

# Radar HRRP Recognition Based on Discriminant Information Analysis

JIANSHENG FU<sup>1</sup>      XIAOHONG DENG<sup>2</sup>      WANLIN YANG<sup>1</sup><sup>1</sup>College of Electronic Engineering, University of Electronic Science and Technology of China  
No.4, Section 2, North Jianshe Road, Chengdu  
CHINA<sup>2</sup>Sichuan Post and Telecommunication College  
No.536, Jing kang Road, Chengdu  
CHINA[fujiansheng2010@126.com](mailto:fujiansheng2010@126.com)[50561736@qq.com](mailto:50561736@qq.com)[wlyang@uestc.edu.cn](mailto:wlyang@uestc.edu.cn)

*Abstract:* In radar HRRP target recognition, the quality and quantity of Discriminant Information (DI), which one is more important? Accompanied with this issue, the paper proceeds to delve into DI analysis, and accordingly, three fundamental DI extraction models are proposed, i.e., PGA, PIB and AIB. Among these models, PIB and AIB both aim to obtain Between-class DI (B-DI) from individual standpoints while PGA obtains Among-class DI (A-DI) from a general viewpoint; PGA and PIB are both used for passive recognition while AIB for active recognition. In order to externalize these models, we conduct Generalized Discriminant Analysis (GDA) into them, and two GDA variations come forth, i.e., PIB-based GDA (PIB-GDA) and AIB-based GDA (AIB-GDA). Theoretical analyses and experimental results indicate as follows. Firstly, although PGA prevails in pattern recognition, but the implementation prospect is hardly optimistic on account of the weak anti-fading ability of A-DI. Compared with PGA, PIB and AIB are both more suitable to multi-class discrimination due to the relative stability of B-DI. Secondly, in general, PIB-GDA is inferior to AIB-GDA but superior to GDA to many challenges, such as computational efficiency, target quantity, aspect and sample variation, noise disturbance, etc.

*Key-Words:* discriminant information, feature extraction, target recognition, generalized discriminant analysis.

## 1 Introduction

Mathematical speaking, mutual information (MI) represents the reduction of uncertainty in one random variable when the value of another related random variable is known, so it can expose a dependency between two random variables even when that dependency is nonlinear [1]. As a fundamental concept of information theory, MI always appears in pattern recognition for feature extraction [2]–[4], whereas it is impractical to calculate MI in high dimensions due to the unfeasible number of samples and long operating time [5], [6]. In order to avoid this complex time-consuming process, a simple concept, namely, Discriminant Information (DI), is proposed for radar HRRP target recognition, here DI is defined as the information which can denote part of its class' characters and can be used to discriminate its class from others. The essence of DI analysis is to select and utilize special DI components via applying a certain Feature Extraction Method (FEM), thereby obtaining the related Feature Templates (FTs) which contain the appointed DI components. Not only can different DI utilizations result in different classification performances, but also a proper DI analysis is very conducive to feature extraction. Nevertheless

we are usually subject to many statistical FEMs but neglect the DI components they aim for, and as a result, the thought of DI analysis is seldom accorded due respect from HRRP-based Radar Automatic Target Recognition (RATR) communities.

A raw HRRP is the amplitude of the coherent summations of the complex time returns from target scatters in each range resolution cell, which represents the projection of the complex returned echoes from the target scattering centers onto the radar Line Of Sight (LOS). Among several kinds of wideband radar target signatures, HRRP is a promising signature and easier to be acquired, but highly sensitive to time-shift and target-aspect variation, so how to extract robust and effective feature from it becomes a key problem in HRRP-based RATR. During past years, many RATR communities confirmed that raw HRRP contains some target structure signatures, such as scatter distribution, size, etc [7]–[10], and accordingly, a number of pretreatment methods, such as discrete Fourier transform [11], wavelet transform [12], etc, have been adopted to extract its time-shift invariants, such as amplitude, spectra, etc [3], [4], [8], [9], [13]–[17], as the feature dataset for the discriminant analysis and recognition.

In this paper, we mainly concern with discriminant analysis. Generally speaking, the chief aim of discriminant analysis is to obtain the FTs which contain DI as much as possible, and synchronously, to adjust the recognition speed as per the practical demand. There are multifarious FEMs for discriminant analysis [3], [4], [8], [9], [13]–[33]. As the traditional decomposition methods, Principal Component Analysis (PCA) and Kernel PCA (KPCA) always appear in pattern recognition [18]–[21], but both obtain feature information from an energy viewpoint instead of DI, so they overlook the redundancy information in principal components. As the classical discriminant methods, Linear Discriminant Analysis (LDA) and kernel Generalized Discriminant Analysis (GDA) have been widely applied for feature extraction and dimensionality reduction [22]–[26], but both are prevalently designed to obtain Among-class DI (A-DI) at the expense of Between-class DI (B-DI), thereby losing the relative DI components between two random targets and resulting in a dissatisfied classification performance sometimes.

Therefore, DI continually decreases along the whole feature extraction process. In raw data preprocess, we can obtain the time-shift invariant DI but abandon the useful DI in variant components. In discriminant analysis process, we prefer to A-DI while A-DI is only one subset of DI. Furthermore, different DI components always have different stabilities of information content, and different FEMs usually have different abilities of DI extraction. When the number of classes is increasing, some DI components, such as B-DI, can keep relatively stable, while others, such as A-DI, lessen sharply. Even to an excellent FEM which can almost make full use of a certain DI component, if this component only occupies slight part of DI, the recognition still dissatisfies us, so rational selection and utilization of DI is vital in radar HRRP target recognition. Usually, different recognition styles may lead to different DI selections and utilizations, but we are accustomed to carrying out a recognition process from the standpoint of a certain spectator who seems “out of this collectivity”. We define this recognition style as passive recognition. Admittedly, passive recognition suffers from partial loss of B-DI. In order to obtain more B-DI, a new recognition style called active recognition is proposed in which each class is personified to perform recognizing behavior on behalf of her related target [34].

All in all, DI is the most important factor which directly determines the final recognition result, nevertheless it is always neglected by the fact that it has categories and each category has limited capacity. In this paper, sorted by DI extraction area, DI selection standpoint and recognition style, there are three

proposed DI extraction models, i.e., PGA, PIB and AIB<sup>1</sup>. Due to the huge storage requirement and computation burden in radar HRRP target recognition, some computational analysis is provided in the experiment. Additionally, topological diagram is adopted for DI analysis [35], and 1-NN ruler is applied for template match [29].

The rest is organized as follows. In Section 2, we proceed to delve into DI analysis, and three fundamental DI extraction models are proposed, i.e., PGA, PIB and AIB. In Section 3, we conduct GDA into the models and two GDA variations come forth, i.e., PIB-based GDA (PIB-GDA) and AIB-based GDA (AIB-GDA). In Section 4, three recognition processes are designed corresponding to the three DI extraction models. In Section 5, a seven-simulated-plane system and a three-measured-plane system are offered to evaluate the performances. Finally, some conclusions are made in Section 6.

## 2 Discriminant Information Analysis

Throughout this paper, we assume that the given training HRRP space  $\{\mathbf{X}|\mathbf{x}_i, i=1,2,\dots,M\}$  with  $M$  HRRPs, and each HRRP is represented as a  $n$ -dimensional real column vector. Let  $g$  be the total number of classes,  $m_\xi$  be the  $\xi^{\text{th}}$  class' HRRP number,  $\{\mathbf{X}_\xi|\mathbf{x}_{\xi,j}, j=1,2,\dots,m_\xi\}$  denote the  $\xi^{\text{th}}$  class' HRRP subset, and  $\mathbf{m}$  be the HRRP number vector, thus we have  $\mathbf{m}=[m_1 \ m_2 \ \dots \ m_g]$ ,  $M=\sum_{\xi=1}^g m_\xi$  and  $\mathbf{X}=[\mathbf{X}_1 \ \mathbf{X}_2 \ \dots \ \mathbf{X}_g]$ . The details of HRRP datasets for experiments are offered in Section 5. All the implementations were based on MATLAB 7.1 and performed on a 3.06-GHz Pentium(R)-4 machine which runs Windows XP operation system and has 1-GB EMS memory.

When a process is difficult to understand but can be vividly described by a function, note that this function can't be used for calculation, so we define it as the abstract function of this process. When an algorithm needs many complicated formulas to demonstrate its calculation process, usually, a single function can be used to represent that algorithm, and we define it as synthesis function of that algorithm.

---

<sup>1</sup> PGA is defined as the DI extraction model in which A-DI is extracted from a general viewpoint and used for passive recognition. PIB is defined as the DI extraction model in which B-DI is extracted from the standpoints of individual classes and used for passive recognition. AIB is defined as the DI extraction model in which B-DI is also extracted from the standpoints of individual classes but used for active recognition.

## 2.1 Definition and Analysis

HRRP feature information is defined as the information that is contained in HRRPs and can be used to describe the characters of its related class, including, but not limited to, size and scatter distribution. According to the definition, DI can be considered as a subset of feature information, and doesn't include noise and other useless or harmful information.

### 2.1.1 Components of DI

In this paper, a simple DI structure is presented in which three fundamental components, i.e., absolute, relative and futile components, are defined according to their different discriminant abilities. Given a system with  $g$  classes, absolute DI component is this kind of DI which belongs to only one class, and can be used to discriminate its related class from all the others, while futile DI component belongs to all the classes, and can't be used to discriminate any class from others. Relative component is defined as the rest which excludes the absolute and futile components. Let  $G_\xi$ ,  $G_{a,\xi}$ ,  $G_{f,\xi}$  and  $G_{r,\xi}$ , respectively, denote class  $\xi$ 's DI aggregate and her absolute, futile and relative components, thus they are given by

$$\begin{cases} G_{a,\xi} = G_\xi - \bigcup_{i \in D(\xi,g)} (G_\xi \cap G_i) \\ G_{f,\xi} = \bigcap_{j=1}^g G_j \\ G_{r,\xi} = G_\xi - (G_{a,\xi} \cup G_{f,\xi}) = \bigcup_{\gamma \in D(\xi,g)} G_{r,\xi,\gamma} \end{cases}, \quad (1)$$

$$(\xi = 1, 2, \dots, g)$$

where  $G_{r,\xi,\gamma}$  is class  $\xi$ 's relative DI component obtained from class  $\gamma$ 's standpoint, and  $D(\xi, g)$  is a subset function for class selecting. They are given by

$$\begin{cases} G_{r,\xi,\gamma} = \begin{cases} G_{r,\xi} & \gamma = \xi \\ G_\xi - G_{a,\xi} - (G_\xi \cap G_\gamma) & \gamma \in D(\xi, g) \end{cases} \\ D(\xi, g) = \{1, 2, \dots, \xi - 1, \xi + 1, \dots, g - 1, g\} \end{cases}. \quad (2)$$

$$(\xi = 1, 2, \dots, g)$$

For example, as shown in Fig. 1 (a, b), there are two topological diagrams to demonstrate the DI components of a three-class system. From Fig. 1 (a), we can discern the absolute, relative and futile DI components clearly. Due to the close-set property between class 2 and 3, compared with  $G_2 \cap G_3$ , their absolute and relative DI components, i.e.,  $G_{a,2}$ ,  $G_{r,2,3}$ ,  $G_{a,3}$  and  $G_{r,3,2}$ , are all relatively small, which can apparently increase the recognition difficulty

between class 2 and 3. Let's consider the DI components from class 1's standpoint. As shown in Fig. 1 (b), the absolute and relative components, i.e.,  $G_{a,1}$  and  $G_{r,1}$ , are both beneficial to her discrimination, while the futile component  $G_{f,1}$  is useless to her discrimination. If class 1's feature information contains the element of  $G_{a,2}$ ,  $G_{a,3}$  or  $G_2 \cap G_3 - G_{f,1}$ , obviously, it is harmful to her discrimination, so we called it bad DI component from class 1's standpoint.

### 2.1.2 Types of DI

On account of different roles in application, there are two fundamental DI types, i.e., A-DI and B-DI. A-DI is defined as the DI subset which can be used to discriminate its related class from all the others, while B-DI is defined as the DI subset which can be used to discriminate its related class from at least one of the others, that is,

$$\begin{cases} G_{A,\xi} = G_{a,\xi} \\ G_{B,\xi} = G_{a,\xi} \cup G_{r,\xi} = \bigcup_{k \in D(\xi,g)} G_{B,\xi,k} \end{cases}, \quad (3)$$

$$(\xi = 1, 2, \dots, g)$$

where  $G_{A,\xi}$  denotes class  $\xi$ 's A-DI,  $G_{B,\xi}$  denotes class  $\xi$ 's B-DI, and  $G_{B,\xi,\gamma}$  denotes class  $\xi$ 's B-DI subset from the standpoint of class  $\gamma$ , which is given by

$$G_{B,\xi,\gamma} = \begin{cases} G_{B,\xi} & \gamma = \xi \\ G_\xi - (G_\xi \cap G_\gamma) & \gamma \in D(\xi, g) \end{cases}. \quad (4)$$

$$(\xi = 1, 2, \dots, g)$$

Hence we can conclude that, in mathematics, one class' B-DI obtained from her own standpoint is equivalent to the union obtained from the standpoints of all other individuals. As indicated by (3), since A-DI is equal to absolute DI component, B-DI can be considered as the union of A-DI and relative DI component. Actually, A-DI is the most usable DI, which always receives intensity attention from HRRP-based RATA communities [17]–[23], [29], [38]. Compared with A-DI, the relative DI component can't be used to discriminate its class from all the others, but at least, it can be used to discriminate its related class from one of the others. As a result, it decreases the quantity of discriminant classes in some sense, and reduces the recognition difficulty to some extent, so it is also beneficial to discrimination. Furthermore, as analyzed in Section 5.2.1, making use of B-DI doesn't increase, or acceptably increases, or even decreases the storage requirement and computation burden.

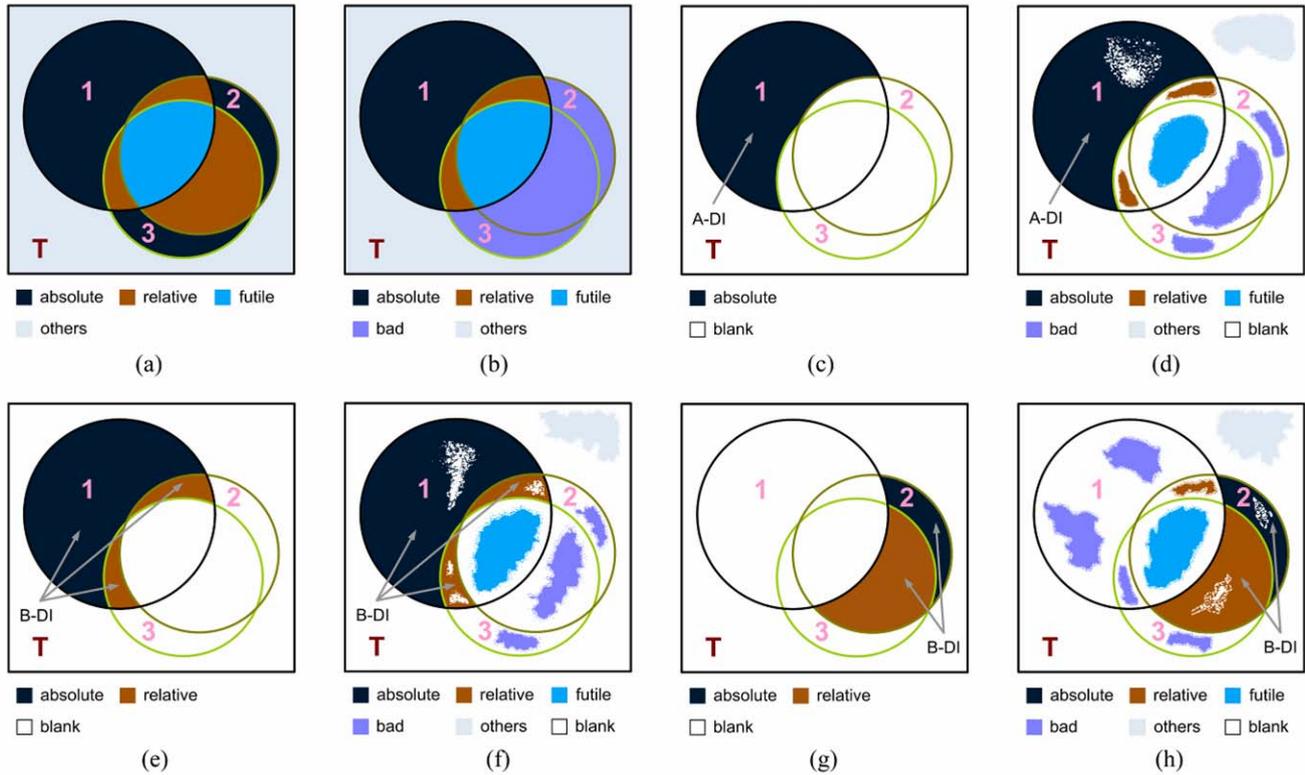


Fig. 1: Using topological diagrams to demonstrate the DI components of a 3-class system. (a) DI components of three classes in theory. (b) DI components of class 1 in theory. (c) PGA: class 1's theoretical A-DI from a general viewpoint. (d) PGA: class 1's practical DI components from a general viewpoint. (e) PIB or AIB: class 1's theoretical B-DI from class 1's standpoint. (f) PIB or AIB: class 1's practical DI components from class 1's standpoint. (g) AIB: class 2's theoretical B-DI from class 1's standpoint. (h) AIB: class 2's practical DI components from class 1's standpoint. (Practical DI components denote the DI components which are obtained by a certain FEM. These sketch maps are just used to demonstrate DI analysis, and aren't obtained from the measured or simulated experiments.)

### 2.1.3 Extraction of DI

Compared with MI that can be obtained via calculating probability density, differential entropy and other related parameters [1]–[6], it is difficult to obtain a certain pure DI component from original dataset, but some FEMs can be designed to obtain the special FTs which contain the appointed DI components, and therefore, their abilities of DI extraction can be estimated by the recognition performance. Nevertheless, from the viewpoint of statistics, it is impossible or impractical for a certain FEM to obtain the FTs which contain the whole selected DI components, and moreover, it is reasonable for the FTs containing other DI components, or even other useless or harmful information, so an abstract function for DI extraction can be defined by

$$\begin{cases} \{\Theta_{a,\xi}, \Theta_{r,\xi}, \Theta_{f,\xi}, O_{\xi}\}_{FT} = \mathfrak{R}^{\varphi}(\mathbf{X}_{\xi}, G_{\Omega,\xi}) \\ \Theta_{\tau,\xi} \subset G_{\tau,\xi} \quad (\tau = a, r, f) \end{cases}, \quad (5)$$

$$(\xi = 1, 2, \dots, g)$$

where  $\mathfrak{R}^{\varphi}(\mathbf{X}_{\xi}, G_{\Omega,\xi})$  denotes an abstract function for FEM  $\mathfrak{R}$  to obtain class  $\xi$ 's DI components corresponding to a given DI extraction model  $\varphi$ ,  $\Omega$  denotes the DI type, such as A-DI and B-DI, and  $\{\Theta_{a,\xi}, \Theta_{r,\xi}, \Theta_{f,\xi}, O_{\xi}\}_{FT}$  is defined as the information aggregate in class  $\xi$ 's FTs, here  $\Theta_{a,\xi}$ ,  $\Theta_{r,\xi}$ ,  $\Theta_{f,\xi}$  and  $O_{\xi}$ , respectively, denote absolute, relative, futile DI components and other information obtained by FEM  $\mathfrak{R}$ . Note that, by (5), FEM  $\mathfrak{R}$  aims to class  $\xi$ 's theoretical  $G_{\Omega,\xi}$ , but in practice it obtains  $\{\Theta_{a,\xi}, \Theta_{r,\xi}, \Theta_{f,\xi}, O_{\xi}\}_{FT}$ .

### 2.2 DI Extraction Models

Based on the analysis above, we can find that, more DI components may lead to more DI extraction area, and better FEM for a certain DI component can obtain more DI of this kind. So there are two main aspects to deal with it, that is to say, by what model to design DI extraction, and by which FEM to externalize the model. In this subsection, we mainly

Table 1: Abbreviative notations of three DI extraction models

| abbreviative notation | recognition style | selection standpoint | extraction area |
|-----------------------|-------------------|----------------------|-----------------|
| PGA                   | passive           | general              | A-DI            |
| PIB                   | passive           | individual           | B-DI            |
| AIB                   | active            | individual           | B-DI            |

concern with the models, and their actualizations are offered in Section 3. According to various DI extraction areas and selection standpoints, there are three fundamental DI extraction models, i.e., PGA, PIB and AIB, here the DI extraction area is defined as the total scale of the selected DI components that the model aims to obtain. Showing in Table 1 is the three abbreviative notations corresponding to the three DI extraction models, which are sorted by DI extraction area, DI selection standpoint and recognition style.

### 2.2.1 PGA

PGA is defined as the DI extraction model in which A-DI is extracted from a general viewpoint of collectivity and used for passive recognition. According to (5), its abstract DI extraction function is

$$\left\{ \begin{array}{l} \{\Theta_{a,\xi}, \Theta_{r,\xi}, \Theta_{f,\xi}, O_{\xi}\}_{FT} = \mathfrak{R}^{PGA}(\mathbf{X}_{\xi}, G_{A,\xi}) \\ \Theta_{A,\xi}^{PGA} = \Theta_{a,\xi} \subset G_{A,\xi} \end{array} \right., \quad (\xi = 1, 2, \dots, g) \quad (6)$$

where  $\Theta_{A,\xi}^{PGA}$  denotes class  $\xi$ 's practical A-DI extracted by FEM  $\mathfrak{R}$ . For example, there are two sketch maps for PGA in Fig. 2 (c, d), from which class 1's theoretical A-DI and practical DI components are demonstrated, respectively, by two topological diagrams.

One view worth pointing out is that, PGA prevails in statistical pattern recognition despite the inefficiency of DI extraction and utilization innately existing in it. With respect to the FTs extracted by a FEM under certain global optimization criterion, they can be statistically discriminated on each projection axis of feature extraction [28], [29], so from the viewpoint of statistics, each single value in them can be used to discriminate its class from all the others. According to Section 2.1.2, the value contains the element of A-DI, and the FTs contain the practical A-DI. During past years, there were many statistical FEMs proposed for A-DI extraction by PGA mode, such as LDA, GDA, the Direct-LDA (DLDA), the Modified GDA (MGDA), the Kernel Direct Discriminant Analysis (KDDA), etc [2], [5], [17], [22]–

[27]. Even though these FEMs have been proven successful in many applications, their usages are still a bit inanimate, that is, the prevalent applications of them were usually designed to obtain the A-DI among all classes while neglecting the B-DI between two random classes, thereby potentially losing the relative DI components and possibly resulting in a dissatisfied recognition performance.

Another view worth emphasizing is that, PGA is crag-fast in multi-target recognition. Although PGA originally aims to solve a multi-class discriminant problem, but with the number of classes increasing, the content of A-DI reduces urgently in theory, thereby sharply increasing the difficulty and uncertainty of information extraction in practice. The detailed analysis is offered in Section 2.3 by a simple mathematical model.

### 2.2.2 PIB

Admittedly, A-DI can be used to discriminant one class from all the others, but it only occupies part of DI, and can be considered as a subset of B-DI, so making use of B-DI may perform better than A-DI does. In order to obtain B-DI, a new DI extraction model called PIB is proposed in which each class' B-DI is selected from her own standpoint and also used for passive recognition. The abstract DI extraction function is given by

$$\left\{ \begin{array}{l} \{\Theta_{a,\gamma}, \Theta_{r,\gamma}, \Theta_{f,\gamma}, O_{\gamma}\}_{FT} = \mathfrak{R}^{PIB}(\mathbf{X}_{\xi}, G_{B,\xi,\gamma}) \quad \text{s.t. } \xi = \gamma \\ \Theta_{B,\gamma}^{PIB} = (\Theta_{a,\gamma} \cup \Theta_{r,\gamma}) \subset G_{B,\gamma} \end{array} \right., \quad (\gamma = 1, 2, \dots, g) \quad (7)$$

where  $\Theta_{B,\gamma}^{PIB}$  denotes class  $\gamma$ 's practical B-DI obtained by FEM  $\mathfrak{R}$  from her own standpoint. Showing in Fig. 1 (e, f) are the two sketch maps for PIB, in which class 1's theoretical B-DI and practical DI components are demonstrated by two topological diagrams respectively.

Although (7) can be considered as an intuitive notation for PIB in theory, but it is not easy to realize in application. Since one class' B-DI obtained from her own standpoint is equivalent to the union obtained from the standpoints of all the others, a variation for PIB is provided by

$$\left\{ \begin{array}{l} \{\Theta_{a,\gamma,\xi}, \Theta_{r,\gamma,\xi}, \Theta_{f,\gamma,\xi}, O_{\gamma,\xi}\}_{FT} = \mathfrak{R}_{\xi}^{PIB}(\mathbf{X}_{\gamma}, G_{B,\gamma,\xi}) \\ \Theta_{B,\gamma,\xi}^{PIB} = (\Theta_{a,\gamma,\xi} \cup \Theta_{r,\gamma,\xi}) \subset G_{B,\gamma,\xi} \\ \Theta_{B,\gamma}^{PIB} = \bigcup_{k \in D(\gamma,g)} \Theta_{B,\gamma,k}^{PIB} \subset \bigcup_{k \in D(\gamma,g)} G_{B,\gamma,k} \Leftrightarrow G_{B,\gamma} \end{array} \right., \quad (\xi \in D(\gamma,g), \gamma = 1, 2, \dots, g) \quad (8)$$

where  $\Theta_{B,\gamma,\xi}^{PIB}$  denotes class  $\gamma$ 's practical B-DI obtained by FEM  $\mathfrak{R}$  from class  $\xi$ 's standpoint.

Let's analyze the DI extraction areas of PGA and PIB. According to the definitions of A-DI and B-DI, although the DI extraction area of PIB is larger than that of PGA, but the quality of PGA's DI extraction area is higher than PIB's. In some sense, the DI extraction areas of PGA and PIB can be considered as two different reflections of DI, that is, PGA emphasizes the quality while PIB prefers to the size; PGA emphasizes common differentia among all classes while PIB prefers to individual differentia between two random classes.

### 2.2.3 AIB

Compared with PIB, individual AIB also aims for B-DI, but all class' B-DI is selected from one class' standpoint and used for active recognition. Here is the abstract DI extraction function of individual AIB:

$$\left\{ \begin{array}{l} \left\{ \left\{ \Theta_{a,\gamma}, \Theta_{r,\gamma}, \Theta_{f,\gamma}, O_{\gamma} \right\}_{FT} = \mathfrak{R}_{\gamma}^{PIB}(\mathbf{X}_{\xi}, G_{B,\xi,\gamma}) \right. \\ \left. \Theta_{B,\xi,\gamma}^{AIB} = \Theta_{B,\gamma}^{AIB} = (\Theta_{a,\gamma} \cup \Theta_{r,\gamma}) \subset G_{B,\gamma} \right. \\ \left. \left\{ \left\{ \Theta_{a,\xi,\gamma}, \Theta_{r,\xi,\gamma}, \Theta_{f,\xi,\gamma}, O_{\xi,\gamma} \right\}_{FT} = \mathfrak{R}_{\gamma}^{AIB}(\mathbf{X}_{\xi}, G_{B,\xi,\gamma}) \right. \right. \\ \left. \left. \Theta_{B,\xi,\gamma}^{AIB} = (\Theta_{a,\xi,\gamma} \cup \Theta_{r,\xi,\gamma}) \subset G_{B,\xi,\gamma} \right. \right. \\ \left. \left. (\gamma = 1, 2, \dots, g), \right. \right. \end{array} \right. \quad \xi = \gamma \quad \xi \in D(\gamma, g) \quad (9)$$

where  $\Theta_{B,\xi,\gamma}^{AIB}$  denotes class  $\xi$ 's practical B-DI obtained by FEM  $\mathfrak{R}$  from class  $\gamma$ 's standpoint. For example, showing in Fig. 1 (e-h) are the four sketch maps from class 1's standpoint, in which class 1 and class 2's theoretical and practical DI components are demonstrated, respectively, by four diagrams.

Let's consider class  $\xi$ 's DI extraction areas from different individual standpoints. From her own standpoint, the DI extraction area is  $G_{B,\xi}$ , while from class  $\gamma$ 's standpoint, the DI extraction area is  $G_{B,\xi,\gamma}$ . We compare the two areas by

$$\left\{ \begin{array}{l} G_{B,\xi,\gamma} = G_{B,\xi} \quad \gamma = \xi \\ G_{B,\xi,\gamma} \subset G_{B,\xi} \quad \gamma \in D(\xi, G) \end{array} \right. \quad (\gamma = 1, 2, \dots, g), \quad (10)$$

so we can find that the DI extraction area of PIB is larger than that of individual AIB, that is,

$$\left\{ \begin{array}{l} G_{B,L,\gamma}^{AIB} \subset G_B^{PIB} \\ \text{s.t.} \left\{ \begin{array}{l} G_B^{PIB} = \bigcup_{i=1}^g G_{B,i} \\ G_{B,L,\gamma}^{AIB} = \bigcup_{j=1}^g G_{B,j,\gamma} \end{array} \right. \quad (\gamma = 1, 2, \dots, g), \end{array} \right. \quad (11)$$

where  $G_B^{PIB}$  denotes PIB's total DI extraction area,

and  $G_{B,L,\gamma}^{AIB}$  denotes individual AIB's total DI extraction area from class  $\gamma$ 's standpoint.

An explain to this phenomenon is given as that, in individual AIB, class  $\gamma$  is personified on behalf of her related class, so there are some individual bias and selfness inevitably existing in her judgement [34]. It is reasonable for her to select the B-DI relating to her class while neglecting the rest. For example, as shown in Fig. 1 (g), from class 1's standpoint, class 2's B-DI subset  $G_{B,2,1}$  is selected while  $G_{B,2,3} - G_{A,2}$  is abandoned. In summary, individual AIB's DI extraction area is smaller than PIB's, besides which there are some personal biases in individual AIB, so by the same FEM, PIB should perform better than individual AIB does.

Fortunately, not only is AIB designed for individual discrimination, but also for global discrimination. Let's analyze global AIB's total DI extraction area  $G_B^{AIB}$  by

$$\begin{aligned} G_B^{AIB} &= \bigcup_{k=1}^g G_{B,L,k}^{AIB} = \bigcup_{k=1}^g \left( \bigcup_{i=1}^g G_{B,i,k} \right) \\ &= \bigcup_{i=1}^g \left( \bigcup_{k=1}^g G_{B,i,k} \right) = \bigcup_{i=1}^g \left( \bigcup_{k \in D(\xi, g)} G_{B,i,k} \right), \quad (12) \\ &= G_B^{PIB} \end{aligned}$$

that is to say further, global AIB is equivalent to PIB in theory.

## 2.3 Comparison of Information Content

As proven above, since global AIB is equivalent to PIB in theory, we only analyze the change of DI content in the three models, i.e., PGA, PIB and individual AIB. Due to the range and diversity of the targets in real applications, there are many formidable obstacles to estimate their DI performances, but an ideal mathematical model can be designed to test the general performance. In this paper, a mathematical model is provided in which each class has the same statistical property and can be considered as an isotropy subset:

$$\left\{ \begin{array}{l} f\left(\bigcap_{i=1}^k G_{\xi_i}\right) = \rho(1-\sigma)^{k-1} \\ f\left(\bigcup_{i=1}^k G_{\xi_i}\right) = \rho\left(1 + \sigma \frac{1-\sigma^{k-1}}{1-\sigma}\right) \end{array} \right. \quad \left( \begin{array}{l} k = 1, 2, \dots, g \\ 0 < \sigma < 1 \end{array} \right)$$

$$\text{s.t.} \quad \left\{ \begin{array}{l} \xi_1, \xi_2, \dots, \xi_k \in \{1, 2, \dots, g\} \\ \text{and } \xi_1 \neq \xi_2 \neq \dots \neq \xi_k \end{array} \right., \quad (13)$$

where  $f(*)$  is defined as an abstract function to obtain the DI content,  $\rho$  denotes the original infor-

mation content of each class, and  $\sigma$  denotes the utilization rate of DI between two classes.

Under this suppositional model, the content of  $G_{A,\xi}$ ,  $G_{B,\xi}$  and  $G_{B,\xi,\gamma}$  can be obtained by

$$\left\{ \begin{array}{l} \rho_{A,\xi} = f(G_{A,\xi}) = \rho\sigma^{g-1} \\ \rho_{B,\xi} = f(G_{B,\xi}) = \rho(1-(1-\sigma)^{g-1}) \\ \rho_{B,\xi,\gamma} = f(G_{B,\xi,\gamma}) \quad (\xi=1,2,\dots,g) \\ \quad = \begin{cases} \rho(1-(1-\sigma)^{g-1}) & \gamma = \xi \\ \rho\sigma & \gamma \in D(\xi,g) \end{cases} \end{array} \right. \quad (14)$$

where  $\rho_{A,\xi}$ ,  $\rho_{B,\xi}$  and  $\rho_{B,\xi,\gamma}$ , respectively, denote the information content of  $G_{A,\xi}$ ,  $G_{B,\xi}$  and  $G_{B,\xi,\gamma}$ .

Due to the same mathematical model that the three DI extraction models get involved with, the influence of the original information content can be overlooked. Suppose that  $\rho$  is equal to 1, then the average usable DI contents of the three models can be estimated by

$$\left\{ \begin{array}{l} \rho_A^{\text{PGA},g} = \frac{1}{g} \sum_{k=1}^g \rho_{A,k} = \sigma^{g-1} \\ \rho_B^{\text{PIB},g} = \frac{1}{g} \sum_{k=1}^g \rho_{B,k} = 1-(1-\sigma)^{g-1} \\ \rho_{B,I}^{\text{AIB},g} = \frac{1}{g} \sum_{k=1}^g \rho_{B,k,\gamma} = \frac{1-(1-\sigma)^{g-1}}{g} + \frac{g-1}{g}\sigma \end{array} \right. \quad (\gamma=1,2,\dots,g), \quad (15)$$

where  $\rho_A^{\text{PGA},g}$ ,  $\rho_B^{\text{PIB},g}$  and  $\rho_{B,I}^{\text{AIB},g}$ , respectively, denote the appointed DI content of PGA, PIB and AIB.

In some sense, pattern recognition can be considered as to solve a comparison and match problem, so the quality of DI is more important than its content [28], [29]. But in application, it is a bit difficult to extract the relatively slight DI components from a huge dataset, which not only improves the extraction difficulty, but also depresses the information accuracy. As demonstrated by (15), with the number of classes increasing, in terms of DI content, PIB can almost make full use of DI, individual AIB also keeps optimistically stable, while PGA suffers from an exponentially decay.

### 3 Model Actualization Analysis

Generally, in pattern recognition, feature extraction can be considered as the process of deriving useful DI from some original signals, here DI has a more

compact representation and can be adopted for a certain task, such as template match and target recognition. As aforementioned in Section 2.1.3, FT can be considered as a carrier of DI. Since it is difficult to extract pure DI component from the original datasets, a certain FEM  $\mathfrak{R}$  can be applied to obtain the special FTs which contain the designed DI components. As a nonlinear extension of LDA via kernel trick, GDA has been proved often achieving better recognition performance than other nonlinear methods due to the perfect capability of DI extraction [25], [26], so we apply it in the three DI extraction models as the fundamental discriminant analysis unit, and accordingly, two new GDA variations come forth, i.e., PIB-GDA and AIB-GDA.

#### 3.1 Application of GDA in PGA

Given two HRRP subspaces  $\mathbf{X}_w$  and  $\mathbf{X}_v$ , we define the kernel function  $k(\mathbf{x}_{w,i}, \mathbf{x}_{v,j})$  corresponding to a given nonlinear mapping  $\Phi$  by

$$k(\mathbf{x}_{w,i}, \mathbf{x}_{v,j}) = \langle \Phi(\mathbf{x}_{w,i}), \Phi(\mathbf{x}_{v,j}) \rangle \quad (w, v = 1, 2, \dots, g) \quad (16)$$

Note that there are many kernel functions but each one must meet Mercer condition [29]–[32]. We apply Gaussian kernel function for kernel calculating by

$$\left\{ \begin{array}{l} \varphi_{i,j} = k(\mathbf{x}_{w,i}, \mathbf{x}_{v,j}) = \exp\left(-\frac{\|\mathbf{x}_{w,i} - \mathbf{x}_{v,j}\|^2}{\sigma^2}\right) \\ \mathbf{K}_{w,v} = (\varphi_{i,j})_{\substack{i=1,2,\dots,m_w \\ j=1,2,\dots,m_v}} \triangleq \mathbb{k}(\mathbf{X}_w, \mathbf{X}_v) \end{array} \right. \quad (w, v = 1, 2, \dots, g) \quad (17)$$

where  $\mathbf{K}_{w,v}$  denotes the kernel matrix of  $\mathbf{X}_w$  by  $\mathbf{X}_v$ , and the symbol  $\mathbb{k}(*,*)$  denotes the kernel synthesis function. In this paper, we suppose that  $\sigma^2$  is equal to 0.5.

As explored in [25], GDA is originally designed to solve a multi-class discriminant problem for passive recognition. The essence of GDA is to find an optimal transformation by maximizing the between-class distance and minimizing the within-class distance, thereby obtaining the FTs which contain the practical A-DI. According to a variant of Fisher's kernel criterion variant [29]–[31], it aims to solve an optimization problem:

$$J(\mathbf{u}_{\text{opt}}) = \arg \max_{\mathbf{u}} \frac{\mathbf{u}^T(\mathbf{QWQ})\mathbf{u}}{\mathbf{u}^T(\mathbf{QQ})\mathbf{u}}, \quad (18)$$

where the kernel symmetric matrix  $\mathbf{Q}$  is obtained by  $\mathbf{Q} = \mathbf{K} - \mathbf{1}_M \mathbf{K} - \mathbf{K} \mathbf{1}_M + \mathbf{1}_M \mathbf{K} \mathbf{1}_M$ , and the block diagonal matrix  $\mathbf{W} = \text{diag}(\mathbf{1}_{m_1}, \mathbf{1}_{m_2}, \dots, \mathbf{1}_{m_g})$ , here the mean value matrix  $\mathbf{1}_\tau$  is defined as a  $\tau \times \tau$  matrix with terms all equal to  $1/\tau$ , and the kernel matrix  $\mathbf{K}$  is given by  $\mathbf{K} = \mathbb{k}(\mathbf{X}, \mathbf{X})$ .

Let's rank the coefficient vectors  $\mathbf{u}_i$  conforming to their related cost values  $J(\mathbf{u}_i)$  in descending order, and select the front  $g-1$  ones as the Feature Extraction Subspace (FES)  $\mathbf{U}$ . Then for a given HRRP  $\mathbf{y}$ , its feature vector can be obtained by

$$\mathbf{z} = (\mathbb{k}(\mathbf{y}, \mathbf{X}) \times \mathbf{U})^T, \quad (19)$$

where  $\mathbf{z}$  is a  $(g-1)$ -dimensional column vector, which contains  $\mathbf{y}$ 's A-DI and can be considered as  $\mathbf{y}$ 's feature vector.

Now according to (16)–(19), we can acquire the FT matrix  $\mathbf{A}^{\text{PGA}}$  and the feature matrix  $\mathbf{Z}^{\text{PGA}}$  by

$$\left\{ \begin{array}{l} \mathbf{A}^{\text{PGA}} = [\mathbf{A}_1^{\text{PGA}} \quad \mathbf{A}_2^{\text{PGA}} \quad \dots \quad \mathbf{A}_g^{\text{PGA}}] \\ \{\mathbf{A}^{\text{PGA}}, \mathbf{Z}^{\text{PGA}}\} \triangleq \mathbb{S}_{\text{GDA}}(\mathbf{Y}, \mathbf{X}, \mathbf{m}) \end{array} \right., \quad (20)$$

where  $\mathbb{S}_{\text{GDA}}(\mathbf{Y}, \mathbf{X}, \mathbf{m})$  is defined as the GDA synthesis function of PGA for  $\mathbf{X}$  and  $\mathbf{Y}$  to obtain their feature matrixes  $\mathbf{A}^{\text{PGA}}$  and  $\mathbf{Z}^{\text{PGA}}$ ,  $\{\mathbf{A}^{\text{PGA}}, \mathbf{Z}^{\text{PGA}}\}$  denotes an aggregate with two matrix elements ( $\mathbf{A}^{\text{PGA}}$  and  $\mathbf{Z}^{\text{PGA}}$ ) in it, and  $\mathbf{A}^{\text{PGA}}$ 's subset  $\mathbf{A}_\xi^{\text{PGA}}$  is the  $\xi^{\text{th}}$  ( $\xi=1, 2, \dots, g$ ) class' FT matrix. Note that the matrix  $\mathbf{Y}$  only denotes a given HRRP subset which may be comprised of some training HRRPs or test ones, and of course  $\mathbf{Y}$  can be a one vector matrix when it contains only one HRRP.

### 3.2 Application of GDA in PIB

As an extension version of Kernel Fisher Discriminant Analysis (KFDA) [33], not only can GDA be designed for multi-class discrimination, but also used as a KFDA variation to solve the two-class pattern recognition problem. According to Section 2.1, B-DI can be considered as the DI components between two random classes, so we apply GDA to obtain B-DI by

$$\left\{ \begin{array}{l} \mathbf{X}_{\gamma, \xi} = [\mathbf{X}_\gamma \quad \mathbf{X}_\xi] \\ \mathbf{m}_{\gamma, \xi} = [m_\gamma \quad m_\xi] \\ \{\mathbf{B}_{\gamma, \xi}, \mathbf{Z}_{\gamma, \xi}\} = \mathbb{S}_{\text{GDA}}(\mathbf{Y}, \mathbf{X}_{\gamma, \xi}, \mathbf{m}_{\gamma, \xi}) \end{array} \right., \quad (21)$$

$(\xi \in D(\gamma, g), \quad \gamma = 1, 2, \dots, g)$

where  $\mathbf{X}_{\gamma, \xi}$  denotes  $\mathbf{Y}$ 's feature matrix discriminated by class  $\gamma$  and  $\xi$ . Note that  $\mathbf{X}_{\gamma, \xi}$  is only a row vector and each element denotes a feature value of the related HRRP.

Let's consider GDA for two-class discriminant analysis. The two classes are related to each other and can be discriminated on a feature extraction axis. According to (21), when  $\mathbf{Y}$  is equal to  $\mathbf{X}_\gamma$ , its feature matrix  $\mathbf{Z}_{\gamma, \xi}$  contains the practical B-DI  $\Theta_{\text{B}, \gamma, \xi}^{\text{PIB}}$ , which is class  $\gamma$ 's B-DI obtained by GDA from class  $\xi$ 's standpoint, and vice versa. In terms of optimization, when  $\mathbf{Y}$  doesn't belong to class  $\gamma$  or  $\xi$ , the DI component in  $\mathbf{Z}_{\gamma, \xi}$  is defective, and moreover, some additional redundant information enters into  $\mathbf{Z}_{\gamma, \xi}$ , thereby decreasing the discriminant efficiency. Now we arrange  $\mathbf{Z}_{\gamma, \xi}$  from different individual standpoints by

$$\mathbf{Z}_\gamma^{\text{PIB}} = [\mathbf{Z}_{1, \gamma}^T \quad \mathbf{Z}_{2, \gamma}^T \quad \dots \quad \mathbf{Z}_{\gamma-1, \gamma}^T \quad \mathbf{Z}_{\gamma+1, \gamma}^T \quad \dots \quad \mathbf{Z}_{g, \gamma}^T]^T, \quad (22)$$

$(\gamma = 1, 2, \dots, g)$

where  $\mathbf{Z}_\gamma^{\text{PIB}}$  is  $\mathbf{Y}$ 's feature matrix from all other classes' standpoints except class  $\gamma$ .

In accordance with the analysis in Section 2.2,  $\mathbf{Z}_\gamma^{\text{PIB}}$  can be considered as  $\mathbf{Y}$ 's feature matrix from class  $\gamma$ 's standpoint. When  $\mathbf{Y}$  is equal to  $\mathbf{X}_\gamma$ , its feature matrix  $\mathbf{Z}_\gamma^{\text{PIB}}$  becomes class  $\gamma$ 's FT matrix  $\mathbf{A}_\gamma^{\text{PIB}}$ , which contains the practical B-DI  $\Theta_{\text{B}, \gamma}^{\text{PIB}}$ . Note that  $\mathbf{X}_\gamma$  is a subset of  $\mathbf{X}$  and can be labeled by  $\mathbf{m}$  and  $\gamma$ . According to (21), (22), a GDA synthesis function from class  $\gamma$ 's standpoint is given by

$$\{\mathbf{A}_\gamma^{\text{PIB}}, \mathbf{Z}_\gamma^{\text{PIB}}\} \triangleq \mathbb{S}_\gamma^{\text{PIB}}(\mathbf{Y}, \mathbf{X}, \mathbf{m}) \quad (\gamma = 1, 2, \dots, g), \quad (23)$$

where  $\mathbb{S}_\gamma^{\text{PIB}}(\mathbf{Y}, \mathbf{X}, \mathbf{m})$  is defined as PIB's GDA synthesis function from class  $\gamma$ 's standpoint, and used for  $\mathbf{X}_\gamma$  and  $\mathbf{Y}$  to obtain their related feature matrixes  $\mathbf{A}_\gamma^{\text{PIB}}$  and  $\mathbf{Z}_\gamma^{\text{PIB}}$ , here  $\mathbf{A}_\gamma^{\text{PIB}}$  is defined as class  $\gamma$ 's FT matrix obtained from her own standpoint. As the same goes for PIB, a GDA synthesis function is given by

$$\left\{ \begin{array}{l} \mathbf{A}^{\text{PIB}} = [\mathbf{A}_1^{\text{PIB}} \quad \mathbf{A}_2^{\text{PIB}} \quad \dots \quad \mathbf{A}_g^{\text{PIB}}] \\ \mathbf{Z}^{\text{PIB}} = [\mathbf{Z}_1^{\text{PIB}} \quad \mathbf{Z}_2^{\text{PIB}} \quad \dots \quad \mathbf{Z}_g^{\text{PIB}}] \\ \{\mathbf{A}^{\text{PIB}}, \mathbf{Z}^{\text{PIB}}\} \triangleq \mathbb{S}^{\text{PIB}}(\mathbf{Y}, \mathbf{X}, \mathbf{m}) \end{array} \right., \quad (24)$$

where  $\mathbb{S}^{\text{PIB}}(\mathbf{Y}, \mathbf{X}, \mathbf{m})$  is defined as PIB's GDA synthesis function for  $\mathbf{X}$  and  $\mathbf{Y}$ , respectively, to obtain their feature matrix  $\mathbf{A}^{\text{PIB}}$  and  $\mathbf{Z}^{\text{PIB}}$  from an overall standpoint. Obviously, to each HRRP, there are  $g$

feature vectors corresponding to  $g$  classes. Note that the total calculation process to obtain  $\{\mathbf{A}^{\text{PIB}}, \mathbf{Z}^{\text{PIB}}\}$  is defined as PIB-GDA in this paper.

### 3.3 Application of GDA in PGA

As described above, by PIB, class  $\gamma$  only obtains her own FTs, so she can't estimate which class a test HRRP belongs to. If each class is personified to perform recognizing behavior on behalf of her related target, what will happen? In order to discriminate a new HRRP, she needs the FTs of all classes so as to match and classify it. But how can she obtain the FTs of all the classes? In this paper, a new model called AIB is proposed in which each class obtains the FTs of all classes from her own standpoint, and of course, the FTs should contain the appointed DI components. Let's consider the FT matrix obtained by (22) from class  $\gamma$ 's standpoint. When  $\mathbf{Y}$  is equal to  $\mathbf{X}_\xi$ , its feature matrix  $\mathbf{Z}_\gamma^{\text{PIB}}$  becomes class  $\xi$ 's FT matrix  $\mathbf{A}_{\xi,\gamma}^{\text{AIB}}$ , which contains the B-DI  $\Theta_{\text{B},\xi,\gamma}^{\text{AIB}}$  and can be used for template match. By the same, class  $\gamma$  can obtain the FTs of all classes. We construct a GDA synthesis function from class  $\gamma$ 's standpoint by

$$\begin{cases} \mathbf{A}_{1,\gamma}^{\text{AIB}} = [\mathbf{A}_{1,1,\gamma}^{\text{AIB}} \ \mathbf{A}_{1,2,\gamma}^{\text{AIB}} \ \cdots \ \mathbf{A}_{1,g,\gamma}^{\text{AIB}}] \\ \{\mathbf{A}_{1,\gamma}^{\text{AIB}}, \mathbf{Z}_{1,\gamma}^{\text{AIB}}\} \triangleq \mathbb{S}_{1,\gamma}^{\text{AIB}}(\mathbf{Y}, \mathbf{X}, \mathbf{m}) \end{cases} \quad (\gamma=1,2,\dots,g), \quad (25)$$

where  $\mathbb{S}_{1,\gamma}^{\text{AIB}}(\mathbf{Y}, \mathbf{X}, \mathbf{m})$  is defined as class  $\gamma$ 's GDA synthesis function for  $\mathbf{X}$  and  $\mathbf{Y}$  to obtain their related feature matrixes  $\mathbf{A}_{1,\gamma}^{\text{AIB}}$  and  $\mathbf{Z}_{1,\gamma}^{\text{AIB}}$  by individual AIB. Obviously,

$$\mathbf{Z}_{1,\gamma}^{\text{AIB}} = \mathbf{Z}_\xi^{\text{PIB}} \quad \text{s.t. } \gamma = \xi \quad (\xi = 1, 2, \dots, g). \quad (26)$$

As aforementioned in Section 2.2, the reason that the FTs can be used for template match is the DI contained in them, but different kinds of FTs may contain different types of DI, so they may have different recognition performances. Compared with PIB, the DI extraction area of individual AIB is a bit limited due to the personal biases, so some useful DI may be lost. Moreover, some additional redundancy information enters into the FTs of individual AIB, which directly depresses the recognition efficiency.

Let's consider the DI performances of PGA and individual AIB in the measured experiment. As shown in Fig. 2, in order to compare the DI contents, the feature distributions are projected onto the four planes with the same scale. From Fig. 2 (a, b), we can find that, the distance between the distribution centers of An-26 and Jiang by PGA, and the distance

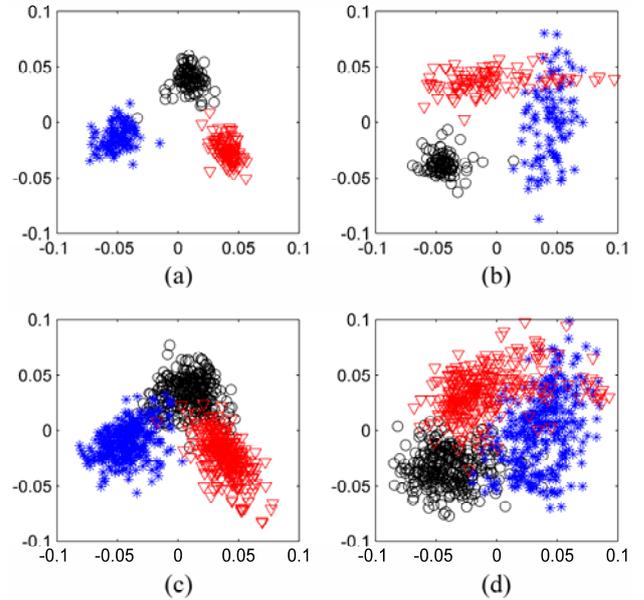


Fig. 2: Feature distributions from different standpoints in the measured experiment. (a) Distributions of 300 FTs by PGA from a general viewpoint. (b) Distributions of 300 FTs by AIB from An-26's standpoint. (c) Distributions of 900 test samples in PGA's FES from a general viewpoint. (d) Distributions of 900 test samples in AIB's FES from An-26's standpoint. ('o': An-26, '\*': Jiang, '∇': Yark-42)

between the distribution centers of An-26 and Yark-42 by PGA, respectively, are shorter than the related distances by AIB, which indicates that the DI content in PGA is smaller than that in individual AIB. As shown in Fig. 2 (a), no matter from horizontal or vertical direction, the FT distribution of the three airplanes can be statistically discriminated due to the A-DI in them. While in Fig. 2 (b), from the vertical direction, although the FT distributions of Yark-42 and An-26 can be easily discriminated, but it's very difficult to discriminate both of them from Jiang, so the practical B-DI contained in them is obvious. Moreover, compared with PGA for passive recognition, individual AIB is designed for active recognition. Although a single class can estimate which class a test HRRP belongs to in some sense, she prefers to judge whether it belongs to her own class or not. Showing in Fig. 2 (c, d) is the projections of test samples obtained by PGA and individual AIB respectively, from which the performance differences between PGA and individual AIB are apparent.

Since individual AIB has many native shortcomings, in order to obtain more B-DI and achieve a general performance, it is necessary to synthesize the

the FT matrixes  $\mathbf{A}_{1,\gamma}^{\text{AIB}}$  by

$$\begin{cases} \mathbf{A}^{\text{AIB}} = [\mathbf{A}_{1,1}^{\text{AIB}} & \mathbf{A}_{1,2}^{\text{AIB}} & \cdots & \mathbf{A}_{1,g}^{\text{AIB}}] \\ \mathbf{Z}^{\text{AIB}} = [\mathbf{Z}_{1,1}^{\text{AIB}} & \mathbf{Z}_{1,2}^{\text{AIB}} & \cdots & \mathbf{Z}_{1,g}^{\text{AIB}}] \\ \{\mathbf{A}^{\text{AIB}}, \mathbf{Z}^{\text{AIB}}\} \triangleq \mathbb{S}^{\text{AIB}}(\mathbf{Y}, \mathbf{X}, \mathbf{m}) \end{cases}, \quad (27)$$

where  $\mathbb{S}^{\text{AIB}}(\mathbf{Y}, \mathbf{X}, \mathbf{m})$  is defined as the GDA synthesis function of global AIB from the standpoints of all classes, which is used for  $\mathbf{X}$  and  $\mathbf{Y}$  to obtain their feature matrixes  $\mathbf{A}^{\text{AIB}}$  and  $\mathbf{Z}^{\text{AIB}}$  respectively. Note that  $\mathbf{A}^{\text{AIB}}$  is a  $(g-1) \times gM$  matrix while  $\mathbf{A}^{\text{PIB}}$  is a  $(g-1) \times M$  matrix. Obviously,  $\mathbf{Z}^{\text{AIB}}$  is equal to  $\mathbf{Z}^{\text{PIB}}$ . The total calculation process to obtain  $\{\mathbf{A}^{\text{AIB}}, \mathbf{Z}^{\text{AIB}}\}$  is defined as AIB-GDA in this paper.

## 4 Recognition and Analysis

Once we obtain the FTs contained the appointed DI components, we can use them for the upcoming recognition. As the simplest and the most attractive pattern classification criterions, 1-NN rule is usually used for template matching and image classifying [33], [34]. Now we apply 1-NN rule as follows.

### 4.1 Passive Recognition

The thought of passive recognition is widely spread in our living. It is reasonable for a person to select the most significant FTs which not only can be used to discriminate a class from all the others, but also doesn't take sides with any class. Note that the recognized classes don't take part in recognition, and the person who discriminates them seems "out of the collectivity", so we defined it as passive recognition. There are two DI extraction models for passive recognition, i.e., PGA and PIB.

Given a test HRRP  $\mathbf{e}$ , we can obtain the FT matrixes  $\{\mathbf{A}_{\xi}^{\text{PGA}} | \mathbf{a}_{\xi,j}^{\text{PGA}}, j=1,2,\dots,m_{\xi}\}$  and its feature vector  $\mathbf{s}_{\mathbf{e}}^{\text{PGA}}$  by (20). Then we apply 1-NN rule to estimate  $\mathbf{e}$  by

$$\begin{cases} d_{\mathbf{e},\xi}^{\text{PGA}} = \min_{j=1,2,\dots,m_{\xi}} \|\mathbf{s}_{\mathbf{e}}^{\text{PGA}} - \mathbf{a}_{\mathbf{e},\xi,j}^{\text{PGA}}\| \\ \theta_{\mathbf{e}}^{\text{PGA}} = \arg \min_{i=1,2,\dots,g} d_{\mathbf{e},i}^{\text{PGA}} \end{cases} \quad (\xi = 1, 2, \dots, g), \quad (28)$$

where  $d_{\mathbf{e},\xi}^{\text{PGA}}$  denotes the Nearest Euclidean Distance (NED) between  $\mathbf{e}$  and class  $\xi$ , and  $\theta_{\mathbf{e}}^{\text{PGA}}$  indicates that  $\mathbf{e}$  belongs to class  $\theta_{\mathbf{e}}^{\text{PGA}}$  by PGA.

Also  $\mathbf{e}$ 's feature matrix  $\{\mathbf{S}_{\mathbf{e}}^{\text{PIB}} | \mathbf{s}_{\mathbf{e},\gamma}^{\text{PIB}}\}$  and the FT matrixes  $\{\mathbf{A}_{\gamma}^{\text{PIB}} | \mathbf{a}_{\gamma,j}^{\text{PIB}}, j=1,2,\dots,m_{\gamma}\}$  can be obtained

by (24). Then we apply 1-NN rule to estimate  $\mathbf{e}$  by

$$\begin{cases} d_{\mathbf{e},\gamma}^{\text{PIB}} = \min_{j=1,2,\dots,m_{\xi}} \|\mathbf{s}_{\mathbf{e},\gamma}^{\text{PIB}} - \mathbf{a}_{\mathbf{e},\gamma,j}^{\text{PIB}}\| \\ \theta_{\mathbf{e}}^{\text{PIB}} = \arg \min_{i=1,2,\dots,g} d_{\mathbf{e},i}^{\text{PIB}} \end{cases} \quad (\gamma = 1, 2, \dots, g), \quad (29)$$

where  $d_{\mathbf{e},\gamma}^{\text{PIB}}$  denotes the NED between  $\mathbf{e}$  and class  $\gamma$ , and  $\theta_{\mathbf{e}}^{\text{PIB}}$  indicates that  $\mathbf{e}$  belongs to class  $\theta_{\mathbf{e}}^{\text{PIB}}$ .

### 4.2 Active Recognition

Not only can the test HRRP  $\mathbf{e}$  be used for passive recognition, but also for active recognition. From class  $\gamma$ 's standpoint,  $\mathbf{e}$ 's feature vector  $\mathbf{s}_{\mathbf{e},\gamma}^{\text{AIB}}$  and the FT matrixes  $\{\mathbf{A}_{1,\xi,\gamma}^{\text{AIB}} | \mathbf{a}_{1,\xi,\gamma,j}^{\text{AIB}}, j=1,2,\dots,m_{\xi}\}$  can be obtained by (25). Then the recognition result is given by

$$\begin{cases} d_{\mathbf{e},\xi,\gamma}^{\text{AIB}} = \min_{j=1,2,\dots,m_{\xi}} \|\mathbf{s}_{\mathbf{e},\gamma}^{\text{AIB}} - \mathbf{a}_{1,\xi,\gamma,j}^{\text{AIB}}\| \\ \theta_{1,\mathbf{e},\gamma}^{\text{AIB}} = \arg \min_{i=1,2,\dots,g} d_{\mathbf{e},i,\gamma}^{\text{AIB}} \end{cases} \quad (\xi, \gamma = 1, 2, \dots, g), \quad (30)$$

where  $d_{\mathbf{e},\xi,\gamma}^{\text{AIB}}$  denotes the NED between  $\mathbf{e}$  and class  $\xi$ , and  $\theta_{1,\mathbf{e},\gamma}^{\text{AIB}}$  indicates that  $\mathbf{e}$  belongs to class  $\theta_{1,\mathbf{e},\gamma}^{\text{AIB}}$  from the standpoint of class  $\gamma$  by individual AIB.

In accordance with the front analysis, it is impossible for  $\mathbf{A}_{1,\xi,\gamma}^{\text{AIB}}$  to contain the full appointed DI  $G_{B,\xi,\gamma}$ , besides which some redundancy information exists in  $\mathbf{A}_{1,\xi,\gamma}^{\text{AIB}}$ . But in some sense, the practical DI  $\Theta_{B,\xi,\gamma}^{\text{AIB}}$  in  $\mathbf{A}_{1,\xi,\gamma}^{\text{AIB}}$ , can be considered as the reflection of  $G_{B,\xi,\gamma}$ , and accordingly, the NED  $d_{\mathbf{e},\xi,\gamma}^{\text{AIB}}$  can be considered as the comparison result of  $\Theta_{B,\xi,\gamma}^{\text{AIB}}$  between  $\mathbf{e}$  and class  $\xi$ . According to Section 2.2, since one class' B-DI obtained from her own standpoint is equivalent to the union obtained from the standpoints of all the other individuals, in order to obtain an equitable result, we synthesize the NEDs from the standpoints of all classes by

$$\begin{cases} d_{\mathbf{e},\xi}^{\text{AIB}} = d_{\mathbf{e},\xi,\xi}^{\text{AIB}} + \frac{1}{g-1} \sum_{\gamma \in D(\xi,g)} d_{\mathbf{e},\xi,\gamma}^{\text{AIB}} \\ \theta_{\mathbf{e}}^{\text{AIB}} = \arg \min_{i=1,2,\dots,g} d_{\mathbf{e},i}^{\text{AIB}} \end{cases} \quad (\xi = 1, 2, \dots, g), \quad (31)$$

where  $d_{\mathbf{e},\xi}^{\text{AIB}}$  denotes the NED between  $\mathbf{e}$  and class  $\xi$  from a global standpoint, and  $\theta_{\mathbf{e}}^{\text{AIB}}$  denotes that  $\mathbf{e}$  belongs to class  $\theta_{\mathbf{e}}^{\text{AIB}}$  by global AIB. Obviously,

$$d_{\mathbf{e},\xi,\gamma}^{\text{AIB}} = d_{\mathbf{e},\xi}^{\text{AIB}} \quad \text{s.t. } \gamma = \xi \quad (\xi = 1, 2, \dots, g), \quad (32)$$

so AIB-GDA can obtain more B-DI than PIB-GDA

Table 2: The correct recognition rates (%) of three airplanes by AIB in the measured experiment

| standpoint | AN    | JIANG | YARK  | Average |
|------------|-------|-------|-------|---------|
| AN         | 91.67 | 81.67 | 90.33 | 87.89   |
| JIANG      | 80.33 | 87.67 | 84.33 | 84.11   |
| YARK       | 82.67 | 90.33 | 94.33 | 89.11   |
| Global     | 94.00 | 92.67 | 95.00 | 93.89   |

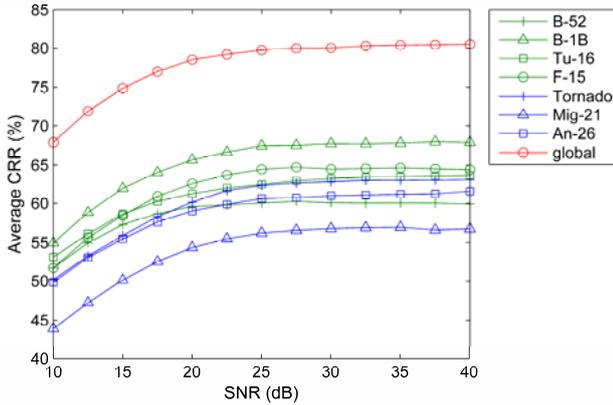


Fig. 3: AIB's average CRRs from different standpoints in the simulated experiment.

does in application, despite that PIB is equivalent to AIB in theory.

As shown in Table 2, we can find that, for a single class from her standpoint, her own Correct Recognition Rate (CRR) is always higher than the CRR of her class estimated by others, while from the standpoints of all classes; the CRRs are much more stable and rational. Also we can analyze the average CRRs corresponding to the Signal Noise Ratio (SNR) and achieve a similar conclusion, as shown in Fig. 3. In summary, global AIB performs better than individual AIB does. At this point in the text, as the intergradation of global AIB, individual AIB has finished its analysis task, so AIB all means global AIB except additional depiction in the next.

### 5 Experiments and Analysis

According to Section 3, there are three discriminant algorithms, i.e., GDA, PIB-GDA and AIB-GDA, which are the specific realizations corresponding to the three DI extraction models, i.e., PGA, PIB and AIB. In this section, several experiments are conducted to evaluate their recognition performances on target quantity, aspect variation, noise disturbance, and so on.

Table 3: Parameters in the measured experiments

| radar parameters | center frequency | 5520 MHz  |            |
|------------------|------------------|-----------|------------|
|                  | bandwidth        | 400 MHz   |            |
|                  | PRF              | 400 Hz    |            |
| planes           | length (m)       | width (m) | height (m) |
| AN               | 23.80            | 29.20     | 8.58       |
| JIANG            | 14.39            | 15.90     | 4.57       |
| YARK             | 36.38            | 34.88     | 9.83       |

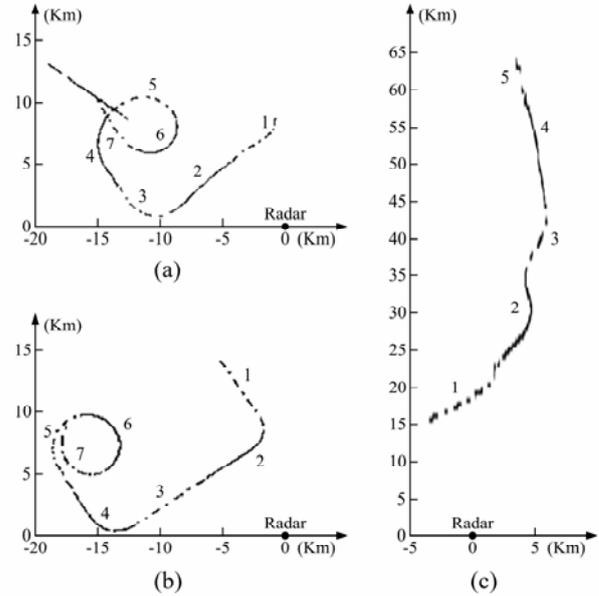


Fig. 4: The projections of target trajectories onto the ground plane. (a) An-26. (b) Jiang. (c) Yark-42.

### 5.1 Description of Experiment Datasets

The original experiment databases includes the HRRPs of the three measured airplanes and the seven simulated airplanes. The details are described as follows.

#### 5.1.1 Measured HRRP Datasets

As described in Table 3, the data used to evaluate the recognition performance were measured by a C-band ISAR radar with bandwidth 400 MHz [4], [13]–[15], [17], [19]–[21], [26], and the projections of target flying trajectories onto the ground plane are shown in Fig. 4, from which the aspect angle of an airplane can be estimated according to its relative position to radar. Since the trajectories of them are different, the data of each airplane are divided into several segments. In this paper, the original measured datasets are selected as per the aspect angle variation of each segment, sampled along the related target trajectory from the 1<sup>st</sup> to 7<sup>th</sup> segments, and

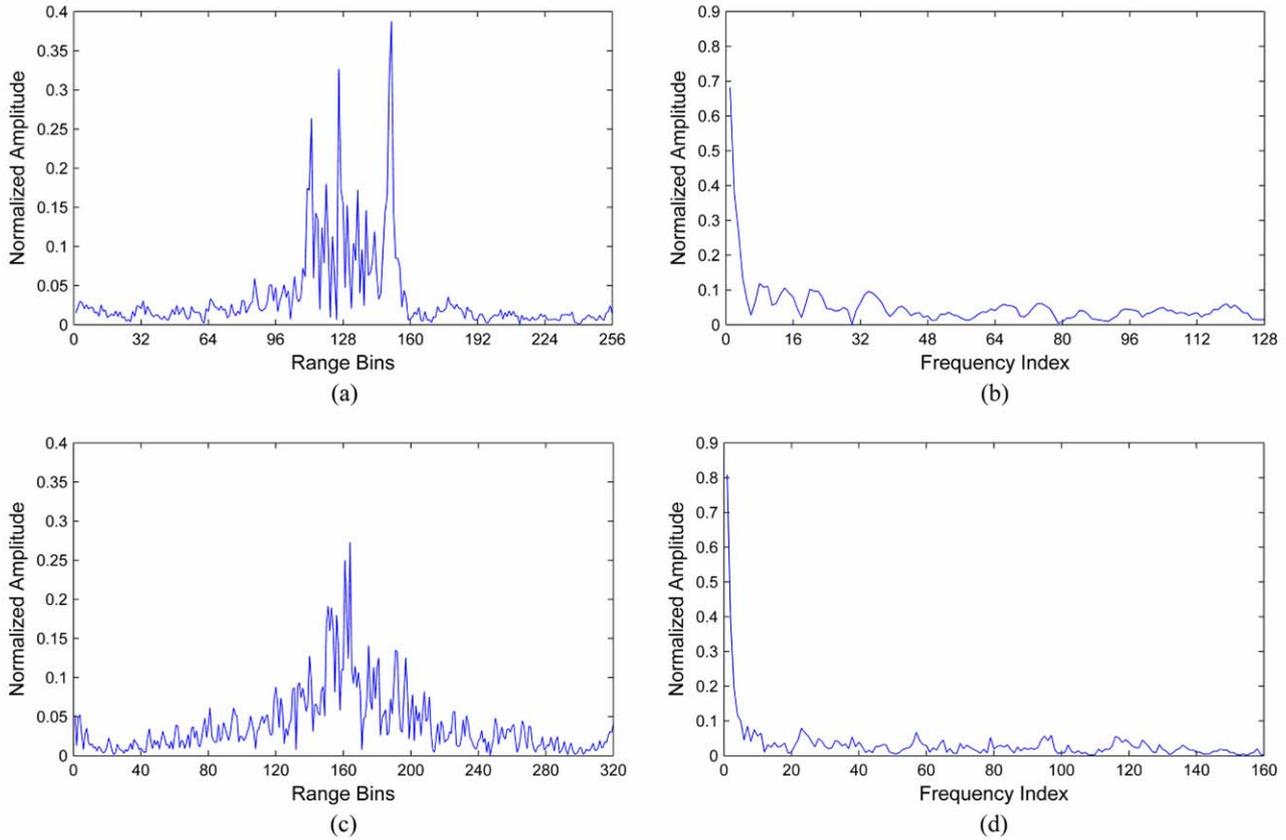


Fig. 5: A measured HRRP and a simulated HRRP for normalized amplitudes in the data preprocessing. (a) The waveform of measured HRRP. (b) The power spectrum of measured HRRP. (c) The waveform of simulated HRRP. (d) The power spectrum of simulated HRRP.

Table 4: Parameters in the simulated experiments

| radar parameters |                    |           |       |
|------------------|--------------------|-----------|-------|
|                  | center frequency   | 5520 MHz  |       |
|                  | bandwidth          | 400 MHz   |       |
|                  | sampling frequency | 800 MHz   |       |
|                  | PRF                | 1000 Hz   |       |
| planes           | length (m)         | width (m) | scale |
| B-52             | 49.50              | 56.40     | 1:1   |
| B-1B             | 44.80              | 23.80     | 1:1   |
| Tu-16            | 33.80              | 33.00     | 1:1   |
| F-15             | 19.43              | 13.05     | 1:1   |
| Tornado          | 16.72              | 13.91     | 1:1   |
| Mig-21           | 15.76              | 7.15      | 1:1   |
| An-26            | 23.80              | 29.21     | 1:1   |

taken from different data segments, including the 1<sup>st</sup>, 2<sup>nd</sup>, 4<sup>th</sup> and 7<sup>th</sup> segments of An-16, the 1<sup>st</sup>, 2<sup>nd</sup>, 4<sup>th</sup> and 7<sup>th</sup> segments of Jiang (Cessna Citation S/II), and the 1<sup>st</sup>, 2<sup>nd</sup>, 4<sup>th</sup> and 5<sup>th</sup> segments of Yark-42. Each target provides 400 HRRP samples as its original dataset, among which the training set is selected at 4-HRRP interval, and the rest 300 HRRPs is used as the test set for the measured experiments except ad-

ditional description.

### 5.1.2 Simulated HRRP Datasets

As mentioned above, the measured HRRPs we have at hand only includes three targets with each target four data segments, so the measured data available in radar HRRP target recognition incompletely covers all of the target-aspect angles, but it exerts little mathematical influence upon the discriminant algorithms presented in this paper. Furthermore, in order to obtain the generalization performance, we simulate radar backscattering data of seven airplanes by a program [36], [37]. The simulated parameters are offered in Table 4. As these seven airplanes are all symmetrical in horizontal, we only simulate azimuth 0°~180° at interval 0.25° and the elevation angle is initialized to 0°. Note that in each simulated experiment, we choice HRRP with azimuth at interval 0.5° as the training subsets, and the related remainders as test ones except additional depiction.

### 5.1.3 Raw Data Preprocessing

Due to the time-shift and amplitude-scale sensitivity of a raw HRRP, its  $l_2$ -norm normalized power spect-

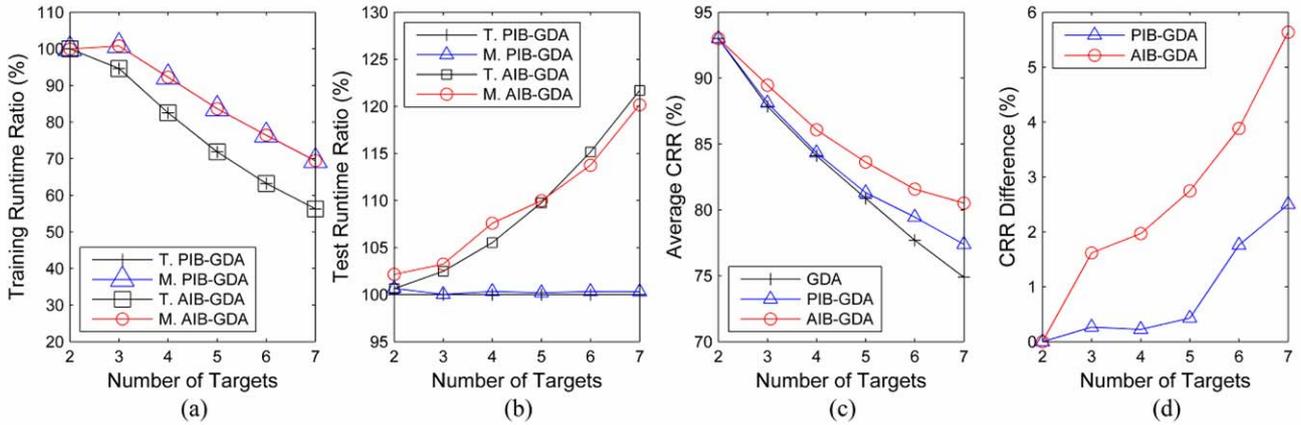


Fig. 6: Computational efficiency and recognition performance versus number of targets in the simulated experiment. (a) Training runtime ratio versus number of targets. (b) Test runtime ratio versus number of targets. (c) Average CRR versus number of targets. (d) CRR difference versus number of targets. (“T. PIB-GDA” denotes “PIB-GDA in theory”, “P. PIB-GDA” denotes “PIB-GDA in practice”, “T. AIB-GDA” denotes “AIB-GDA in theory”, and “P. AIB-GDA” denotes “AIB-GDA in practice”).

rum feature was used to perform classification. As the power spectrum is symmetrical real bilateral spectra, it is enough to use half of the features. As shown in Fig. 5, a raw measured HRRP and a raw simulated HRRP are selected, randomly, to demonstrate the normalized amplitudes of their waveforms and power spectrums respectively.

## 5.2 Experiments on Target Quantity

Since the proposed models and algorithms are designed for multi-target recognition, it is necessary to test their performances according to the change of target quantity.

### 5.2.1 Computational Efficiency

In some sense, the computational efficiencies can be evaluated by the related runtime ratios of PIB-GDA and AIB-GDA compared with GDA. As mentioned in Section 3, GDA is originally designed for multi-class discriminant analysis, but it can be also used for two-class discriminant analysis. As a result, it needs  $C_g^2$  GDA units in PIB-GDA or AIB-GDA, here the total process of  $C_g^2$  GDA units is defined as distributed GDA (D-GDA) in this paper, and  $C_g^\xi$  is the total combination number about selecting  $\xi$  members out of  $g$  members, which is given by

$$C_g^\xi = C(\xi, g) = \frac{\prod_{i=1}^g i}{\left(\prod_{j=1}^{g-\xi} j\right) \left(\prod_{k=1}^{\xi} k\right)}. \quad (33)$$

Let's compare D-GDA with GDA for a  $g$ -class case. The time and space complexities of D-GDA are ab-

out  $c_t$  and  $4/g^2$  times of GDA's respectively [38], here  $c_t$  can be estimated by

$$c_t \approx 2 \cdot \frac{g-1}{g} \cdot \frac{2m+n}{gm+n} \quad \text{s.t.} \begin{cases} m_k = m \\ k = 1, 2, \dots, g \end{cases}, \quad (34)$$

where the definition of each parameter is referred in Section 2.

Now we can estimate the training runtime ratios of PIB-GDA and AIB-GDA by

$$c_t^{\text{AIB},g} \approx c_t^{\text{PIB},g} \approx \frac{44(g-1)}{9g^2 + 4g}, \quad (35)$$

where  $c_t^{\text{PIB},g}$  and  $c_t^{\text{AIB},g}$ , respectively, denote the theoretical training runtime ratios of PIB-GDA and AIB-GDA compared with GDA.

From Fig. 6 (a) we can find that, no matter in theory or the simulated experiment, the differences between AIB-GDA and GDA become more and more obvious with the number of targets increasing, while the training runtime ratios of PIB-GDA and AIB-GDA are very similar. Even though the training runtime isn't of the most important as per the practical demand, the EMS memory is still vital in radar HRRP target recognition due to the huge storage requirement and computation burden which may lead to the program error “out of memory”.

Let's analyze the test processes of PIB-GDA, AIB-GDA and GDA. Specially, their test processes can be divided into two main sub-processes, i.e., feature extraction and template match. Compared with GDA, PIB-GDA and AIB-GDA have the sim-

Table 5: Each target's CRR by the three algorithms in the measured and simulated experiments

| Algorithms | The measured experiment |       |       |       |         |        |       |         | The simulated experiment |       |       |         |
|------------|-------------------------|-------|-------|-------|---------|--------|-------|---------|--------------------------|-------|-------|---------|
|            | Airplanes               |       |       |       |         |        |       | Average | Airplanes                |       |       | Average |
|            | B-52                    | B-1B  | Tu-16 | An-26 | Tornado | Mig-21 | F-15  |         | AN                       | JIANG | YARK  |         |
| GDA        | 91.11                   | 60.00 | 85.83 | 80.83 | 66.11   | 68.06  | 72.22 | 74.88   | 92.67                    | 85.33 | 96.67 | 91.56   |
| PIB-GDA    | 95.00                   | 69.72 | 92.50 | 77.78 | 69.72   | 64.44  | 72.50 | 77.38   | 93.67                    | 90.33 | 94.00 | 92.67   |
| AIB-GDA    | 95.56                   | 70.28 | 91.67 | 86.39 | 72.50   | 72.22  | 75.00 | 80.52   | 94.00                    | 92.67 | 95.00 | 93.89   |

ilar calculation of feature extraction. Although the computation of template matching in AIB-GDA is about  $g$  times of that in PIB-GDA or GDA, it only occupies small part of the total test process. By the same token, their theoretical test runtime ratios can be estimated by

$$r_t^{\text{PIB},g} \approx 1, \quad (36)$$

and

$$r_t^{\text{AIB},g} \approx \frac{n + g(g-1)}{n + g - 1} = \frac{160 + g(g-1)}{160 + g - 1}, \quad (37)$$

where  $r_t^{\text{PIB},g}$  and  $r_t^{\text{AIB},g}$ , respectively, denote the theoretical test runtime ratios of PIB-GDA and AIB-GDA compared with GDA. As shown in Fig. 6 (b), despite that the test runtime ratios of AIB-GDA are acceptably higher than these of GDA, but if parallel and distributed computing is adopted in AIB-GDA, apparently, the test computational efficiency can be improved greatly.

### 5.2.2 Recognition Performance

Fig. 6 (c) shows the average CRRs of the three algorithms according to the number of targets. Compared with GDA, the CRR differences of other two algorithms are shown in Fig. 6 (d). Apparently, the CRR of PIB-GDA is higher than GDA's while lower than AIB-GDA's. Moreover, with the number of targets increasing, the CRR difference between AIB-GDA and PIB-GDA, and the difference between PIB-GDA and GDA become more and more obvious. As aforementioned in Section 2.2, PIB and AIB emphasize B-DI between two random classes while PGA prefers to A-DI among all classes. One view worth emphasizing is that, due to the stability of theoretical B-DI in PIB and AIB, when the number of classes is increasing, the practical B-DI of one class extracted by a certain FEM will become more and more stable, so from the viewpoint of classification ability, PIB-GDA and AIB-GDA have

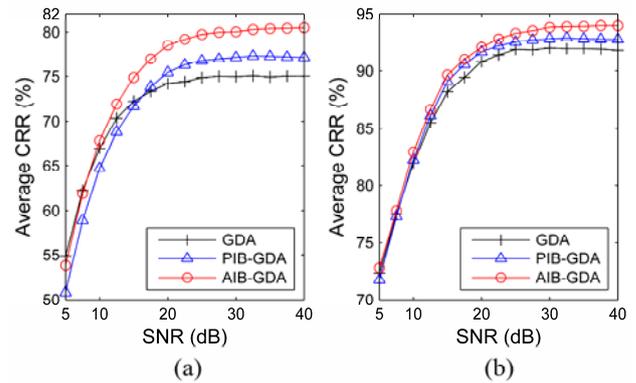


Fig. 7: Recognition performance on noise disturbance. (a) Average CRR versus SNR in the simulated experiment. (b) Average CRR versus SNR in the measured experiment. (Each SNR point repeats 100 times to obtain the average CRRs.)

certain anti-fading character, which can be incompletely corroborated by Fig. 6c. Compared with PIB and AIB, when the number of class is increasing, the theoretical A-DI in PGA reduces sharply, and accordingly, the A-DI extraction becomes more and more difficult, and the practical A-DI becomes weaker and weaker, so the recognition performance of GDA becomes worse urgently. As shown in Table 5, when the airplane number arrives at 7, the CRR difference between AIB-GDA and GDA arrives at 5.65%.

### 5.3 Experiments on Noise Disturbance

In mathematics, HRRP can be considered as a function of target scatters, target distance, radar antenna gain, radar receiver gain, meteorology, etc, so the sources of noise are complex and difficult to analyze. Therefore, it is very important to build a robust and stable model to noise disturbance. In order to obtain a general performance, a Rayleigh noise, with SNR varying from 5 dB to 40 dB at interval 2.5 dB, is added to the original HRRP datasets, including the

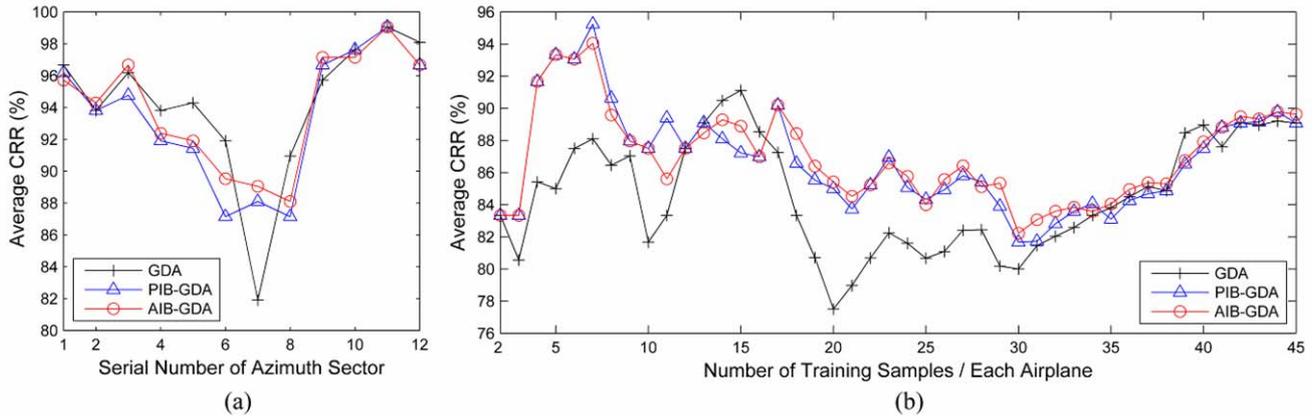


Fig. 8: Recognition performance on aspect variation and training samples.

(a) Three algorithms' CRRs versus the serial number of azimuth sector in the simulated experiment. (In this simulated trial, the serial number varies from 1 to 12, representing the azimuth sector 0–15°, 15–30°, 30–45°, 45–60°, 60–75°, 75–90°, 90–105°, 105–120°, 120–135°, 135–150°, 135–150°, 135–150°, 150–165° and 165–180° respectively. In each sector, each simulated airplane has 60 HRRPs, which are divided into two equal subsets at azimuth interval 0.5°. One is used as the training set and another as the test ones.)

(b) Three algorithms' CRRs versus the number of training samples in the measured experiment. (In this measured trial, each airplane's training samples vary from 2 to 45, selected from its 400-HRRP original dataset in ascending order and at interval 5 HRRPs. For example, when we select  $k$  ( $k = 2, 3, \dots, 45$ ) HRRPs from an airplane's original dataset as the training set, actually, the front  $5k$  HRRPs are selected as the sub-experiment dataset, in which the  $(5(i-1)+3)^{\text{th}}$  ( $i = 1, 2, \dots, k$ ) HRRPs are used as the training set, and the rest as the test ones.)

simulated and measured datasets, to test the noise influence on the three algorithms, as shown in Fig. 8 respectively. One fact worth pointing out is that, in real application, some efficient methods are usually adopted to denoise the original signals, so we usually obtain the HRRPs with high SNR before discriminant analysis. As shown in Fig.7, when the SNR is greater than about 15 dB, the recognition qualities of the three algorithms are labeled from high to low by AIB-GDA, PIB-GDA and GDA.

#### 5.4 Experiments on Training Samples and Aspect Variation

Due to the HRRP's target-aspect sensitivity [4], [7]–[10], aspect variation is one of the main challenges in HRRP-based RATR, which directly changes the distances between target scatters and radar receiver. When the distance change exceeds the Range Resolution Cell (RRC), Range Cell Migration (RCM) appears and the HRRP may change acutely. Even though the aspect fluctuation is within RRC, the HRRP's change still can't be neglected. Moreover, different azimuth sectors always lead to different recognition performances [20]. Perhaps in some sectors the target can keep a relatively stable geometry shape onto the radar LOS, so the HRRP changes

little between two neighbor sampling points. But in other azimuth sectors, the shapes of targets onto the radar LOS may change urgently even within a little aspect fluctuation, so the HRRP may vary sharply, and accordingly, the recognition may become worse.

Some experiments are designed to test the recognition performances of the three algorithms on the variation of aspect or training samples, as shown in Fig. 8. Let's analyze the recognition performance on aspect variation and training samples. Firstly, although the three CRR curves all fluctuate very much due to the high aspect sensitivity of HRRP, but they have different aspect stabilities, which are labeled from high to low by AIB-GDA, PIB-GDA and GDA. Secondly, in terms of CRR, PIB-GDA is superior to GDA but inferior to AIB-GDA in the mass. Thirdly, as many statistical FEMs suffer from the Small Sample Size (SSS) problem when the number of the samples is much smaller than the dimension of the sample space [22]–[27], the three algorithms also suffer from it, and the suffer degrees are sorted from low to high by AIB-GDA, PIB-GDA and GDA.

As mentioned in [26], although GDA has been proved as an excellent FEM, it also suffers from the so-called SSS problem congenitally. When it is used in the DI extraction models as the fundamental FEM

unit, it brings this problem to these models synchronously. From Fig. 8, we can find that, the other two algorithms can greatly depress its bad impression on recognition despite that they can't resolve this problem essentially.

## 6 Conclusions

The three proposed DI extraction models, i.e., PGA, PIB and AIB, mainly focus on the issue of DI selection and exaction in radar HRRP target recognition. Although A-DI is the most usable information for classification, but it has weak anti-fading ability, so PGA is a bit crag-fast in multi-class discrimination under certain global optimization criterion. Due to the relative stability of B-DI, PIB and AIB both more suit for multi-class discrimination than PGA does, and furthermore, PIB is equivalent to AIB in theory but inferior to AIB in application. All the conclusions are tested by the recognition performances of their related actualizations, i.e., GDA, PIB-GDA and AIB-GDA. Experiment results on the measured and simulated HRRP datasets indicate that, as two variations of GDA, the presented AIB-GDA and PIB-GDA both more suit for multi-class discrimination, that is to say further, in terms of computational efficiency, AIB-GDA is acceptably lower than PIB-GDA but higher than GDA; as for recognition performances, generally, PIB-GDA is inferior to AIB-GDA but superior to GDA to many challenges, such as target quantity, aspect variation, SSS problem, noise disturbance, etc. All the analyses and results enlighten us that, in multi-target recognition, even if the quantity of DI is more important than its quantity, the quantity of DI should not be overlooked at least.

### References:

- [1] T. M. Cover, and J. A. Thomas, *Elements of Information Theory (Second Ed)*, Wiley, 2006.
- [2] Z. Nenadic, Information Discriminant Analysis: Feature Extraction with An Information-Theoretic Objective, *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 29, No. 8, 2007, pp. 1394–1407.
- [3] J. Gudnason, J. Cui, and M. Brookes, HRR Automatic Target Recognition from Superresolution Scattering Center Features, *IEEE Trans. Aerosp. Electron. Syst.*, Vol. 45, No. 4, 2009, pp. 1512–1524.
- [4] L. Du, H. Liu, and Z. Bao, Radar HRRP Statistical Recognition: Parametric Model and Model Selection, *IEEE Trans. Signal Process.*, Vol. 56, No. 5, 2008, pp. 1931–1944.
- [5] S. Petridis, and S. J. Perantonis, On the Relation Between Discriminant Analysis and Mutual Information for Supervised Linear Feature Extraction, *Pattern Recogn.*, Vol. 37, No. 5, 2004, pp. 857–874.
- [6] M. Gurban, and J.-P. Thiran, Information Theoretic Feature Extraction for Audio-Visual Speech Recognition, *IEEE Trans. Signal Process.*, Vol. 57, No. 12, 2009, pp. 4765–4776.
- [7] D. R. Weher, *High Resolution Radar (Second Ed.)*, Artech House, Norwood, MA, 1995.
- [8] S. P. Jacobs, *Automatic Target Recognition Using High-Resolution Radar Range Profiles*, Ph. D. Dissertation, Washington Univ., St. Louis, MO, 1999.
- [9] 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, 2000, pp. 364–382.
- [10] M. Xing, Z. Bao, B. Pei, Properties of High-Resolution Range Profiles, *Opt. Eng.*, Vol. 41, No. 2, 2002, pp. 493–504.
- [11] A.V. Oppenheim, and R.W. Schaffer, *Discrete-Time Signal Processing, (Third Ed.)*, Prentice Hall, 2008.
- [12] C. S. Burrus, R. A. Gopinath, and H. Guo, *Introduction to Wavelets and Wavelet Transforms: A Primer*, Prentice Hall, New Jersey, 1997.
- [13] L. Du, H. Liu, Z. Bao, and M. Xing, Radar HRRP Target Recognition Based on Higher Order Spectra, *IEEE Trans. Signal Process.*, Vol. 53, No. 7, 2005, pp. 2359–2368.
- [14] X.-D Zhang, Y. Shi, and Z. Bao, A New Feature Vector Using Selected Bispectra for Signal Classification With Application in Radar Target Recognition, *IEEE Trans. Signal Process.*, Vol. 49, No. 9, 2001, pp. 1875–1885.
- [15] B. Pei, and Z. Bao, Logarithm Bispectrum-Based Approach to High Radar Range Profile for Automatic Target Recognition, *Pattern Recogn.*, Vol. 35, No. 11, 2002, pp. 2643–2651.
- [16] D. E. Nelson, J. A. Starzyk, and D. D. Ensley, Iterated Wavelet Transformation and Signal Discrimination for HRR Radar Target Recognition, *IEEE Trans. Syst. Man Cybern., Part A: Syst. Humans.*, Vol. 33, No. 1, 2002, pp. 52–57.
- [17] B. Chen, H. Liu, J. Chai, and Z. Bao, Large Margin Feature Weighting Method via Linear Programming, *IEEE Trans. Knowl. Data Eng.*, Vol. 21, No. 10, 2009, pp. 1475–1488.
- [18] B. Schölkopf, A. Smola, and K. R. Müller, Nonlinear Component Analysis as A Kernel Eigenvalue Problem, *Neural Comput.*, Vol. 10, 1998, pp. 1299–1319.

- [19] B. Chen, H. Liu, and Z. Bao, PCA and Kernel PCA for Radar High Range Resolution Profiles Recognition, *Proc. IEEE Int. Conf. Radar*, 2005, pp. 528–533.
- [20] L. Du, H. Liu, Z. Bao, and J. Zhang, Radar Automatic Target Recognition Using Complex High-Resolution Range Profiles, *IET Radar Sonar Navig.*, Vol. 1, No. 1, 2007, pp. 18–26.
- [21] B. Chen, H. Liu, and Z. Bao, A Kernel Optimization Method Based on the Localized Kernel Fisher Criterion, *Pattern Recogn.*, Vol. 41, No. 3, 2008, pp. 1098–1109.
- [22] L.-F. Chen, H.-Y.M. Liao, M.-T. Ko, J.-C. Lin, and G.-J. Yu, A New LDA-based Face Recognition System Which Can Solve the Small Sample Size Problem, *Pattern Recogn.*, Vol. 33, No. 10, 2000, pp. 1713–1726.
- [23] H. Yu, and J. Yang, A Direct LDA Algorithm for High-Dimensional Data—with Application to Face Recognition, *Pattern Recogn. Lett.*, Vol. 34, 2004, pp. 2067–2070.
- [24] M. Loog, and R. P. W. Duin, Linear Dimensionality Reduction via a Heteroscedastic Extension of LDA: The Chernoff Criterion, *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 26, No. 6, 2004, pp. 732–739.
- [25] G. Baudat, and F. Anouar, Generalized Discriminant Analysis Using a Kernel Approach, *Neural Computation*, Vol. 12, No. 10, 2000, pp. 2385–2404.
- [26] H. Liu, and W. Yang, Modified Generalized Discriminant Analysis for Radar HRRP Recognition, *Proc. Int. Conf. Microwave Millimeter Wave Technol.*, 2007, pp. 1–4.
- [27] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, Face Recognition Using Kernel Direct Discriminant Analysis Algorithms, *IEEE Trans. Neural Networks*, Vol. 14, No. 1, 2003, pp. 117–126.
- [28] S. Theodoridis, and K. Koutroumbas, *Pattern Recognition (4<sup>th</sup> Ed.)*, Academic Press, 2009.
- [29] B. Schölkopf, C. J. C. Burges, and A. J. Smola, *Advance in Kernel Methods: Support Vector Learning*, MIT Press, 1999.
- [30] J. Mercer, Functions of Positive and Negative Type and Their Connection with the Theory of Integral Equations, *Philos. Trans. Roy. Soc. London A*, Vol. 209, 1909, pp. 415–446.
- [31] G. Baudat, and F. Anouar, Feature Vector Selection and Projection Using Kernels, *Neurocomputing*, Vol. 55, No. 1–2, 2003, pp. 21–38.
- [32] H. Xiong, M. N. S. Swamy, and M. O. Ahmad, Optimizing the Kernel in the Empirical Feature Space, *IEEE Trans. Neural Networks*, Vol. 16, No. 2, 2005, pp. 460–474.
- [33] S. Mika, G. R. Vatsch, J. Weston, B. Schölkopf, and K. R. Müller, Fisher Discriminant Analysis with Kernels, in: *Y.-H. Hu, J. Larsen, E. Wilson, S. Douglas (Eds.), Neural Networks for Signal Processing IX, IEEE, New York*, 1999, pp. 41–48.
- [34] J. Fu, K. Liao, D. Zhou, and W. Yang, Modeling Recognizing Behavior of Radar High Resolution Range Profile Using Multi-Agent System, *WSEAS Trans. Inf. Sci. Appl.*, Vol. 7, No. 5, 2010, pp. 629–640.
- [35] J. R. Munkres, *Topology (Second Ed.)*, Prentice Hall, 2000.
- [36] S. A. Gorshkov, S. P. Leschenko, V. M. Orlenko, S. Y. Sedyshev, and Y. D. Shirman, *Radar Target Backscattering Simulation Software and User's Manual*, Artech House, 2002.
- [37] Y. D. Shirman, *Computer Simulation of Aerial Target Radar Scattering Recognition, Detection, and Tracking*, Artech House, 2002.
- [38] M. A. Weiss, *Data Structures and Algorithm Analysis in C++ (Third Ed.)*, Addison Wesley, 2006.