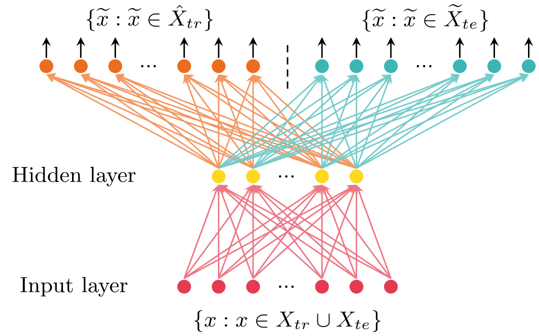

Fig. 4. Illustration of the shared-hidden-layer autoencoder (SHLA) on the training set and test set [17].

The shared-hidden-layer autoencoder (SHLA) is an efficient transfer learning method and utilized here to obtain the feature set with common knowledge from dictionary $\mathbf{D}^*$ [17]. The structure of the SHLA is shown in Fig. 4.

As shown, it sets the target values to be equal to the input, so as to automatically find common feature representations for both training data and test data in an unsupervised way. In response to an input example x, the hidden representation $h(x)$ is

$$h(x) = f(W_1 x + b_1) \tag{7}$$

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

The network output completes the reconstruction process, which takes the hidden representation h, and maps it back to a reconstruction $\widetilde{x}$:

$$\widetilde{x} = f(W_2 h(x) + b_2) \tag{8}$$

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

Note that the SHLA shares the same parameters for the mapping from the input layer to the hidden layer, but uses independent parameters for the reconstruction process. Given a training set of examples X_{tr}, and a test set of examples X_{te}, the two objective functions, are formed as follows:

$$J_{tr}(\theta_{tr}) = \sum_{x \in \chi^{tr}} \|x - \tilde{x}\|^2 \tag{9}$$

$$J_{te}(\theta_{te}) = \sum_{x \in \chi^{te}} \|x - \tilde{x}\|^2 \tag{10}$$

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 $\{W_1, b_1\}$.

In the end, the overall objective function is obtained by joining the distance for the two sets, given by:

$$J_{SA}(\theta_{SA}) = J_{tr}(\theta_{tr}) + \gamma J_{te}(\theta_{te}) \tag{11}$$

where $\theta_{SA} = \left\{W_1, W_2^{tr}, W_2^{te}, b_1, b_2^{tr}, b_2^{te}\right\}$ are the parameters to be optimized during training, the hyper-parameter γ controls the strength of the regularization. Training the SHLA is equivalent to training a basic autoencoder, and the standard back-propagation algorithm can be applied.

To minimize the objective function, the shared hidden layer is biased to make the distribution induced by the training set

TABLE I
INFORMATION OF KNOWN RADAR TARGET

Plane	Length(m)	Width (m)	Height (m)
B-52	49.5	56.4	12.40
TU-16	34.8	33	9.85
F-15	19.45	13.5	5.68
Yark-42	36.38	34.88	9.83
Cessna Citation S/II	14.40	15.90	4.57
An-26	23.80	29.20	9.83

as similar as possible to the distribution induced by the target set. This helps to regularize the functional behavior of the autoencoder.

IV. SIMULATION RESULTS AND DISCUSSIONS

A. Selected Task and Data

In the simulation, we will examine the moving target recognition rate of the proposed approach. In addition, the robust against SNR changing will be studied. In order to test the generalization performance of the recognition methods, the training data cover only one target-aspect angle of the test data, while the test data cover all target-aspect angles.

We investigate the performance of the proposed framework on two parts of data. One is the simulated HRRP data of airplanes, including B-52, TU-16 and F-15. The other is the measured HRRP data from three real airplanes, which insists of Yark-42, Cessna Citation S/II and An-26. The information of targets is shown in Table I.

Due to motion, the target-aspect angle of a target varies. Therefore, the measured data of Yark-42, Cessna Citation S/II, and An-26 cover 5, 7 and 7 target-aspect angles, respectively and the simulated data of targets insist of 6 target-aspect angles.

B. Experimental Setup and Evaluation Metrics

We evaluate the proposed approach using the simulated data and measured data, respectively. In dictionary establishing stage, a target-aspect angle of each plane is chosen randomly used to train the dictionary. After dictionary learning, the moving targets can be characterized by using atoms in the dictionary. The common knowledge which can be used in different target-aspect angles will be found among these atoms. In the decompose process, the feature set including the common knowledge is built.

In the transfer learning, SHLA is adapted. In SHLA, the attempted hyper-parameter and weight decay values were the following: $\gamma \in \{0.1, 0.3, 0.5, 1, 2, 3\}$, $\lambda \in \{0.0001, 0.001, 0.01, 0.1\}$. Because the size of learned feature set is relevant to the number of hidden units m, influence of m will be studied first. As classifier, we use linear SVMs with a fixed penalty factor C = 0.5 as the basic supervised learner.

To evaluate the proposed approach, we examine two aspects: generation in different target-aspect angles and robustness against SNR changing. Therefore, we examine the recognition rate on not only the training target-aspect data but also all target-aspect data and in different SNR circumstances. In order to evaluate the proposed approach in this paper, the approaches proposed in [3], [6], [7], and [9] are used for comparison. In [3], a dynamic system which models the short dependency between consecutive samples of HRRP in segments for moving target recognition is proposed (which is termed as DS-ARTR). In [6], a novel target recognition method termed as orthogonal maximum margin projection subspace is proposed for HRRP-based radar target recognition (which is termed as OKMMPS-ARTR). In [7], a radar target recognition approach based on sparse representation and time-frequency feature is proposed (which is termed as SP-ARTR). Moreover, for moving target recognition, a closely related learning technique to transfer learning is the multitask learning framework. Therefore, a radar target recognition approach based on multitask learning [9] is adapted in the comparison (which is termed as MT-ARTR).

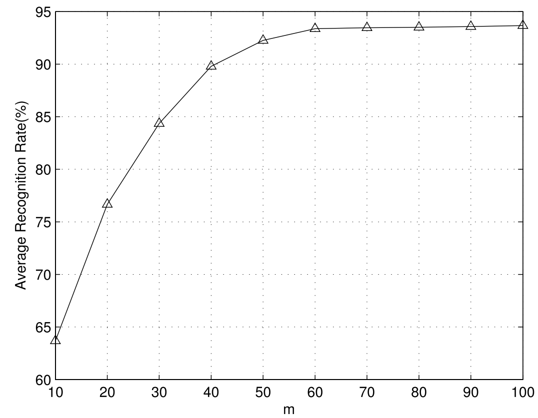

Fig. 5. Recognition rate vs. the number of hidden units m.

C. Results

At first, we evaluate the performance of proposed scheme in term of the recognition rate over different size of the feature set, namely m. Actually, m is also the number of the hidden nodes of SHLA, which can influence the generation in different target-aspect angles of moving targets.

As shown in Fig. 5, as m grows from 10 to 100, the average recognition rate increases from about 63% to 93%. This suggests that the proposed approach can get a satisfactory result to recognize the radar targets in different target-aspect angles with a small number of features. Moreover, it is a rapid increase process from $m = 10$ to $m = 60$. Especially, when $m = 60$, the average recognition receive about 92%. This result can prove the proposed approach's validity well. It also means that we can use target information in one target-aspect angle to recognize others.

To make a compromise between recognition and computing complexity, we let $m = 60$ in the following simulations.

In Fig. 6, we evaluate the *Average Recognition Ratio* (ARR) performance of the proposed scheme over different target-aspect angles using the simulated data set. We can know that the proposed approach achieve high ARRs over all target-aspect angles. Since only one target-aspect angle data are used for training, ARR on the whole test data set vary from about 90% to 98%. In the domain (target-aspect angles)

TABLE II
CONFUSION MATRIXES COMPARISON

	Multitask learning			Transfer learning		
	Yark-42	Cessna Citation S/II	An 26	Yark-42	Cessna Citation S/II	An 26
Yark-42	92.36	4.82	2.82	92.52	4.41	3.07
Cessna Citation S/II	5.59	91.10	3.31	3.74	93.04	3.22
An 26	3.28	2.92	93.80	2.92	2.53	94.55
Average recognition rate	92.42			93.37		

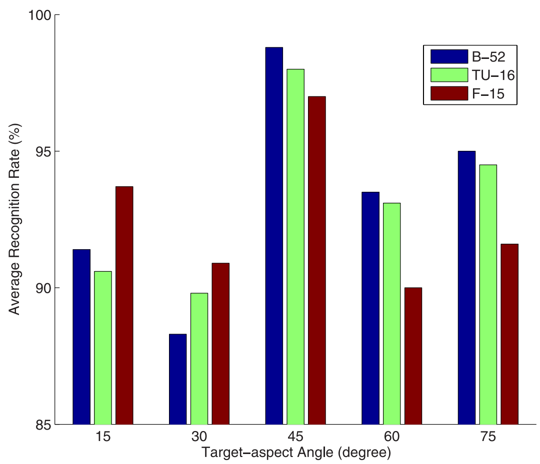

Fig. 6. Average recognition rates of the simulated target data with different TAAs.

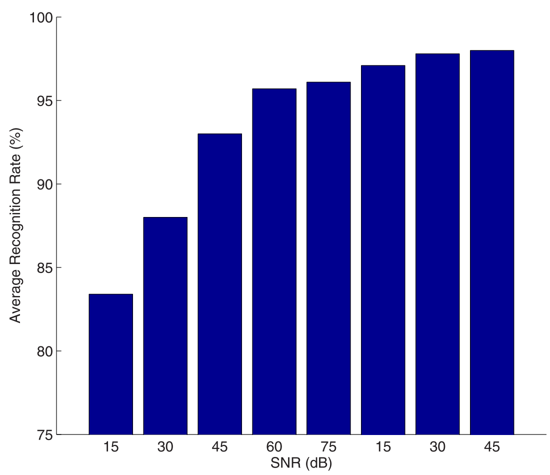

Fig. 7. Average recognition rates against different SNRs.

of 45 degree, ARR achieves the high performance, namely, around 98%. Moreover, in each domain, ARR performance is very close on three types targets. It can be proved that the common knowledge among different target-aspect angles is found through the transfer learning.

Fig. 7 shows the performance of the proposed scheme over different SNRs on the simulated data set. It can be observed that the performance of the proposed scheme have obvious increasing recognition rates, as the SNR increases. In high SNR environment, the proposed scheme can get good performance. ARR is higher than 90% with 20 dB SNR. Especially, when SNR increases to 40 dB, the recognition

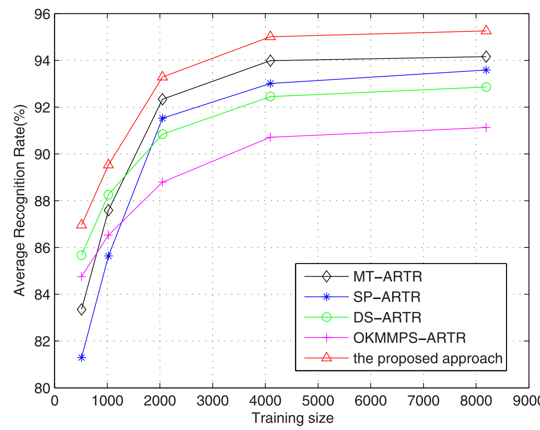

Fig. 8. Average recognition rates against different training set size.

rate is approximate to 95%. In lower SNR environment, our scheme can also have a good performance. More specifically, it has a recognition more than 80% in 10 dB SNR and 90% in 20 dB SNR, which can show that the proposed scheme has good robust against noise variation.

In Fig. 8, we evaluate the *Average Recognition Ratio* (ARR) performance of the proposed scheme over different training set size. We can know that the ARR of approaches in comparison increase as the training size increases. When training size is above 2000, all approaches obtain more than 90% recognition rates and when training size is above 8000, all approaches can get more than 92% recognition rates. The two deep learning approaches, multitask and transfer learning, have good performance in this stage. They have higher average recognition rates than other approaches. Compared with existing approaches, the proposed approach in this paper has a better performance in this comparison. More specifically, it obtains a recognition more than 93% with 2048 training samples and 95% with 4096 training samples, which reflects the validity of the proposed approach.

In moving target recognition, a similar approach with transfer learning is the framework based on multitask learning (proposed in, which is also a deep learning-based approach. Therefore, we compare the multitask learning-based scheme proposed in [9] and our scheme. A confusion matrix with 2048 training samples is given in Table II. We can know that the proposed scheme outperforms the multitask learning-based scheme in ARR performance. The ARR of the proposed approach is higher than that of the multitask learning framework by about 1.5% for the given test dataset.

In Fig. 9, we evaluate the ARR performance of the proposed scheme over different SNRs. In a low SNR condition (5 dB),

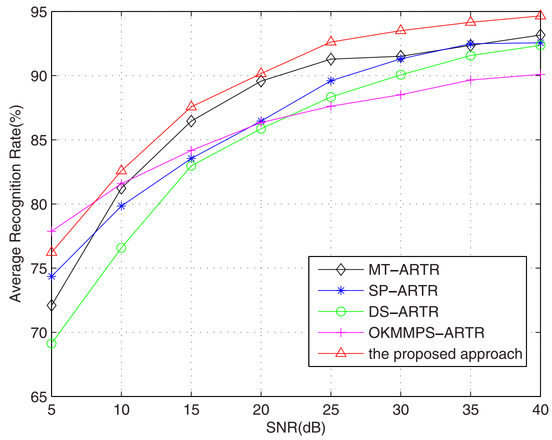

Fig. 9. Average recognition rate comparison under different SNR environments.

recognition rates of schemes have acceptable performances (close to or higher than 70%), and the scheme OKMMPS-ARTR has the highest ARR. As the SNR grows, the average recognition rates of all schemes increase obviously. Especially, the proposed scheme in this paper outperform other schemes when SNR is above 10 dB. In a high SNR condition (higher than 30 dB), most schemes have average recognition rates more than 90%. The schemes of SP-ARTR and MT-ARTR have similar performances, and converge when SNR grows to 35 dB. Obviously, the proposed scheme has the best performance in the comparison. More specifically, it has average recognition rates of 80%, 90% and 95% under 10 dB, 20 dB and 40 dB SNR conditions, respectively. Therefore, we can draw the conclusion that the proposed scheme is effective under both low and high SNR conditions.

V. Conclusions and Future Work

This paper has studied the problem of moving target recognition. We considered that the key issue of this problem is how to eliminate the affect of target-aspect angle. Therefore, we tackle the problem in two steps. First, we structured a 3D over-complete dictionary based on cubic higher-order autocorrelation function to represent moving targets. Next, we extracted the target feature by using the SHLA, which contains the common knowledge among different target-aspect angle targets. Therefore, we used limited target samples from one target-angle to train the moving target recognition system and recognized targets with different target-aspect angles. Finally, simulation results have been presented to demonstrate the performance of proposed scheme. From the simulation results, we have verified the validity of proposed scheme and the efficiency. Then we have analyzed the performance effect under SNR variation, which proves that the proposed approach has noise robustness. In our future work, we will study the shift sensibility of moving target recognition.

References

[1] H.-J. Li and S.-H. Yang, "Using range profiles as feature vectors to identify aerospace objects," *IEEE Trans. Antennas Propag.*, vol. 41, no. 3, pp. 261–268, Mar. 1993.

[2] R. Mitchell and J. J. Westerkamp, "Robust statistical feature based aircraft identification," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 35, no. 3, pp. 1077–1094, Jul. 1999.

[3] A. Ajorloo, M. Hadavi, M. H. Bastani, and M. M. Nayebi, "Radar HRRP modeling using dynamic system for radar target recognition," *Radioengineering*, vol. 23, no. 1, pp. 121–127, 2014.

[4] M. Kaleem, A. Guergachi, and S. Krishnan, "Empirical mode decomposition based sparse dictionary learning with application to signal classification," in *Proc. IEEE Digit. Signal Process. Signal Process. Edu. Meeting (DSP/SPE)*, Aug. 2013, pp. 18–23.

[5] A. Ajorloo, M. Hadavi, M. M. Nayebi, and M. H. Bastani, "Statistical modeling of consecutive range profiles for radar target recognition," in *Proc. 14th Int. Radar Symp. (IRS)*, vol. 2, Jun. 2013, pp. 608–613.

[6] D. Zhou, "Orthogonal maximum margin projection subspace for radar target HRRP recognition," *EURASIP J. Wireless Commun. Netw.*, vol. 2016, no. 72, 2016, pp. 1–11.

[7] Y. Kong and C. Wang, "Radar target recognition based on dictionary of time-frequency feature and nonnegative sparse decomposition," in *Proc. 12th Int. Bhurban Conf. Appl. Sci. Technol. (IBCAST)*, Jan. 2015, pp. 672–674.

[8] F. Zhu, X.-D. Zhang, Y.-F. Hu, and D. Xie, "Nonstationary hidden Markov models for multiaspect discriminative feature extraction from radar targets," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 2203–2214, May 2007.

[9] L. Du, P. Wang, H. Liu, M. Pan, F. Chen, and Z. Bao, "Bayesian spatiotemporal multitask learning for radar HRRP target recognition," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3182–3196, Jul. 2011.

[10] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[11] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[12] W. Dai, T. Xu, and W. Wang, "Simultaneous codeword optimization (SimCO) for dictionary update and learning," in *Proc. 49th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Sep. 2011, pp. 920–927.

[13] T. Kanamori, S. Hido, and M. Sugiyama, "Efficient direct density ratio estimation for non-stationarity adaptation and outlier detection," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 809–816.

[14] J. Deng, Z. Zhang, E. Marchi, and B. Schuller, "Sparse autoencoder-based feature transfer learning for speech emotion recognition," in *Proc. Humaine Assoc. Conf. Affective Comput. Intell. Interact. (ACII)*, Sep. 2013, pp. 511–516.

[15] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.

[16] Y. Bengio, "Deep learning of representations for unsupervised and transfer learning," *Unsupervised Transf. Learn. Challenges Mach. Learn.*, vol. 7, p. 19, 2012.

[17] J. Deng, R. Xia, Z. Zhang, Y. Liu, and B. Schuller, "Introducing shared-hidden-layer autoencoders for transfer learning and their application in acoustic emotion recognition," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2014, pp. 4818–4822.

Zhutian Yang received the M.E. degree and Ph.D. degree from the Harbin Institute of Technology, China, in 2008 and 2012, respectively. He was a Visiting Research Associate with King's College London in 2015. He is currently a Lecturer with the Harbin Institute of Technology. He has authored over ten papers in journals and international conferences. His general research interests include full-duplex radio technology, signal processing algorithms, and machine learning.

Jun Deng received the bachelor's degree in electronic and information engineering from Harbin Engineering University in 2009, and the master's degree in information and communication engineering from the Harbin Institute of Technology, Harbin, China, in 2011. He is currently pursuing the Ph.D. degree with the MISP Group, Technische Universität München, Munich, Germany. His interests are machine learning methods such as transfer learning with an application preference to emotion recognition in speech.

Arumugam Nallanathan (S'97–M'00–SM'05) served as the Head of Graduate Studies with the School of Natural and Mathematical Sciences, King's College London, from 2011 to 2012. He was an Assistant Professor with the Department of Electrical and Computer Engineering, National University of Singapore from 2000 to 2007. He is currently a Professor of Wireless Communications with the Department of Informatics, King's College London (University of London). He has co-authored over 300 papers. His research interests include 5G technologies, millimeter wave communications, cognitive radio and relay networks. He is a co-recipient of the Best Paper Awards presented at the 2007 IEEE International Conference on Ultra-Wideband (ICUWB 2007) and the IEEE International Conference on Communications 2016 (ICC 2016). He is a Distinguished Lecturer of the IEEE Vehicular Technology Society.

He served as the Chair of the Signal Processing and Communication Electronics Technical Committee of the IEEE Communications Society, the Technical Program Co-Chair (MAC track) for the IEEE WCNC 2014, the Co-Chair of the IEEE GLOBECOM 2013 (Communications Theory Symposium), the Co-Chair of the IEEE ICC 2012 (Signal Processing for Communications Symposium), the Co-Chair of the IEEE GLOBECOM 2011 (Signal Processing for Communications Symposium), a Technical Program Co-Chair of the IEEE International Conference on UWB 2011 (IEEE ICUWB 2011), the Co-Chair of the IEEE ICC 2009 (Wireless Communications Symposium), the Co-Chair of the IEEE GLOBECOM 2008 (Signal Processing for Communications Symposium), and the General Track Chair of the IEEE VTC 2008. He received the IEEE Communications Society SPCE Outstanding Service Award 2012 and the IEEE Communications Society RCC Outstanding Service Award 2014. He is an Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS and the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. He was an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS (2006–2011), the IEEE WIRELESS COMMUNICATIONS LETTERS, and the IEEE SIGNAL PROCESSING LETTERS.