

Radar HRRP Statistical Recognition With Local Factor Analysis by Automatic Bayesian Ying-Yang Harmony Learning

Lei Shi, Penghui Wang, Hongwei Liu, *Member, IEEE*, Lei Xu, *Fellow, IEEE*, and Zheng Bao, *Senior Member, IEEE*

Abstract—Radar high-resolution range profiles (HRRPs) are typical high-dimensional, non-Gaussian and interdimension dependently distributed data, the statistical modelling of which is a challenging task for HRRP based target recognition. Assuming the HRRP data follow interdimension dependent Gaussian distribution, factor analysis (FA) was recently applied to describe radar HRRPs and a two-phase procedure was used for model selection, showing promising recognition results. Besides the interdimensional dependence, this paper further models the non-Gaussianity of the radar HRRP data by local factor analysis (LFA). Moreover, since the two-phase procedure suffers from extensive computation and inaccurate evaluation on high-dimensional finite HRRPs, we adopt an automatic Bayesian Ying-Yang (BYY) harmony learning, which determines the component number and the hidden dimensionalities of LFA automatically during parameter learning. Experimental results show incremental improvements on recognition accuracy by three implementations, progressively from a two-phase FA, to a two-phase LFA, and then to an automatically learned LFA by BYY harmony learning.

Index Terms—Automatic model selection, Bayesian Ying-Yang (BYY) harmony learning, high-range resolution profile (HRRP), local factor analysis (LFA), radar automatic target recognition.

I. INTRODUCTION

RADAR automatic target recognition (RATR) is to identify the unknown target from its radar-echoed signatures. A high-resolution range profile (HRRP) is the coherent summation amplitudes of the complex time returns from target scatterers in each range cell, which contains target structure information such as target size and scattering distribution. Therefore, radar HRRP target recognition has received intensive attention from

the RATR community [1]–[3]. Bayesian classifiers are widely used for this typical pattern recognition task [4], [5].

Radar HRRP data is typically *high-dimensional, non-Gaussian, and interdimension dependently* distributed, the statistical modelling of which is a challenging task [6]–[8]. Nonparametric methods suffer from the curse of dimensionality in this application [9], and parametric methods are usually preferred [4], [6] and also considered in this paper. In the literature, many efforts have been made on two important topics. The first is how to determine a family of parametric models that can appropriately describe HRRP distribution, and the second is how to appropriately learn the model based on finite HRRP training samples.

On selecting an appropriate parametric model, early efforts [1], [7], [8] assumed that the range cells (dimensions) in HRRP are independently Gaussian distributed, which nevertheless was later found inappropriate based on physical arguments and results of empirical investigations [6], [10]. There are mainly two streams of efforts for improvement. On one hand, papers [5], [6], [10] extended the independent Gaussian to independent (mixture of) Gamma distribution. On the other hand, papers [4], [11] considered dependence among dimensions by a Gaussian factor analysis (FA) model. However, none of these efforts considers both non-Gaussian and interdimensional dependence simultaneously. This paper is thus motivated to consider both by mixture of factor analyzers (MFA) [12], or called local factor analysis (LFA) (see [13, Sec. 3.2]). Possible alternative choices may include (mixture of) correlated Gamma, (mixture of) independent component analysis and (mixture of) correlated log-normal, which deserve future studies.

Once a family of parametric models has been chosen, we approach the learning task, which consists of parameter learning for determining all unknown parameters given a model scale k and model selection for choosing an appropriate scale k to avoid under-fitting or over-fitting [14]. For LFA model, this k consists of the number of components and the local hidden dimensionalities. Parameter learning is usually implemented by the expectation-maximization (EM) algorithm [15], [16] under the maximum-likelihood (ML) principle. Model selection is conventionally tackled by a two-phase procedure with the help of a model selection criterion, such as Akaike's Information Criterion (AIC) and Bayesian Inference Criterion (BIC) [4], [14]. The work in [4] implements a two-phase procedure with AIC and BIC on FA. This paper implements a two-phase procedure on LFA instead, which improves the HRRP recognition performance in [4].

Manuscript received November 12, 2009; revised April 07, 2010 and September 05, 2010; accepted September 25, 2010. Date of publication October 18, 2010; date of current version January 12, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Deniz Erdogmus. This work was supported by a grant from the Research Grant Council of the Hong Kong SAR (Project No: CUHK4180/10E).

L. Shi and P. Wang are with Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, China, and also with the National Laboratory of Radar Signal Processing, Xidian University, Xi'an, China (e-mail: shil@cse.cuhk.edu.hk; wangpenghui@mail.xidian.edu.cn).

H. Liu and Z. Bao are with National Laboratory of Radar Signal Processing, Xidian University, Xi'an, China (e-mail: hwliu@xidian.edu.cn; zhbao@xidian.edu.cn).

L. Xu is with Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, China (e-mail: lxu@cse.cuhk.edu.hk).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSP.2010.2088391

However, for high-dimensional data such as HRRPs, this two-phase implementation inevitably suffers from extensive computation and unreliable estimated values of the criterion (see [13, Sec. 2.1] for a detailed discussion). To relieve these problems, this paper further adopts an automatic Bayesian Ying-Yang (BYY) harmony learning algorithm [13], [17], [18] for LFA, which automatically determines the number of components and local hidden dimensionalities during parameter learning (see the interpretations in [13, Eqs. (3) and (4)]). Based on the LFA model, this automatic algorithm not only obtains a better HRRP target recognition performance, but is also much faster than the two-phase procedure.

In addition, there exists one crucial target-aspect sensitivity in HRRP data [4], [5], [11]. The works in [4] and [19] partitioned consecutive HRRP samples into several aspect-frames and then modelled HRRPs in different frames separately. Focusing on parametric modelling, paper [4] proposed a heuristic partition mechanism with the help of AIC and BIC. The effort in [19] proposed another adaptive frame segmentation method by evaluating the curvature of HRRP manifold in a nonparametric way. Following [4], this paper further implements the heuristic partition mechanism under the BYY harmony model selection criterion.

The rest of this paper is organized as follows. In Section II, we give some background knowledge about HRRP based target recognition. Section III gives a review of existing works and justifies the motivations of this work. We implement a two-phase procedure on LFA instead of FA, and compare its HRRP recognition performances with [4]. In Section IV, we further introduce an automatic BYY harmony learning algorithm on LFA, and describe the heuristic aspect-frame partition based on a BYY harmony criterion. Section V systematically compares the HRRP target recognition performances of the two-phase procedures on FA and LFA, and the automatic BYY learning algorithm for LFA. Finally, some concluding remarks are drawn in Section VI.

II. BACKGROUND FOR RATR BASED ON HRRP

A. Target Recognition Based on HRRP

For a high resolution radar, the wavelength of radar signal is far smaller than the target size, and the electromagnetism characteristic of the target can be described by the scattering center model approximately [20], [21]. According to this model, a target consists of many scatterers distributed in several range cells along the radar line of sight (LOS). Intuitively, an HRRP can be viewed as a projection of radar returns from these scatterers onto the radar LOS, as illustrated in Fig. 1. Formally, an HRRP $\mathbf{x} \in \mathcal{R}^d$ is the coherent summation amplitudes of the complex returns from target scatterers in each range cell, where d is the number of range cells, i.e. the dimensionality of HRRP. Measured HRRPs are usually high-dimensional, e.g., $d = 256$ in [4], [11], and this paper.

Carrying essential information about the target's overall size, shape and structure, HRRP is widely used as a discriminative feature for radar automatic target recognition (RATR) [1], [2], [22], [23]. For the task of HRRP based radar target recognition, Bayesian classifiers have been widely investigated in the RATR

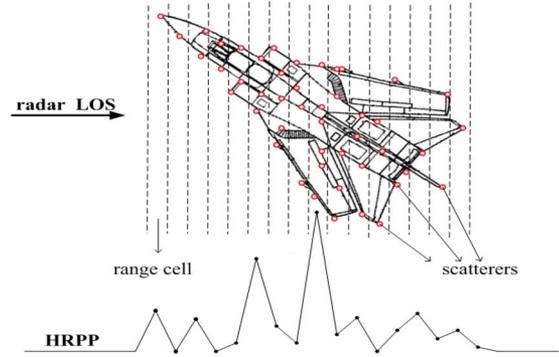


Fig. 1. Radar returns from the scatterers on the target are projected onto the LOS, resulting in an HRRP.

community [4], [10], [11]. For a C -class classification task, an HRRP sample \mathbf{x}_t is assigned to the class $\hat{c}(\mathbf{x}_t) \in \{1, \dots, C\}$ with the maximum a posteriori (MAP), i.e.

$$\hat{c}(\mathbf{x}_t) = \arg \max_c p(c|\mathbf{x}_t) = \arg \max_c [p_c(\mathbf{x}_t)p(c)] \quad (1)$$

where $p_c(\mathbf{x}) = p(\mathbf{x}|c)$ is the conditional probability density given class c and $p(c)$ is a prior over class c . To implement (1), we need to estimate $p_c(x)$ for each class. In the literature, two kinds of methods can be used for density estimation, i.e., nonparametric and parametric. However, since the HRRP data is high dimensional, nonparametric method not only is time and storage consuming, but is also insufficient or inaccurate to describe HRRP data distribution due to the sample sparseness [9], therefore impractical. Following [4], [10], this paper considers parametric modeling.

B. Sensitivities and Aspect-Frames Partition

Before radar HRRP statistical modeling, three sensitivity issues need to be considered, namely, translation [5], [24], amplitude-scale [5] and target-aspect sensitivities [4], [5], [11]. The former two could be commonly tackled by translation alignment [25], [26] and amplitude-scale normalization [4], [5], respectively. Similar to [4], we will assume all HRRPs in each frame have been slide correlation aligned and L_2 normalized, without further specification.

The target-aspect sensitivity [5], [6] refers to the issue that the HRRP distribution varies as the target-aspect angle changes. To tackle this problem, we adopt the divide-and-conquer policy in [4], [5], [11], i.e., consecutive HRRP samples are partitioned into different sectors, named as aspect-frames, and then all aspect-frames are modelled separately. More precisely, HRRPs of a target c are divided into K_c frames, and a model is used to describe $p_c(\mathbf{x}|j)$ in each frame j . Then, the conditional probability $p_c(\mathbf{x}_t)$ in (1) for classifying a testing sample \mathbf{x}_t is calculated as

$$p_c(\mathbf{x}_t) \triangleq \max_j [p_c(\mathbf{x}_t|j)p_c(j)] \quad (2)$$

where $p_c(j)$ is a prior probability of the j th aspect-frame in target c , which is assigned to be proportional to its sample size in this frame. For each pair of c and j , the density $p_c(\mathbf{x}|j)$ will be denoted as $p(\mathbf{x})$ in the sequel for simplicity.

plane	length (m)	width (m)	height (m)
<i>Yak-42</i>	36.38	34.88	9.83
<i>Cessna</i>	14.40	15.90	4.57
<i>An-26</i>	23.80	29.20	9.83

Fig. 2. Descriptions of the plane target parameters of the radar HRRP data corpus used in experiments.

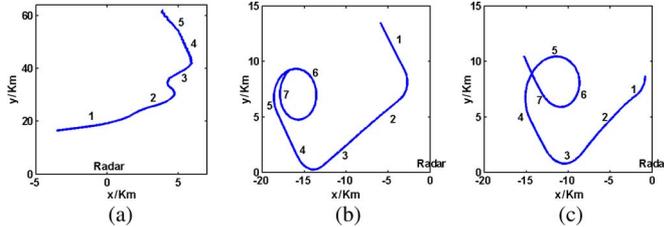


Fig. 3. Projections of target trajectories onto ground plane. (a) *Yak-42*. (b) *Cessna*. (c) *An-26*.

To partition the consecutive training samples into aspect-frames, both the frame number and the partition positions need to be determined. Efforts in [1], [5], [7], [8] conducted the segmentation with equal intervals. The work [4] improves this equal-interval partition by a heuristic partition approach, which sequentially determines aspect-frame boundaries based on AIC or BIC. For a comparison, this paper adopts the same heuristic partition mechanism as described in Section III-B of [4] and restated by the following two nested levels:

- 1) In the inner level, a set of candidate aspect-frames is built. A model is learned based on the samples in each candidate frame.
- 2) In the outer level, the candidate frame with the optimal value of a criterion, such as AIC, BIC and BYY harmony criterion, is selected as the current frame.

C. Data Set Description

All experimental investigations in this paper will be made based on the measured HRRP data corpus of 3 planes, including *Yak-42*, *Cessna Citation S/II*, and *An-26*. The bandwidth of the radar signal is 400 MHz, and the center frequency is 5520 MHz. The dimensionality of the data, i.e., the number of range cells in an HRRP, is $d = 256$. Fig. 2 shows the parameters of the plane targets. The projections of target trajectories onto ground plane are segmented as shown in Fig. 3, where each segment contains 26 000 samples. The training and testing segments are chosen based on the two reasons: 1) the training data cover almost all the target-aspect angles of the testing data; 2) it is preferred to choose training and testing data differently for a reasonable evaluation of the generalization performance. Finally, the 2nd and the 5th segments of *Yak-42*, the 6th and the 7th segments of *Cessna Citation S/II*, the 5th and the 6th segments of *An-26* are taken as the training samples, with the remaining left for testing.

In [4], the experiments were made on one particular subset uniformly sampled from this corpus, where the training and testing sample sizes are, respectively, 3200 and 2000 for each target. In this paper, we test all methods on not only that subset but also other uniformly randomly sampled data sets of the same size.

III. STATISTICAL MODELS AND LEARNING APPROACHES

A. Choosing Statistical Models

Several early efforts [1], [7], [8] assumed the range cells in HRRP are *independently Gaussian* distributed, i.e., $p(\mathbf{x}) = \prod_{i=1}^d p(x_i)$ and each $p(x_i) = G(x_i|\mu_i, \sigma_i^2)$, where $G(\cdot)$ denotes Gaussian density. This assumption was later found inappropriate based on physical arguments and empirical investigations [5], [6]. To tackle this problem, the development falls into two streams as follows.

The first stream assumes $p(\mathbf{x}) = \prod_{i=1}^d p(x_i)$ with $p(x_i)$ of each range cell being a non-Gaussian distribution. For instance, Webb *et al.* [6], [10] considered a Gamma mixture model for $p(x_i)$. In [5], a compounded model was introduced, where $p(x_i)$ is either a Gamma distribution or a Gaussian mixture distribution.

The second stream considers the interdimensional statistical dependence of range cells [4], [11]. The authors in [4] suggested to model HRRP in a frame by Gaussian factor analysis, assuming that a d -dimensional observable variable \mathbf{x} is generated via a linear mapping from an m -dimensional independent Gaussian factor \mathbf{y} with $m < d$, plus a Gaussian noise which is independent of the factors. The resulted $p(\mathbf{x})$ by FA is still a Gaussian but the degrees-of-freedom of its covariance matrix is greatly reduced, which is more suitable for high dimensional data sets with small sample sizes like radar HRRPs [4]. Moreover, principal component analysis (PCA) is a special case of FA under the maximum-likelihood principle [12], [18].

Nevertheless, none of these efforts considers both non-Gaussian and interdimensional dependence simultaneously. This paper is thus motivated to extend FA [4] to LFA. We consider a d -dimensional observable variable \mathbf{x} from $p(\mathbf{x}) = \sum_{l=1}^k \alpha_l p(\mathbf{x}|l)$, with $\forall \alpha_l \geq 0$, $\sum_{l=1}^k \alpha_l = 1$, and each component $p(\mathbf{x}|l)$ is an FA

$$\begin{aligned} p(\mathbf{x}|\mathbf{y}, l) &= G(\mathbf{x}|\mathbf{U}_l \mathbf{y} + \boldsymbol{\mu}_l, \boldsymbol{\Psi}_l), \\ p(\mathbf{y}|l) &= G(\mathbf{y}|\mathbf{0}, \boldsymbol{\Lambda}_l), \quad \text{s.t. } \mathbf{U}_l^T \mathbf{U}_l = \mathbf{I}_{m_l}. \end{aligned} \quad (3)$$

For each component l , α_l is the mixing weight, m_l is the local dimensionality of the hidden factor \mathbf{y} , \mathbf{U}_l is a $d \times m_l$ rectangular orthogonal loading matrix. Moreover, $\boldsymbol{\Lambda}_l$ is an $m_l \times m_l$ diagonal covariance, $\boldsymbol{\mu}_l$ is a d -dimensional mean vector, $\boldsymbol{\Psi}_l$ is a diagonal covariance matrix for noises. Combining FA and Gaussian Mixture Model (GMM), LFA is equipped with both interdimensional dependence and non-Gaussianity.

B. Two-Phase Model Selection on LFA

Given a family of parametric models, the task of modeling $p(\mathbf{x}|\boldsymbol{\Theta}_k)$ consists of estimating the parameters $\boldsymbol{\Theta}_k$ and selecting an appropriate model scale \mathbf{k} based on training samples $\mathbf{X}_N = \{\mathbf{x}_t\}_{t=1}^N$, where \mathbf{k} refers to the hidden dimensionality m for FA [4], the number of components for Gamma mixture model [10], the component number k and the local hidden dimensionalities $\{m_l\}_{l=1}^k$ for LFA. One widely used method to estimate parameters $\boldsymbol{\Theta}_k$ is the maximum-likelihood learning, i.e. $\hat{\boldsymbol{\Theta}}_k = \arg \max_{\boldsymbol{\Theta}_k} \prod_{t=1}^N p(\mathbf{x}_t)$ (if samples $\{\mathbf{x}_t\}_{t=1}^N$ are assumed independently identical distributed), usually implemented by an EM algorithm [16]. A particular EM algorithm for LFA model has

TABLE I
THE CONFUSION MATRICES AND ACRRs OF LFA SELECTED BY AIC AND BIC. FOR SHORT, *Y* STANDS FOR *Yak-42*, *C* FOR *Cessna*, AND *A* FOR *An-26*. THE ROW LABELED BY (*) SHOWS THE ACRRs IN [4] OF FA BY AIC AND BIC

	Equal Interval Partition						Heuristic Partition					
	AIC			BIC			AIC			BIC		
Plane	<i>Y</i>	<i>C</i>	<i>A</i>	<i>Y</i>	<i>C</i>	<i>A</i>	<i>Y</i>	<i>C</i>	<i>A</i>	<i>Y</i>	<i>C</i>	<i>A</i>
<i>Y</i>	100	0.90	5.20	100	0.35	3.75	100	0.20	2.70	100	0.20	1.85
<i>C</i>	0	96.55	0.65	0	97.60	0.60	0	98.15	0.50	0	99.10	0.45
<i>A</i>	0	2.55	94.15	0	2.05	95.65	0	1.65	96.80	0	0.70	97.70
ACRR	96.90			97.75			98.32			98.93		
(*)FA	94.50			96.78			95.53			98.73		

been formulated in [15]. However, one remaining but important issue is how to appropriately determine \mathbf{k} to avoid over-fitting or under-fitting, which is a typical example of model selection [14].

Conventionally, model selection can be addressed and treated in a *two-phase procedure*. In the first phase, a candidate range \mathcal{M} for \mathbf{k} is pre-specified and assumed to contain the optimal model scale \mathbf{k}^* . For each $\mathbf{k} \in \mathcal{M}$, parameter estimates $\hat{\Theta}_{\mathbf{k}}$ are obtained by the EM algorithm [12], [16]. In the second phase, $\hat{\mathbf{k}}^*$ is selected as an estimate to \mathbf{k}^* via $\hat{\mathbf{k}}^* = \arg \min_{\mathbf{k} \in \mathcal{M}} J(\hat{\Theta}_{\mathbf{k}}, \mathbf{k})$, where $J(\hat{\Theta}_{\mathbf{k}}, \mathbf{k})$ is a model selection criterion such as AIC and BIC [14], [27] as follows:

$$J_{*IC}(\hat{\Theta}_{\mathbf{k}}, \mathbf{k}) = -2L(\mathbf{X}_N | \hat{\Theta}_{\mathbf{k}}) + C_{*IC}D(\mathbf{k}),$$

with $C_{AIC} = 2$ and $C_{BIC} = \ln N$, (4)

where $L(\mathbf{X}_N | \hat{\Theta}_{\mathbf{k}})$ is the log-likelihood of training samples \mathbf{X}_N based on the ML-estimated parameters $\hat{\Theta}_{\mathbf{k}}$, and $D(\mathbf{k})$ is the number of free parameters with a scale \mathbf{k} . Considering LFA for $p(\mathbf{x} | \Theta_{\mathbf{k}})$ with $\mathbf{k} = \{k, \{m_l\}_{l=1}^k\}$, we have $D(\mathbf{k}) = 2kd + k - 1 + \sum_{l=1}^k m_l(d - ((m_l - 1)/2))$, where the number of free parameters in $\{\alpha_l\}$ is $k - 1$, in $\{\mu_l\}$ and $\{\Psi_l\}$ are both kd , in $\{\mathbf{U}_l\}$ is $\sum_{l=1}^k (dm_l - m_l(m_l + 1)/2)$, and in $\{\Lambda_l\}$ is $\sum_{l=1}^k m_l$.

In experiments, for the equal interval partition, we set 35 aspect-frames for *Yak-42*, 50 for *Cessna Citation S/II*, and 50 for *An-26*, same as in [4]. For the heuristic partition, the criterion AIC or BIC is used simultaneously in the outer level to determine frames and in the inner level to select LFA models. On the same data set as in [4], the recognition accuracy by LFA based on a two-phase procedure is reported in Table I, in terms of confusion matrix and average correct recognition rate (ACRR). As shown, LFA-AIC/BIC outperforms FA-AIC/BIC taken from [4].

However, since the candidate scale range \mathcal{M} for LFA is unknown and difficult to be pre-specified appropriately, usually it has to be searched among a wide range. Consequently, the two-phase procedure still has the following two key problems:

- 1) Enumerating \mathbf{k} for parameter estimation requires a huge computation.
- 2) As the model scale \mathbf{k} increases, parameter estimation is less reliable so that the evaluation of the criterion is less accurate for selecting an appropriate $\hat{\mathbf{k}}^*$.

Readers are referred to [13, Sec. 2.1] for a systematic discussion on the two-phase procedure. Subsequently, we tackle these problems by automatic BYY harmony learning.

IV. AUTOMATIC BYY HARMONY LEARNING FOR HRRP BASED TARGET RECOGNITION

A. Automatic BYY Harmony Learning on LFA

Firstly proposed in 1995 [17] and then systematically developed over a decade, Bayesian Ying-Yang harmony learning provides a general statistical learning framework for parameter learning and model selection under a best harmony principle. BYY harmony learning on typical structures leads to new model selection criteria, new techniques for implementing regularization and a class of algorithms that implement automatic model selection during parameter learning.

Denoting the observation as \mathbf{X} and its inner representation as $\mathbf{R} = \{\mathbf{Y}, \Theta\}$, two types of decomposition $p(\mathbf{X}, \mathbf{R}) = p(\mathbf{R} | \mathbf{X})p(\mathbf{X})$ and $q(\mathbf{X}, \mathbf{R}) = q(\mathbf{X} | \mathbf{R})q(\mathbf{R})$ are called Yang machine and Ying machine, respectively. Such a Ying-Yang pair is called a BYY system, as illustrated by the left part of Fig. 4. The best harmony principle is to mathematically maximize the following harmony measure in a general form:

$$\begin{aligned} H(p||q) &= \int p(\mathbf{R} | \mathbf{X})p(\mathbf{X}) \ln [q(\mathbf{X} | \mathbf{R})q(\mathbf{R})] d\mathbf{X}d\mathbf{R} \\ &= \int p(\Theta | \mathbf{X})H(p||q, \Theta)d\Theta, \\ H(p||q, \Theta) &= \int p(\mathbf{Y} | \mathbf{X}, \Theta)p(\mathbf{X}) \ln [q(\mathbf{X} | \mathbf{Y}, \Theta)q(\mathbf{Y} | \Theta)] \\ &\quad \times d\mathbf{X}d\mathbf{Y} + \ln q(\Theta). \end{aligned} \quad (5)$$

Different from maximizing the likelihood function, an important nature of maximizing $H(p||q)$ is that it leads to not only a best matching between the Ying-Yang pair, but also a compact model with a least complexity. Such an ability can be observed and investigated from several perspectives, and here we only introduce one of them due to space limit. On one hand, maximizing $H(p||q)$ forces Ying machine $q(\mathbf{X}, \mathbf{R})$ to match Yang machine $p(\mathbf{X}, \mathbf{R})$. Due to a finite sample size and practical constraints imposed on the Ying-Yang structures, a perfect equality may not be really reached but still be approached as possible as it can. At this equality, $H(p||q)$ becomes the negative entropy that describes the complexity of the system. Further maximizing it actually minimizes the system complexity, which provides a model selection ability.

The above is only a brief introduction on the Bayesian Ying-Yang harmony learning. Readers are referred to [13] for a recent systematic description.

In the sequel, we focus on LFA model described in (3), where each observation \mathbf{x} comes from the hidden representation $\mathbf{Y} = \{\mathbf{y}, l\}$ via the Ying machine

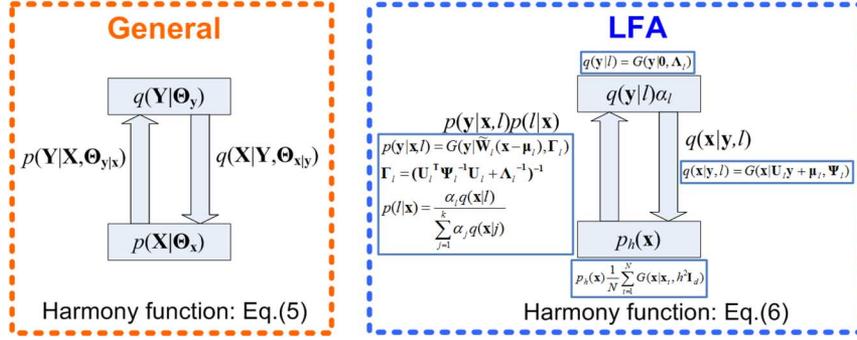


Fig. 4. The BYY system in the general form (left) and in the specific form for LFA (right).

$q(l)q(\mathbf{y}|l)q(\mathbf{x}|\mathbf{y}, l)$ according to (3). A given set of input samples is considered by a smoothed Parzen window density $p_h(\mathbf{x}) = (1/N) \sum_{t=1}^N G(\mathbf{x}|\mathbf{x}_t, h^2 \mathbf{I}_d)$. Moreover, it follows from the ‘‘variety preservation principle’’ (see Eq. (27) in [13]) that we get $p(l|\mathbf{x}) = \alpha_l q(\mathbf{x}|l) / \sum_j \alpha_j q(\mathbf{x}|j)$ and $p(\mathbf{y}|\mathbf{x}, l) = G(\mathbf{y}|\tilde{\mathbf{W}}_l(\mathbf{x} - \boldsymbol{\mu}_l), \Gamma_l)$, where $\tilde{\mathbf{W}}_l$ consists of unknown parameters and $\Gamma_l = [\partial^2 \ln[q(\mathbf{x}|\mathbf{y}, l)q(\mathbf{y}, l)] / \partial \mathbf{y} \partial \mathbf{y}^T]^{-1} = (\mathbf{U}_l^T \boldsymbol{\Psi}_l^{-1} \mathbf{U}_l + \Lambda_l^{-1})^{-1}$ (according to Eq. (31) in [13]). Also, we ignore the priors on $\Theta = \{\{\alpha_l, \boldsymbol{\mu}_l, \boldsymbol{\Psi}_l, \mathbf{U}_l, \Lambda_l, \tilde{\mathbf{W}}_l\}_{l=1}^k, h\}$ except the one on h only, i.e., $q(\Theta) = q(h)$. Furthermore, $p(\Theta|\mathbf{X})$ is considered to be a free structure, and thus maximizing $H(p||q)$ leads to $p(\Theta|\mathbf{X}) = \delta(\Theta - \hat{\Theta})$ with $\hat{\Theta} = \arg \max_{\Theta} H(p||q, \Theta)$.

Algorithm 1: Automatic BYY harmony learning on local factor analysis (BYY-A)

Input: An observed data set $\mathbf{X}_N = \{\mathbf{x}_t\}_{t=1}^N$.

Output: LFA model $\Theta = \{\alpha_l, \boldsymbol{\mu}_l, \mathbf{U}_l, \Lambda_l, \boldsymbol{\Psi}_l, \tilde{\mathbf{W}}_l\}_{l=1}^k$ together with model scales $\{k, \{m_l\}_{l=1}^k\}$.

1 **Initialization:** Randomly initialize a model with large enough k and $\{m_l\}_{l=1}^k$; set $\tau = 0$ and $H(\tau) = -\infty$;

2 **repeat**

3 **Yang-Step:** After randomly picking a sample \mathbf{x}_t , update Yang machine as follows:

4 **for** $l = 1, \dots, k$ **do**

5 $\tilde{\mathbf{W}}_l^{new} = \tilde{\mathbf{W}}_l + \eta p_{l,t} (\Lambda_l^{-1} + \mathbf{U}_l^T \boldsymbol{\Psi}_l^{-1} \mathbf{U}_l)$
 $(\mathbf{W}_l - \tilde{\mathbf{W}}_l) \mathbf{S}_{l,t}$;

6 $\mathbf{S}_{l,t} = \mathbf{e}_{l,t}, \mathbf{e}_{l,t}^T + h^2 \mathbf{I}_d$, $\mathbf{e}_{l,t} = \mathbf{x}_t - \boldsymbol{\mu}_l$,
 $\mathbf{W}_l = \Lambda_l \mathbf{U}_l^T \mathbf{M}_l$, $\mathbf{M}_l = (\mathbf{U}_l \Lambda_l \mathbf{U}_l^T + \boldsymbol{\Psi}_l)^{-1}$;

7 $p_{l,t} = p(l|\mathbf{x}_t)[1 + \delta_L(l, \mathbf{x}_t)]$, $\delta_L(l, \mathbf{x}_t)$
 $= \ln p(l|\mathbf{x}_t) - \sum_{j=1}^k p(j|\mathbf{x}_t) \ln p(j|\mathbf{x}_t)$;

8 **Ying-Step:** Update Ying machine parameters and conduct automatic model selection:

9 **for** $l = 1, \dots, k$ **do**

10 $\alpha_l^{new} = \exp(\beta_l^{new}) / \sum_{j=1}^k \exp(\beta_j^{new})$,
 $\beta_l^{new} = \beta_l + \eta(p_{l,t} - \alpha_l \sum_{j=1}^k p_{j,t})$, $\beta_l = \ln \alpha_l$;

11 \odot **if** $\alpha_l \rightarrow 0$ **then** discard component l , let $k = k - 1$ **and continue**;

12 $\boldsymbol{\mu}_l^{new} = \boldsymbol{\mu}_l + \eta p_{l,t} \mathbf{e}_{l,t}$;

13 $\Lambda_l^{new} = (\tilde{\Lambda}_l^{new})^2$, $\tilde{\Lambda}_l^{new} = (1 - \eta p_{l,t}) \tilde{\Lambda}_l$
 $+ \eta p_{l,t} \text{diag}[\tilde{\mathbf{W}}_l^T \mathbf{S}_{l,t} \tilde{\mathbf{W}}_l]^{1/2}$, $\tilde{\Lambda}_l = \Lambda_l^{1/2}$;

14 \odot **if** $\Lambda_l^{(j)} \rightarrow 0$ **then** discard hidden dimension j of component l and let $m_l = m_l - 1$;

15 $\boldsymbol{\Psi}_l^{new} = (\tilde{\boldsymbol{\Psi}}_l^{new})^2$, $\tilde{\boldsymbol{\Psi}}_l^{new} = (1 - \eta p_{l,t}) \tilde{\boldsymbol{\Psi}}_l +$
 $\eta p_{l,t} \text{diag}[(\mathbf{I}_d - \mathbf{U}_l \tilde{\mathbf{W}}_l) \mathbf{S}_{l,t} (\mathbf{I}_d - \mathbf{U}_l \tilde{\mathbf{W}}_l)^T]^{1/2}$,
 $\tilde{\boldsymbol{\Psi}}_l = \boldsymbol{\Psi}_l^{1/2}$;

16 $\mathbf{U}_l^{new} = \mathbf{U}_l + \eta(\mathbf{G}_{\mathbf{U}_l} - \mathbf{U}_l \mathbf{G}_{\mathbf{U}_l}^T \mathbf{U}_l)$, with
 $\mathbf{G}_{\mathbf{U}_l} = p_{l,t} \boldsymbol{\Psi}_l^{-1} (\mathbf{I}_d - \mathbf{U}_l \tilde{\mathbf{W}}_l) \mathbf{S}_{l,t} \tilde{\mathbf{W}}_l^T$;

17 Normalize \mathbf{U}_l every 5 steps via $\mathbf{U}_l^{new} =$
 $\mathbf{U}_l^{new} (\mathbf{U}_l^{new T} \mathbf{U}_l^{new})^{-1/2}$;

18 **Smoothing-Step:** $h^{2new} = h^{new2}$, $h^{new} = h + \eta \Delta h$,
 $\Delta h = (d/h) - h \sum_{l=1}^k \alpha_l \text{Tr}[\mathbf{M}_l] - (\beta/h^3)$;

19 where $\beta = \sum_{t=1}^N \sum_{\tau=1}^N (\mathbf{x}_\tau - \mathbf{x}_t)^T (\mathbf{x}_\tau - \mathbf{x}_t)$
 $\times \exp[-(1/2h^2)(\mathbf{x}_\tau - \mathbf{x}_t)^T (\mathbf{x}_\tau - \mathbf{x}_t)] / \sum_{t=1}^N \sum_{\tau=1}^N$
 $\times \exp[-(1/2h^2)(\mathbf{x}_\tau - \mathbf{x}_t)^T (\mathbf{x}_\tau - \mathbf{x}_t)]$;

20 **if** another $N/5$ iterations have passed **then** let
 $\tau = \tau + 1$, and calculate $H(\tau)$ by (6);

21 **until** $H(\tau) - H(\tau - 1) < \epsilon H(\tau - 1)$, with $\epsilon = 10^{-5}$
in our implementation;

The above Ying-Yang components for LFA are summarized in the right part of Fig. 4. Substituting them into (5), certain simplification leads us to the following detailed expression:

$$\begin{aligned}
H(p||q, \Theta) &= H_f(p||q, \Theta) + \ln q(h), \\
q(h) &= -\ln \left[\frac{1}{N} \sum_{t=1}^N \sum_{\tau=1}^N G(\mathbf{x}_t|\mathbf{x}_\tau, h^2 \mathbf{I}_d) \right], \\
H_f(p||q, \Theta) &= \frac{1}{2N} \sum_{t=1}^N \sum_{l=1}^k p(l|\mathbf{x}_t) \\
&\times \{2 \ln \alpha_l - (d + m_l) \ln(2\pi) - m_l - \ln |\Lambda_l| - \ln |\boldsymbol{\Psi}_l| \\
&- \text{Tr} \left[\left[\tilde{\mathbf{W}}_l^T \Lambda_l^{-1} \tilde{\mathbf{W}}_l + (\mathbf{I}_d - \mathbf{U}_l \tilde{\mathbf{W}}_l)^T \right. \right. \\
&\quad \left. \left. \times \boldsymbol{\Psi}_l^{-1} (\mathbf{I}_d - \mathbf{U}_l \tilde{\mathbf{W}}_l) \right] \right. \\
&\quad \left. \cdot [(\mathbf{x}_t - \boldsymbol{\mu}_l)(\mathbf{x}_t - \boldsymbol{\mu}_l)^T + h^2 \mathbf{I}_d] \right\}. \tag{6}
\end{aligned}$$

TABLE II
THE CONFUSION MATRICES AND ACRRS BY LFA CLASSIFIER LEARNED BY BYY HARMONY LEARNING

	Equal Interval Partition						Heuristic Partition					
	no smoothing			smoothing			no smoothing			smoothing		
Plane	Y	C	A	Y	C	A	Y	C	A	Y	C	A
Y	100	0.80	3.30	100	0.45	2.80	100	0.15	1.85	100	0.10	1.75
C	0	97.15	0.45	0	98.00	0.45	0	99.20	0.35	0	99.35	0.25
A	0	2.05	96.25	0	1.55	96.75	0	0.65	97.80	0	0.55	98.00
ACRR	97.80			98.25			99.00			99.12		

An implementation of maximizing (6) is summarized in Algorithm 1, for which we address four points as follows.

First, the above (6) for the harmony measure may not be exactly the same as the counterpart given in [13, Fig. 9], due to a different order used in handling the gradient and the integration (or summation). This Algorithm 1 is a gradient-based adaptive implementation, while the paper [13] provides a Ying-Yang alternative iteration as given in Fig. 9 of [13]. Interested readers are further referred to [13, Sec. 4.3].

Second, the use of $p_h(\mathbf{x})$ with $h^2 > 0$ leads to a smoothing regularization as discussed in [28], which is equivalent to adding noises to samples. To determine an appropriate value for the smoothing parameter h , one choice for $q(h)$ is shown in (6) with its detailed updating formulas given in the ‘‘Smoothing-Step’’ in Algorithm 1. Readers are referred to [28] for further details on getting h .

Third, it follows from the interpretations on Eqs. (3)&(4) in [13] that Algorithm 1 is capable of automatic model selection. As the component number k and the hidden dimensionalities $\{m_l\}_{l=1}^k$ are initialized as large enough integers, maximizing $H(p||q, \Theta)$ provides an intrinsic force to push $\alpha_l \rightarrow 0$ if the l th component is extra, and push $\Lambda_l^{(j)} \rightarrow 0$ if the j th hidden dimension of the l th component is extra. In Algorithm 1, the extra sub-structures are removed in the steps labeled by ‘‘ \odot ’’.

Last but not least, remarks are provided on the implementation of Algorithm 1 as follows:

- 1) The initialization of Yang machine parameter $\widetilde{\mathbf{W}}_l$ for each component l is made by $\widetilde{\mathbf{W}}_l = \mathbf{W}_l$ with $\mathbf{W}_l = \Lambda_l \mathbf{U}_l^T (\mathbf{U}_l \Lambda_l \mathbf{U}_l^T + \Psi_l)^{-1}$.
- 2) On Lines 6 ~ 7, the determinant $|\mathbf{U}_l \Lambda_l \mathbf{U}_l^T + \Psi_l|$ and the inverse $(\mathbf{U}_l \Lambda_l \mathbf{U}_l^T + \Psi_l)^{-1}$ is computed indirectly via the following formulas, which are more efficient when $m_l \ll d$:

$$|\mathbf{U}_l \Lambda_l \mathbf{U}_l^T + \Psi_l| = \left| \Lambda_l^{-1} + \mathbf{U}_l^T \Psi_l^{-1} \mathbf{U}_l \right| |\Lambda_l| |\Psi_l|,$$

$$(\mathbf{U}_l \Lambda_l \mathbf{U}_l^T + \Psi_l)^{-1} = \Psi_l^{-1} - \Psi_l^{-1} \mathbf{U}_l \left(\Lambda_l^{-1} + \mathbf{U}_l^T \Psi_l^{-1} \mathbf{U}_l \right)^{-1} \times \mathbf{U}_l^T \Psi_l^{-1}.$$

- 3) If the value of $p(l|\mathbf{x}_t)$ is too small, calculating $p_{l,t}$ on Line 7 may exceed the machine precision due to the computation $\ln p(l|\mathbf{x}_t)$. This can be easily avoided by the following tricky but equivalent formula: $p_{l,t} = p(l|\mathbf{x}_t) + \ln[p(l|\mathbf{x}_t)^{p(l|\mathbf{x}_t)}] - p(l|\mathbf{x}_t) \sum_{j=1}^k \ln[p(j|\mathbf{x}_t)^{p(j|\mathbf{x}_t)}]$.

B. Harmony Criterion Based Heuristic Aspect-Frame Partition

To tackle the target-aspect sensitivity problem, we consider the heuristic aspect-frame partition, which has been restated as

the two nested levels in Section II-B. In the work [4], the first level, i.e., to learn an FA model for each candidate frame, is implemented in a two-phase implementation. Thus, the two levels in [4] (i.e., FA-AIC/BIC) is actually the following three-phase procedure:

- I: Estimate the parameters of FA for each candidate model scale (i.e., the hidden dimensionality of FA);
- II: Select the model scale by the optimal values of AIC/BIC;
- III: Implement Phases I&II on each candidate frame, and select the frame by the optimal value of AIC/BIC.

The inner level in Section II-B actually consists of the above Phases I and II, and the outer level is Phase III. The LFA-AIC/BIC in Section III-B is implemented with LFA taking the place of FA in the above three-phase procedure.

Similarly, the above three-phase procedure may also be implemented with LFA taking the place of FA and with the following BYY harmony criterion J_{BYY} (see [13, Sec. 2.2 and Fig. 5(b)]) taking the place of AIC/BIC:

$$J_{BYY} = -H(p||q, \Theta) + D(\mathbf{k})/(2N) \quad (7)$$

where $D(\mathbf{k})$ is the number of free parameters as described below (4). Experiments show further improvements over LFA-AIC/BIC but not so significant. Still, its Phases II&III suffer from the two key problems addressed at the end of Section III-B.

To relieve the problems, Phases I&II are simultaneously implemented by Algorithm 1 featured with automatic model selection, while the above J_{BYY} takes the place of AIC/BIC in Phase III. We denote the resulted two-phase procedure as LFA-BYY, which not only reduces the computational cost, but also further improves the recognition performance.

V. EXPERIMENTAL RESULTS

First, the HRRP recognition performances of LFA-BYY is investigated, with either heuristic aspect-frame partition or equal interval frame partition, and Algorithm 1 is implemented with or without data smoothing. The recognition accuracies are reported in Table II in terms of confusion matrix and average correct recognition rate. Moreover, Fig. 5 shows the resulted aspect-frames by heuristic partition with data smoothing. The size of each frame is proportional to its sample size, and thus we have 28, 30, and 42 aspect-frames for *Yak-42*, *Cessna* and *An-26*, respectively, in order.

According to the results in Table I and Table II, we summarize the observations as follows:

- 1) The results by FA-AIC/BIC, LFA-AIC/BIC and LFA-BYY show incremental improvements. The recognition accuracy on *Yak-42* is 100% by all approaches, and we

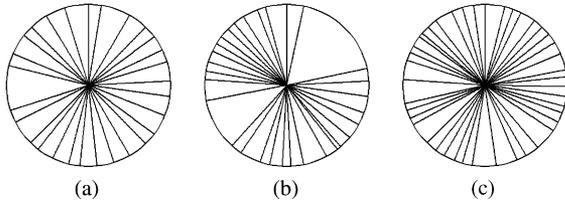


Fig. 5. The heuristically partitioned aspect frames by BYY-A in the inner level and BYY harmony criterion in the outer level. Each separated sector refers to a partitioned aspect-frame, with the angular size proportional to its sample size. The notation K_c represents the number of aspect-frames in corresponding class c . (a) *Yak-42* ($K_Y = 28$). (b) *Cessna* ($K_C = 30$). (c) *An-26* ($K_A = 42$).

only calculate the improvements of the average accuracy on *Cessna* and *An-26*. For the equal interval partition, LFA-AIC/BIC outperforms FA-AIC/BIC by 2.40%/0.97%, and LFA-BYY further outperforms LFA-AIC/BIC by 1.35%/0.50%. For the heuristic partition, LFA-AIC/BIC outperforms FA-AIC/BIC by 2.79%/0.20%, and LFA-BYY further outperforms LFA-AIC/BIC by 0.80%/0.19%.

- 2) The results with data smoothing outperform ones without data smoothing, by 0.29% in average, implying that the data smoothing technique is helpful for sparse data.
- 3) The numbers of resulted frames by LFA-BYY are (28, 30, 42) for the three planes, respectively, which are all smaller than (32, 39, 48) by FA-AIC and (31, 35, 47) by FA-BIC in [4].

We test the recognition performance of FA-AIC, LFA-BIC, and LFA-BYY on 20 data subsets, which are of the same size as in [4] randomly sampled from the corpus. Only the results by FA-BIC and LFA-BIC are reported since we already know from Table I that BIC outperforms AIC based on both FA and LFA. Table III lists the recognition accuracies in terms of mean \pm std together with the average time costs of determining one aspect-frame, from which improvements on recognition accuracies are incrementally obtained, as we proceed from FA-BIC to LFA-BIC and then to LFA-BYY. Especially, we have the following detailed observations:

- 1) The improvement from FA-BIC to LFA-BIC is at the cost of significantly increased time, mainly because Phase I of LFA-BIC is repeated for more times as enumerating the candidate model scale. Consequently, LFA-BIC is not preferred for this task when time cost is emphasized.
- 2) LFA-BYY achieves the best recognition performance at the cost of only 9 additional minutes compared to FA-BIC, with a considerably reduced deviation of recognition accuracies. That is, automatic BYY harmony learning can effectively relieve the key problems in implementing Phases I&II of FA-BIC and LFA-BIC, as addressed at the end of Section III-B.

To further investigate the impact of different initializations to LFA-BYY, we run LFA-BYY on each of the 20 data subsets for 50 times, and the LFA model is initialized randomly in each run. Based on the 1000 trials in total, the classification accuracy in percentage is 97.84 ± 1.95 for equal interval partition, and 98.81 ± 1.13 for heuristic partition, indicating that LFA-BYY is not very sensitive to the initialization.

TABLE III
RECOGNITION ACCURACY ROBUSTNESS (IN PERCENTAGE) AND AVERAGE TIME TO DETERMINE A FRAME (IN MINUTES) ON 20 RANDOM DATA SUBSETS. A THREE-PHASE IMPLEMENTATION BASED ON AIC COSTS THE SAME TIME AS THAT BASED ON BIC

	Equal Interval Partition	Heuristic Partition	Time Cost
FA-BIC	93.47 ± 4.33	95.13 ± 2.40	33.1
LFA-BIC	96.60 ± 2.51	97.28 ± 1.48	542.3
LFA-BYY	97.82 ± 1.93	98.75 ± 1.04	42.6

VI. CONCLUDING REMARKS

As a further investigation of a recent work [4] on radar HRRP data statistical modelling and recognition, this paper considers LFA for modelling both the interdimensional dependence and the non-Gaussian distribution of HRRPs. An automatic BYY harmony learning algorithm is adopted for learning LFA to relieve the difficulties encountered in a two-phase implementation. Moreover, the BYY harmony criterion is employed to replace AIC or BIC in the heuristic aspect-frame partition for tackling the target-aspect sensitivity. Recognition results on the same HRRP data set as [4] show incrementally improved performances. In a two-phase implementation, LFA obtains better recognition performances than FA in [4]. On LFA, the automatic BYY learning outperforms its counterpart of the two-phase implementation in help of AIC or BIC, for both equal interval and heuristic aspect-frame partition.

ACKNOWLEDGMENT

The first two authors would like to give special thanks to S. Tu for his valuable help on the revision of this paper.

REFERENCES

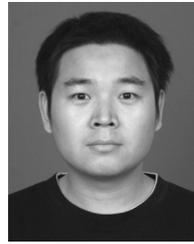
- [1] H.-J. Li and S.-H. Yang, "Using range profiles as feature vectors to identify aerospace objects," *IEEE Trans. Antennas Propag.*, vol. 41, no. 3, pp. 261–268, 1993.
- [2] R. A. Mitchell and R. Dewall, "Overview of high range resolution radar target identification," in *Proc. Autom. Target Recogn. Working Group*, Monterey, CA, 1994.
- [3] A. Zyweck and R. E. Bogner, "Radar target classification of commercial aircraft," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 32, no. 3, pp. 598–606, 1996.
- [4] L. Du, H.-W. Liu, and Z. Bao, "Radar HRRP statistical recognition: Parametric model and model selection," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 1931–1944, 2008.
- [5] L. Du, H.-W. Liu, Z. Bao, and J.-Y. Zhang, "A two-distribution compounded statistical model for radar HRRP target recognition," *IEEE Trans. Signal Process.*, vol. 54, pp. 2226–2238, Jun. 2006.
- [6] K. Copesey and A. R. Webb, "Bayesian Gamma mixture model approach to radar target recognition," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, pp. 1201–1217, Oct. 2003.
- [7] S. Hudson and D. Psaltis, "Correlation filters for aircraft identification from radar range profile," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 29, no. 3, pp. 1741–1748, 1993.
- [8] S. P. Jacobs and J. A. O. Sullivan, "Automatic target recognition using sequences of high resolution radar range-profiles," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 2, pp. 364–381, 2000.
- [9] E. Parzen, "On the estimation of a probability density function and mode," *Ann. Math. Statist.*, vol. 33, pp. 1065–1076, 1962.
- [10] A. R. Webb, "Gamma mixture models for target recognition," *Pattern Recogn.*, vol. 33, pp. 2045–2054, 2000.
- [11] L. Du, H.-W. Liu, and Z. Bao, "Radar HRRP statistical recognition based on hypersphere model," *Signal Process.*, vol. 88, no. 5, pp. 1176–1190, 2008.
- [12] G. E. Hinton, M. Revow, and P. Dayan, "Recognizing handwritten digits using mixtures of linear models," *Advances in NIPS*, pp. 1015–1022, 1994.

- [13] L. Xu, "Bayesian Ying-Yang system, best harmony learning, and five action circling," *J. Frontiers Elect. Electron. Eng. China (A Special Issue on Emerging Themes on Inf. Theory and Bayesian Approach)*, vol. 5, no. 3, pp. 281–328, 2010.
- [14] K. P. Burnham and D. Anderson, *Model Selection and Multi-Model Inference*. New York: Springer, Jul. 2002.
- [15] Z. Ghahramani and G. E. Hinton, "The EM algorithm for mixtures of factor analyzers," Dept. Comput. Sci., Univ. Toronto, Toronto, ON, Canada, Tech. Rep. CRG-TR-96-1, 1996.
- [16] R. Redner and H. Walker, "Mixture densities, maximum likelihood and the EM algorithm," *SIAM Rev.*, vol. 26, no. 2, pp. 195–239, 1984.
- [17] L. Xu, "Bayesian-Kullback coupled Ying-Yang machines: Unified learnings and new results on vector quantization," in *Proc. ICONIP'95*, Beijing, China, 1995, pp. 977–988.
- [18] L. Xu, "Bayesian Ying Yang learning," *Scholarpedia* 2(3):1809 [Online]. Available: http://scholarpedia.org/article/Bayesian_Ying_Yang_Learning_2007
- [19] B. Chen, H.-W. Liu, L. Yuan, and Z. Bao, "Adaptively segmenting angular sectors for radar HRRP automatic target recognition," *EURASIP J. Adv. Signal Process.*, 2008, Article ID 641709.
- [20] W. G. Carrara, R. S. Goodman, and R. M. Majewski, *Spotlight Synthetic Aperture Radar—Signal Processing Algorithms*. Norwood, MA: Artech House, 1995.
- [21] J. B. Keller, "Geometrical theory of diffraction," *J. Opt. Soc. Amer.*, vol. 52, no. 2, pp. 116–130, 1962.
- [22] B. Chen, H.-W. Liu, and Z. Bao, "Optimizing the data-dependent kernel under a unified kernel optimization framework," *Pattern Recogn.*, vol. 41, no. 6, pp. 2107–2119, 2008.
- [23] H.-W. Liu and Z. Bao, "Radar HRR profiles recognition based on SVM with power-transformed-correlation kernel," *Lecture Notes in Comput. Sci.*, vol. 3174(I), pp. 531–536, 2004.
- [24] P. Kosir and R. DeWalt, "Feature alignment techniques for pattern recognition," in *Proc. IEEE Nat. Conf. Aerosp. Electron. NAECON 1994*, Dayton, OH, 1994, vol. 1, pp. 128–132.
- [25] R. G. Pita, M. R. Zurera, R. V. Bueno, F. Lpez-Ferreras, and M. U. Manso, "Study of the performance of different alignment methods in pattern recognition," in *Proc. Eur. Signal Process. Conf. EUSIPCO*, Florence, Italy, 2006.
- [26] J. P. Zwart, R. van der Heiden, S. Gelsema, and F. Groen, "Fast translation invariant classification of HRR range profiles in a zero phase representation," *Proc. Inst. Electr. Eng.—Radar, Sonar, Navigat.*, vol. 150, no. 6, pp. 411–418, 2003.
- [27] H. Akaike, "Factor analysis and AIC," *Psychometrika*, vol. 52, no. 3, pp. 317–332, 1987.
- [28] L. Xu, "Data smoothing regularization, multi-sets-learning, and problem solving strategies," *Neural Netw.*, vol. 16, no. 5–6, pp. 817–825, 2003.



Lei Shi received the B.Eng. degree in computer science and technology from the University of Science and Technology of China, Hefei, in 2005.

He is currently a Ph.D. student with the Department of Computer Science and Engineering, Chinese University of Hong Kong, Shatin, Hong Kong. His research interests include statistical learning and neural computing.



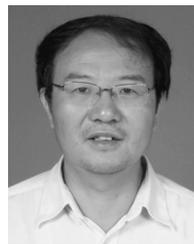
Penghui Wang received the B.Eng. degree in communication engineering from National University of Defense Technology, Changsha, China, in 2005.

He is currently working toward the Ph.D. degree with the National Laboratory of Radar Signal Processing, Xidian University. His research interests include radar automatic target recognition (RATR), radar signal processing, and pattern recognition.



Hongwei Liu (M'00) received the B.Eng. degree from Dalian University of Technology in electronic engineering in 1992, and the M.Eng. and Ph.D. degrees in electronic engineering from Xidian University, Xi'an, China, in 1995 and 1999, respectively.

He is currently the Director and a Professor with the National Laboratory of Radar Signal Processing, Xidian University. His research interests include radar automatic target recognition, radar signal processing, and adaptive signal processing.



Lei Xu (SM'94–F'01) received the Ph.D. degree from Tsinghua University, Beijing, China, in 1987.

He is now a Chair Professor with the Chinese University of Hong Kong, a Chang Jiang Chair Professor with Peking University, China, and an Honorary Professor with Xidian University. He has published a number of well-cited papers on neural networks, statistical learning, and pattern recognition, (e.g., received several thousand citations according to SCI-Expanded and Google Scholar).

Dr. Xu has received some Chinese national academic awards (e.g., 1993 National Nature Science Award) and international awards (e.g., 1995 INNS Leadership Award and the 2006 APNNA Outstanding Achievement Award). He has been elected a Fellow of International Association for Pattern Recognition (2002-), and a member of European Academy of Sciences (2002-).



Zheng Bao (M'80–SM'90) received the B.Eng. degree from the Communication Engineering Institution of China in 1953.

Currently, he is a Professor with Xidian University, Xi'an, China. He is the author or coauthor of six books and has published more than 300 papers. His current research work focuses on the areas of space-time adaptive processing, radar imaging (SAR/ISAR), radar automatic target recognition, over-the-horizon radar signal processing, and passive coherent location.

Prof. Bao is a member of the Chinese Academy of Science.