

基于 LDA 算法的一维距离像特征提取*

李晓辉,黎 湘,郭桂蓉

(国防科技大学 ATR 实验室,湖南长沙 410073)

摘要 在线性分辨分析(LDA)基础上,通过遗传算法寻找 Fisher 准则下最优的线性映射中心和相应的最优映射,使不同模式在特征空间内具有最大的可分离性,并将该算法用于雷达目标一维距离像特征提取与识别中。实验表明,和原算法相比,新算法在特征提取性能和目标正确识别率上有较大提高,说明了算法的有效性。

关键词 目标识别;LDA;遗传算法(GA);一维距离像

中图分类号:TN957.52 文献标识码:A

Feature Extraction of HRRP Based on LDA Algorithm

LI Xiao-hui, LI Xiang, GUO Gui-rong

(ATR Lab, National Univ. of Defense Technology, Changsha 410073, China)

Abstract Finding the right features is a key part of a pattern recognition system. This paper presents a linear feature extraction algorithm which research the optimal pattern centers and projection in Fisher's criterion by genetic algorithm(GA). And then the new algorithm is used in the range profile features extraction and recognition. Experiments show that the new algorithms are more effective than LDA in Fisher's function value and classifying results.

Key words target recognition; LDA; genetic algorithm(GA); range profile

一维距离像可以反映目标的几何形状和结构特点,与低分辨雷达回波相比,能够提供更多目标识别所需的信息。因此,利用一维距离像进行目标识别一直是人们研究的重要内容。但是一维距离像敏感于姿态,在不同的姿态下,同一目标的一维距离像变化很大,不利于识别;而且,目标的一维距离像维数较高,给计算和匹配带来不便。因此,采用适当的维数压缩和稳健的特征提取技术显得极为重要。

数据维数压缩和特征提取是模式识别中的关键技术,有许多方法可以实现。但是,在模式识别中,反映各模式主要特性的信息并不一定是区分不同模式的最有用的信息。所以,需要综合考虑待识别的各类模式的特性,从中提取最有利于减小相同模式样本之间距离和增大不同模式样本之间距离的信息,这对模式识别具有极其重要的意义。其中最有名的是 Fisher 的 LDA^[1]。其主要思想是在 Fisher 准则下,利用线性映射将样本空间映射到特征空间。它的主要缺点是对线性不可分的模式分离效果不佳。由于神经网络良好的非线性映射能力,Carlos Santa Cruz 等人利用神经网络提出了 NLDA 算法^[2],在一定程度上解决了这一问题。但不论是 LDA 还是 NLDA,都有一个共同特点:它们都将各模式样本点的均值作为各自模式默认的中心,这限制了模式分离性能的进一步提高。因此,选取最优模式中心是必要的。文献[3]提出了基于最优聚类中心的目标识别方法,在选取最优模式中心和最优变换上进行了尝试,但由于其将寻找最优模式中心和最优变换分为相对独立的两个步骤,并且其模式中心的选取与样本数据无关,因此在 Fisher 准则下不是最优的。本文在上述 LDA 基础之上,利用遗传算法,在 Fisher 准则下,提出了基于最优模式中心的 LDA 算法,并将其应用于雷达目标一维距离像特征提取和识别中,结果表明了算法的有效性。

1 线性分辨分析(LDA)简述

LDA 主要思想是经过一个线性映射将样本数据映射到特征空间,使得同类模式样本之间相距较近,

* 收稿日期:2005-06-20

作者简介:李晓辉(1974—),男,博士生。

不同模式样本之间相距较远^[4]。

设给定 N 个训练样本,其中包含 C 类模式,其样本数量分别为 $N_i(i=1,2,\dots,C)$, C 类模式表示为 $x_i = \{x_{i1}, x_{i2}, \dots, x_{iN_i}\}, x_{ij}(i=1,2,\dots,C; j=1,2,\dots,N_i)$ 是 n 维向量。于是,各类模式的均值向量为

$$m_i = \frac{1}{N_i} \sum_{j=1}^{N_i} x_{ij} \quad (1)$$

总样本均值向量为

$$m = \frac{1}{C} \sum_{i=1}^C m_i \quad (2)$$

分别计算类内散布矩阵 S_W 和类间散布矩阵 S_B :

$$S_W = \sum_{i=1}^C \sum_{j=1}^{N_i} (x_{ij} - m_i)(x_{ij} - m_i)^T \quad (3)$$

$$S_B = \sum_{i=1}^C (m_i - m)(m_i - m)^T \quad (4)$$

对任意一个 n 维向量 a ,计算 Fisher 函数 $f = \frac{a^T S_B a}{a^T S_W a}$, f 值越大,则表明异类间的差异与同类间的差异之比越大,不同模式之间的线性可分性越强。

设 λ 为矩阵 $S_W^{-1} S_B$ 的最大特征值,而 a 是相应的特征向量,则 a 就是使得 Fisher 函数值最大的映射方向。矩阵 $S_W^{-1} S_B$ 最多有 $C-1$ 个非零特征值,取其最大的 $R(R \leq C-1)$ 个特征值所对应的特征向量组成如下矩阵

$$A = [a_1, a_1, \dots, a_{R}] \quad (5)$$

它所对应的映射空间就是使得 Fisher 函数值最大的 R 维空间。

2 基于最优模式中心的 LDA

传统的 LDA 方法有个明显的缺点是,当模式种类较多时,距离全部样本中心越远的模式对 Fisher 函数值影响越大。LDA 为了得到较大的 Fisher 函数值往往会以减小近距离模式样本之间的距离为代价换取远距离模式样本之间的增大,如图 1 所示。而实际上,近距离模式样本对正确分类影响更大。因此,在进行 LDA 计算 Fisher 函数时,需要对类间散布矩阵 S_B 进行距离加权,重新定义类间散布矩阵 S_B :

$$S_B = \sum_{i=1}^C \frac{N_i (m_i - m)(m_i - m)^T}{d_i^2}$$

其中, d_i 表示第 i 个模式中心到全部样本中心的欧氏距离。

上述 LDA 算法中,将各模式的样本均值作为模式中心,这影响了模式分离的效果,尤其当不同模式样本存在混叠时更加明显。从图 2 可以很直观看出来选择更好的中心对分类性能的提高有很大帮助。寻找最优的模式中心在两个方面有利于提高模式分离能力:

- ① 进一步提高不同模式中心之间的距离,增加类间可分性;
 - ② 在模式样本不是正态分布或均匀分布等中心对称的分布时,新的模式可以增加类内聚集程度。
- 本文利用遗传算法^[5]选择最优的模式中心达到增加同类间的聚合性以及异类间的可分性。算法分为以下几个步骤:

(1) 确定优化的目标函数。目标函数 $f = |S_B|/|S_W|$,算法目的是使目标函数达到最大。

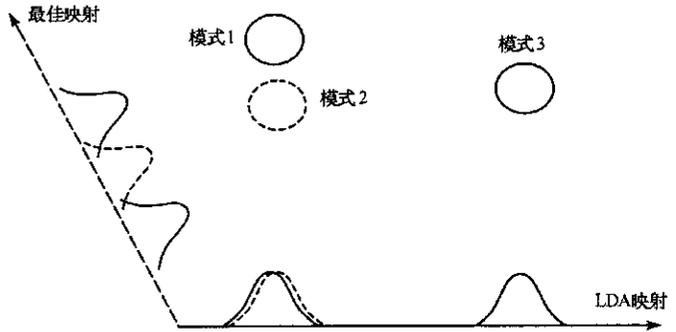


图 1 不同映射模式下 LDA 映射结果示意图

Fig.1 LDA projection in different mode

(2)产生群体。 $t=0$ 时,随机选择 N 组样本,每组含有 C 个样本,作为每类的初始模式中心,即初始群体。其中, N 为人口数目, C 为模式数目。为了加快收敛速度,每类的初始模式中心应在本类样本中进行选择。 $t>0$ 时,群体由进化产生。

(3)编码。对所有初始模式中心进行二进制编码而形成基因码链。

(4)评价。根据不同的编码所对应的模式中心,分别计算其相应的类内和类间散布矩阵 S_W 和 S_B ,利用 LDA 算法计算出各自对应的适应度 $f(i) = |S_B(i)|/|S_W(i)| (i=1, 2, \dots, N)$ 。

(5)选择。按照适应度比例法来选择进行交叉的染色体,各个体的被选择概率为 $P(i) = f(i) / \sum_{j=1}^N f(j)$ 。

同时,为了能在有限的进化过程中防止最优的模式中心被淘汰,这里还采用精华保存法,即适应度最大的染色体不经过交叉直接进入下一代。为保持人口不变,一共产生 N 个新个体。

(6)交叉。本文采用一点交叉的方法对所选择的染色体进行重组。

(7)进化过程达到最大进化世代数则结束算法,否则回到第(2)步。可以看到,为了简化算法,这里没有对各基因进行变异,但对于本文,上述算法已经能够取得很好效果。

3 雷达一维距离像识别中的应用

为了验证改进算法的有效性,本文将其应用于雷达目标一维距离像的特征提取和识别中。实验中所用数据为五种飞机目标缩比模型的外场测量数据。这五种飞机是:F-F117A 隐形战斗机,J-歼轰8,H-幻影2000,W-无人驾驶飞机,Y-运八改预警机。目标缩比模型的尺寸为 $1.5 \sim 2.0\text{m}$,测量系统是毫米波频率步进雷达,工作频率范围为 $34.7 \sim 35.7\text{GHz}$,带宽为 1000MHz ,标称距离分辨率为 0.15m ,步进频率间隔为 2MHz ,极化状态为垂直发射—水平接收。目标的俯仰角为 0° ,方位角范围为 $0^\circ \sim 30^\circ$,方位间隔为 1° 。其一维距离像如图3所示。

可见,同一目标在不同姿态下的一维距离像差别很大,提取不敏感于姿态的特征对于识别非常必要。下面利用 LDA 和基于最优模式中心的 LDA 对其进行特征提取,首先令特征空间为一维,然后利用最小欧氏距离准则进行识别,其中,LDA 算法中将各类在特征空间的中心作为模板,而基于最优模式中心的 LDA 算法将最优模式中心在特征空间中的映射作为各类模板。训练数据由五类飞机奇数姿态角下的一维距离像加噪声构成,非训练数据由五类飞机偶数姿态角下的一维距离像加噪声构成,信噪比均为 25dB 。10 次实验平均结果如图4和表1所示,最优模式中心 LDA 相对于加权 LDA,训练数据和非训练数据的平均正确识别率分别提高了 9.1% 和 8.8% 。因此,新算法在特征提取和分类性能上优于原有算法。

表1 三种算法识别结果

Tab.1 Target recognition results of 3 algorithms

a. 训练数据

	F	J	H	W	Y	平均正确识别率
LDA	0.533	0.933	1	0.6	0.067	0.627
加权 LDA	0.467	0.933	1	0.133	0.6	0.627
最优模式中心 LDA	0.41	0.933	1	0.633	0.612	0.718

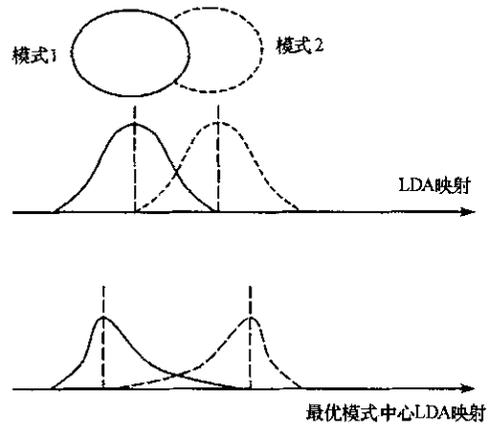


图2 不同模式中心下 LDA 映射结果示意图

Fig.2 LDA projection in different mode centers

b. 非训练数据

	<i>F</i>	<i>J</i>	<i>H</i>	<i>W</i>	<i>Y</i>	平均正确识别率
LDA	0.6	0.733	0.867	0.333	0.067	0.52
加权 LDA	0.33	0.8	0.867	0.067	0.6	0.533
最优模式中心 LDA	0.46	0.733	0.787	0.587	0.54	0.621

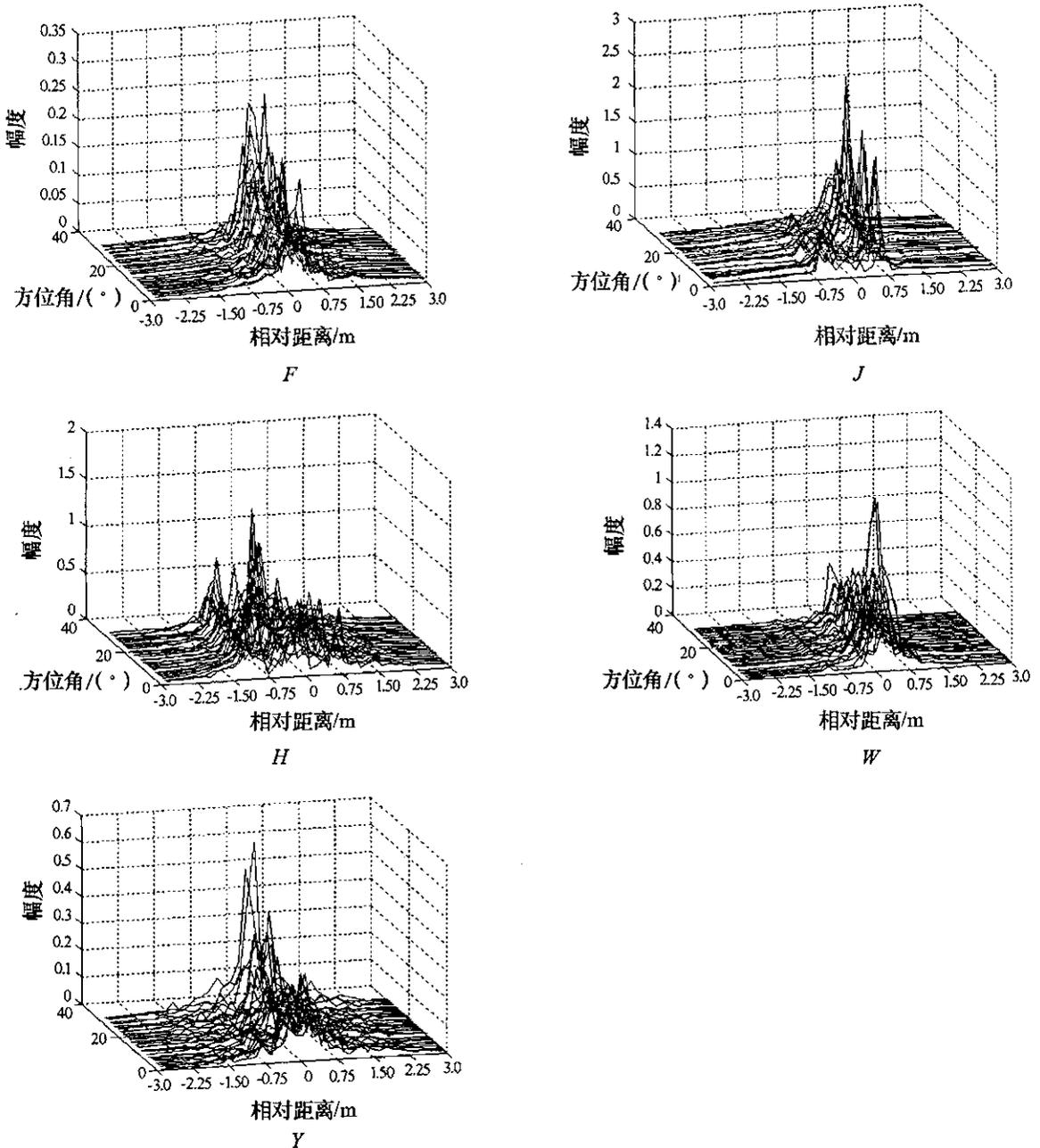


图3 五种飞机在不同姿态角下的一维距离像

Fig.3 Profiles of 5 planes in different direction

本文在不同维数的特征空间也进行了类似的实验,实验平均结果如图5所示。显然,新算法在不同维数的特征空间的特征提取和分类性能上都优于原算法,而且特征空间维数越低,模式之间混叠越严重,效果越明显。

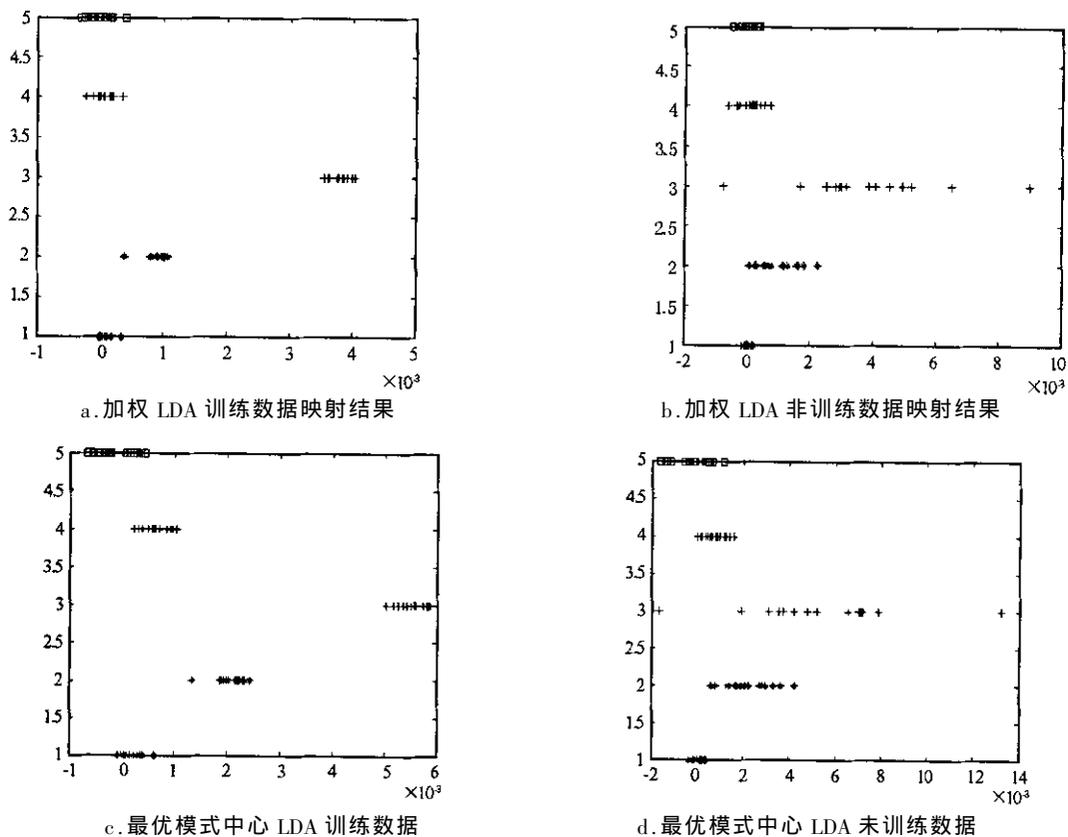


图 4 五类飞机一维特征图

Fig.4 Characteristic of 5 planes in one dimension space

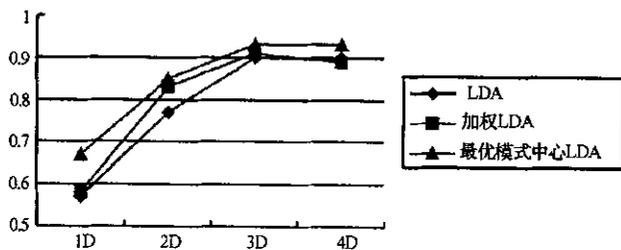


图 5 三种算法在不同特征空间维数下识别结果

Fig.5 Target recognition results of 3 algorithms in different dimension characteristic spaces

4 结论

本文在 LDA 基础上,提出了基于最优模式中心的 LDA 算法,并将其应用于一维距离像特征提取和识别中。实验结果表明,新算法特征提取和分类能力优于原有算法,具有一定的理论和实际意义。需要指出的是,在寻找最优模式中心过程中,遗传算法可能收敛于局部最优解,而不是实际最优模式中心,如何避免该情况和加快搜寻速度是需要进一步研究的问题。

参考文献:

[1] Duda R O ,Hart P E. Pattern Classification and Scene Analysis[M]. New York : Wiley , 1973.

[2] Cruz C S ,Dorrnsoro J R. A Nonlinear Discriminant Algorithm for Feature Extraction and Data Classification[J]. IEEE Transactions on Neural Networks , 1998 (6).

[3] 周代英. 雷达目标一维距离像识别研究[D]. 电子科技大学论文, 2001.

[4] 孙即祥. 现代模式识别[M]. 长沙 :国防科技大学出版社, 2002.

[5] 李士勇. 模糊控制、神经控制和智能控制论[M]. 哈尔滨 :哈尔滨工业大学出版社, 1998.

