

文章编号:1004-9037(2012)01-0000-00

# 基于PCA-LDA-SVM的多普勒雷达车型识别算法

方菲菲<sup>1,2</sup> 余 稳<sup>1,3</sup>

(1. 中国科学院上海微系统与信息技术研究所, 上海, 200050;

2. 中国科学院研究生院, 北京, 100039; 3. 上海慧昌智能交通系统有限公司, 上海, 200233)

**摘要:** 车辆检测和车型识别是智能交通系统(ITS)中的一个重要方面,而目标识别是低分辨率雷达领域的一个难点。该文提出一种用多普勒雷达进行车型识别的方法,把车辆建模成包含多个散射中心的目标体,散射中心与雷达的距离与频谱能量有关,因此同一目标的频谱变化反映了该目标长高等轮廓特征。然后将有效的频谱特征结合主成分分析(PCA)和线性判别分析(LDA)进行降维,再利用支持向量机(SVM)等分类器实现分型。文章对不同识别算法交叉验证的实验结果进行比较,表明基于PCA-LDA-SVM的车型识别算法效果理想,有广泛的应用前景。

**关键词:** 雷达目标识别;多普勒雷达;主成分分析;线性判别分析;支持向量机

中图分类号:TN958.95

文献标识码:A

## Vehicle Recognition algorithm with Doppler Radar Based on PCA-LDA-SVM

Fang Feifei<sup>1,2</sup>, Yu Wen<sup>1,3</sup>

(1. Chinese Academy of Sciences, Shanghai Institute of Microsystem and Information Technology Institute, Shanghai, 200050;

2. the Graduate School of Chinese Academy of Sciences, Beijing, 100039;

3. Shang Hai Huichang Intelligent Transportation System Co. LTD, Shanghai, 200233)

**Abstract:** Vehicle detection and recognition is of great importance to the development of Intelligent Transportation System(ITS), but Target Recognition is a challenging problem for low-resolution Radar. This paper proposes a Vehicle Recognition approach using Doppler Radar, and the spectrum variation of one vehicle reflects its outline. Then, the dimension of effective spectrum feature can be reduced by the methods of Principal Component Analysis(PCA) and Linear Discriminant Analysis(LDA), then vehicles can be classified into three types by classifier algorithms such as Support Vector Machine(SVM), k-Nearest Neighbor(KNN). At last, the paper compares the experiment results of different algorithms by cross validation, and shows the algorithm based on PCA-LDA-SVM can achieve ideal result.

**Key words:** radar target recognition; Doppler radar; principal component analysis; linear discriminant analysis; support vector machine

## 引 言

随着经济的不断发展,人们的生活水平得到极大的提高,但同时也引来一系列问题,交通环境的恶化就是其中一个方面。因此,旨在改善交通环境的智能交通系统(ITS)<sup>[1]</sup>应运而生,其高效的运行

离不开车流量、车速以及车型等交通信息的检测。相比车流量和车速检测的成熟性,车型识别发展稍显滞后,而车型类别无论在交通情况调查还是在高速公路网收费系统多路径确认等方面都很重要,因此近年来成为关注的热点之一。

目前常用于车型识别的方式有感应线圈、视频,红外,超声波,微波等<sup>[2]</sup>。其中微波检测受环境

影响小,可全天候工作,安装方便,虽发展较晚,但可作为车型识别中较有前途的一种方式。用微波雷达进行车型识别<sup>[3]</sup>的常见方式有:多普勒雷达<sup>[2,4]</sup>,高距离分辨率雷达<sup>[5]</sup>,FSR 雷达<sup>[6,7]</sup>或将多普勒雷达和高分辨率雷达相结合<sup>[8]</sup>。高距离分辨率雷达,对带宽要求较高,往往限制了它的使用。FSR 即前向散射雷达,收发分置,能改善系统的灵敏度,但距离分辨率差,工作角度小,因此要求目标与雷达距离很近。

本文采用多普勒雷达进行车型识别,这款雷达以 24.3 GHz 为中心频率,调频带宽 200 MHz,正向安装,采用 4'×6' 的平面阵列天线,最大水平检测距离可达 28 m,发射波瓣角非常窄,有效解决了相邻车道相互干扰和同一车道前后遮挡的问题。交通信息检测算法中,功率谱去噪效果好,目标检测灵敏度高,45°的波束入射角时对应的速度分辨率可达 0.38 km/h,且能较好处理交通拥挤情况。在车辆准确检测的基础上,车型识别算法提取多普勒频谱特征,有效地反映车的轮廓,并可根据用户的需求进行车辆的分型。

特征提取与分类器的设计是车型识别的两个重要环节,前者是分类算法在各领域得到广泛应用的关键,要求特征具有一定的稳定性与不变性。特征提取是在最大程度保持信息量的前提下,降低特征数据的维数,常见的提取方式有主成份分析(PCA),线性判别分析(LDA),独立成分分析(ICA)等。分类器选择也很关键,其设计算法较多,如神经网络,K 最近邻(KNN),支持向量机等,选择准则以适合具体应用为佳。

## 1 多普勒雷达车型识别

多普勒雷达架在 6 m 高处,以 45°的波束入射角射向车道(如图 1(a)),路面上检测区域长为 0.84 m。当目标车辆与雷达有相对运动时,雷达接收的信号与发送的信号在频率上有偏差,为多普勒频移。频移与目标速度的关系为

$$Fd = \frac{2f_0 v \cos(\theta)}{C} \quad (1)$$

式中  $Fd$  为多普勒频移(Hz), $f_0$  为雷达前端发射的 24 GHz 微波信号, $v$  为目标速度(m/s), $C$  为电磁波在空气中的传播速(m/s), $\theta$  为雷达波束入射角。

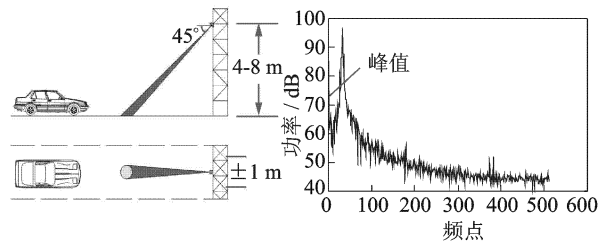
车型识别是以车速和流量的准确检测为前提的,多普勒雷达检测到目标时,信号功率谱显示一个谱峰(图 1(b))。功率谱可辅助车型识别的特性

有:

(1)谱峰的频点对应车辆速度,由车速乘以车经过检测区域的时间,再减去检测区域长就得到车辆长度,该计算过于粗略,中小型车的车长相近难以区分,且零信号的存在导致车辆检测的不连续性,影响车长测量,因此仅车长特征难以较好地区分车型。

(2)最大回波能量,即为功率谱峰值,由雷达方程可知回波能量仅与目标散射截面积的一次成正比,而与距离的四次方成反比,因此回波能量大小可反映车的散射中心与雷达的距离,其间接反映了车的高度。

(3)如把车辆建模成包含多个散射中心的目标,检测过程中谱峰附近频段的能量变化隐含着目标的轮廓特征,这对区分轮廓有差异的中小型车很重要,取有用频段进行分析的方法既排除了速度因素的影响,且无需考虑整个频谱,减小计算量,又能考虑到能量泄漏的存在。因此本文中,这三类特性共同作用,经过有效提取后进行车型分类。



(a) 安装示意图

(b) 回波信号功率谱

图 1 多普勒雷达安装图与回波信号功率谱图

## 2 特征提取与分类器

### 2.1 主成分分析

主成分分析<sup>[9]</sup>的目的是在最小均方意义下,在数据空间中找到一组最能代表原始数据的向量,使数据从原来的  $d$  维空间降维投影到  $M$  维空间( $d > M$ ),并使得降维后保存了数据中主要信息,使其更易于处理。

对  $N$  个  $d$  维的样本  $X = \{x_1, x_2, \dots, x_N\}$ ,用 PCA 算法进行特征提取可按如下步骤进行:

(1)样本  $X$  的散布矩阵为

$$S = \sum_{k=1}^N (x_k - m)(x_k - m)^T \quad (2)$$

式中  $m$  为  $N$  个  $d$  维样本的均值。

(2)散布矩阵的前  $n$  个最大特征值对应的特征向量构成特征空间为

$$A_{\text{eig}} = W^T = [e_1, e_2, \dots, e_n] \quad (3)$$

(3) 将一个测试样本  $x_k$  在特征空间  $A_{\text{eig}}$  中投影, 得到  $n$  个投影值为

$$\text{project\_} X_k = A_{\text{eig}}(x_k - m) \quad (4)$$

以此  $n$  个投影值构成测试样本的特征, 然后利用该特征进行识别。

以上降维时采用的特征向量个数 ( $n$ ) 是个关键的问题, 既要尽可能小, 又要最大限度的保持原始信号的信息。因此引出贡献率的概念, 为  $\sum_{i=1}^{i=n} \lambda_i / \sum_{i=1}^{i=N} \lambda_i$ , 一般来说, 选择  $n$  值使贡献率大于 80% 即可。

## 2.2 基于 Fisher 准则的线性判别分析

与 PCA 一样, LDA 也是特征提取中主流的降维方法, 区别在于 PCA 寻找数据的有效表示方向, 而 LDA<sup>[9]</sup> 寻找数据的有效分类方向, 使得不同类的点在该方向上的投影互相分离, 而同类的点尽量靠近, 即类内离散度最小, 类间离散度最大。

具体数学模型(以两类划分为例)如下。

(a) 类型  $\omega_i$  内样本均值为

$$m_i = \frac{1}{n} \sum_{x \in D_i} x \quad (5)$$

(b) 类型  $\omega_i$  的类内散布为

$$S_i = \sum_{x \in D_i} (x - m_i)(x - m_i)^T \quad (6)$$

(c) 总类内散布矩阵为

$$S_w = S_1 + S_2 \quad (7)$$

(d) 总类间散布矩阵为

$$S_b = (m_1 - m_2)(m_1 - m_2)^T \quad (8)$$

为得到最佳分类方向  $w$ , 可采用基于 Fisher 准则

函数  $J(w) = \frac{w^T S_b w}{w^T S_w w}$ , 使得  $J(w)$  最大必须满足  $S_b w = \lambda S_w w$ , 由于  $S_b w$  总是位于  $m_1 - m_2$  的方向上, 因此当  $S_w$  是非奇异矩阵时, 可取  $w = S_w^{-1}(m_1 - m_2)$ 。这样待测样本在该分类方向上的投影系数为进行识别的特征。

## 2.3 支持向量机

SVM<sup>[10]</sup> 是一种建立在统计学习的 VC 维理论和结构风险最小化基础上的分类方法, 通过引入核函数把低维空间中线性不可分的数据, 表达到高维空间中成为线性可分的数据。SVM 能较好的解决小样本, 非线性和局部极小点的问题。

SVM 用于两类识别, 并通过最大化分类间隔来寻找最优分类超平面, 最终可转化为二次最优规划问题。

## 3 车型识别算法的实现

车型识别的整个流程如框图 2 所示, 对采集的数据进行检测算法后提取有效的特征, 然后使用分类器进行分类, 整个测试过程采用交叉验证的方式。

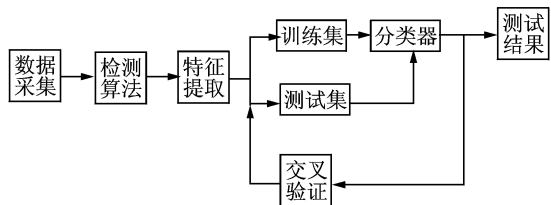


图2 车型检测与识别算法

### 3.1 数据采集及检测算法

雷达采集的数据与视频是配套的, 视频用于记录真实的车道信息, 提供车辆类型。当用目标的频谱特征推测车的轮廓时, 取谱峰及周围  $d$  点的频段进行分析 (512 点 FFT,  $d$  取 11)。雷达检测间隔为 10 ms, 经过检测区域需要  $m$  次检测, 则一个目标的频谱特征  $X_k$  可表示为

$$X_k = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & x_{2d} \\ \cdots & \cdots & \cdots & \cdots \\ x_{m1} & x_{m2} & \cdots & x_{md} \end{bmatrix} \quad (9)$$

因此每个目标对应  $X_k$  和车长 ( $\text{Length}_k$ ), 最大回波能量 ( $\text{Power}_k$ ) 这三组数据,  $X_k$  为  $m \times d$  维, 其中  $x_{pq}$  为第  $p$  次检测第  $q$  个频点能量。  $\text{Length}_k$  和  $\text{Power}_k$  都是一维的数据。

### 3.2 特征提取

由于分类器要求输入样本具有相同的维数, 而  $X_k$  的维数是不定的, 因此需要对  $X_k$  处理实现样本数据的统一。在尽量保持数据统计特征的基础上, 采用两种统一数据的方式。

#### 3.2.1 统一数据

##### (1) 统一数据方法 1 (PCA)

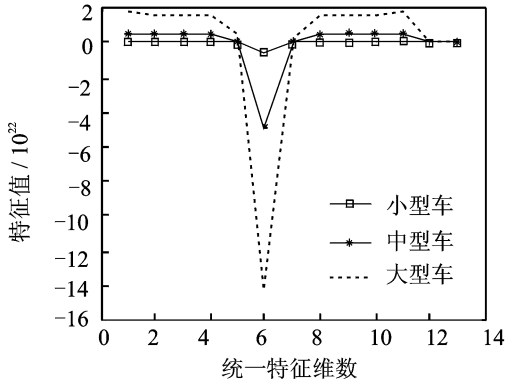
对每个目标的  $X_k$  (式 9) 各自进行主成分分析, 取最大特征值对应的特征向量组成特征空间, 实际处理时贡献率均为 90% 左右, 把每个目标数据在各自特征子空间投影产生的  $d$  维向量称为该目标被提取的特征

$$X p_k = \{x_{k1}', x_{k2}', \cdots, x_{kd}'\} \quad (10)$$

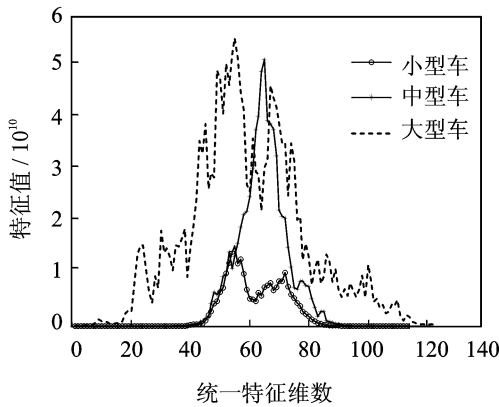
$$X\_PCA = \{X p_1, X p_2, \cdots, X p_N\} \quad (11)$$

$X\_PCA$  为  $N$  个训练样本经 PCA 处理后的特

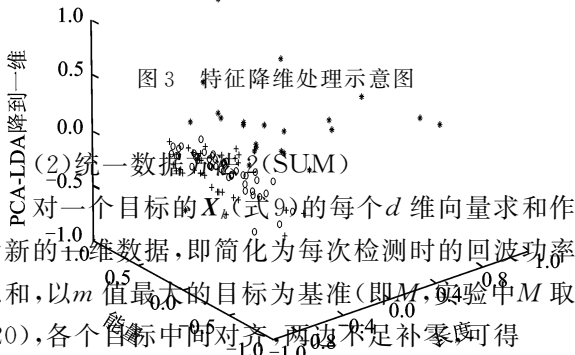
征,求出每一类型车辆的平均特征如图 3(a)所示。



(a)  $X_{PCA}$ : 每一类型车辆的平均特征分布



(b)  $X_{SVM}$ : 每一类型车辆的平均特征分布



(c) 训练样本特征空间三维分布图 (12)  
 (12)  $X_{SUM}$  为  $N$  个训练样本求和统一后的特征, 求出每一类型车辆的平均特征如图 3(b) 所示。

$X_{SUM}$  为  $N$  个训练样本求和统一后的特征, 求出每一类型车辆的平均特征如图 3(b) 所示。

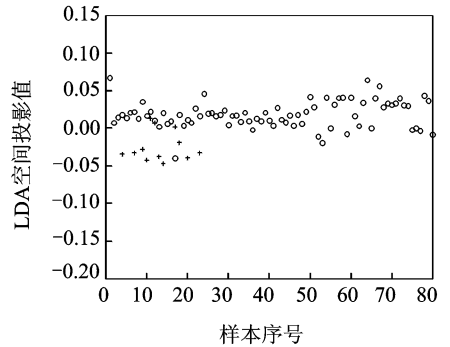
### 3.2.2 特征降维

由于  $Length_k$  和  $Power_k$  均为一维, 而经样本统一后得到新样本的维数都大于  $(d$  维和  $M$  维), 因此需要对其进一步降维, 直到样本的三个特征对分类器的作用均衡。

(1). 对  $N$  个样本  $X_{PCA}$  式 (11), 对其提取 LDA 特征从而降到一维 ( $X_{LDA}$ ), 再结合  $Length_k$  和  $Power_k$  这两个特征, 则每个样本的特征

为  $X'_k = \{X_{LDA_k}, Length_k, Power_k\}$ , 之后输入分类器。全部训练样本特征空间的三维分布如图 3 (c) 所示, 可见大车与中小型车比较容易区分, 而中小型车相差不大, 区分稍困难。这是由于车辆外观和行驶的多样性导致车辆反射面差异很大, 而微波雷达的分辨率相对较低, 使得同一类型的不同车辆的特征变化很大, 当车辆长度较短时这个问题尤为明显。

(2) 对  $N$  个样本  $X_{SUM}$  式 (13), 由于此时每个样本的维数很高 (120 维), 样本数不大时, 无法保证类内散布矩阵的可逆性, 导致直接对其做传统的 LDA 行不通, 只能考虑别的降维方式。如对  $N$  个  $M$  维训练样本做 PCA, 或者直接对每个样本的  $M$  维求和, 即求出目标检测过程中回波功率总和。这两种方法降到一维后, 再结合  $Length_k$  和  $Power_k$  特征构成三维训练样本, 直接输入分类器, 或者对三维样本用 LDA 降到一维, 使得数据最优分类。可以看到大车与中小型车之间区分明显 (图 4(a)), 而与小型车有一部分混叠影响区分 (图 4(b))。



(a) 大与中小型车样本分布示意图

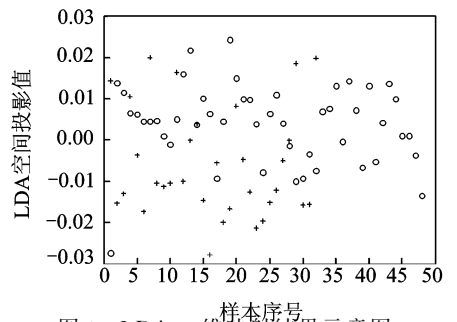


图 4 LDA 一维映射结果示意图  
 (b) 中与小型车样本分布示意图

### 3.3 分类器

分类器的选择有两类: KNN 分类器和 SVM 分类器。为减少样本不均衡的影响, KNN 分类器采用距离加权算法<sup>[11]</sup>, 将较大的权值赋给较近的近邻。具体做法是先将距离标准化, 再用加权核函数  $K(\cdot)$ , 将标准距离转化为待测样本  $x_i$  与周围  $k$  个

近邻  $x_{i(l)}$  同类的概率  $P(x_{i(l)} | x_i)$ , 然后求  $x_i$  为类型  $w_s$  的后验概率, 即  $P(w_s | x_i) = \frac{1}{a} \sum_{l=1}^k P(x_{i(l)} | x_i) I(y(l) = w_s)$ , 此时最大后验概率对应的类别被判定为  $x_i$  的类型。 $k$  值选择使测试正确率尽可能高的值, 测试发现  $k$  为 6 较为合适。

KNN 实现方便, 但是每次测试时的计算量很大, 与训练样本数成正比。而引入 SVM 分类器, 可利用训练好的模型, 较快地实现测试数据的分类。SVM 模型(C-SVM)选用径向基核函数, 模型参数采用遗传算法优化, 多类判别采用 1-V-R 方式(如图 5)。

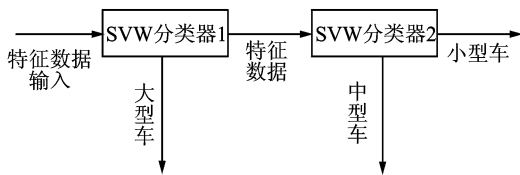


图5 SVM3类车型划分方式

## 4 实验结果与分析

为检测算法的性能, 对数据采用交叉验证的方法: K-CV, 即将原始数据分成  $K$  组( $K$  选为 5), 将每个子集数据分别做一次验证集, 其余的  $K-1$  组子集数据作为训练集, 这样会得到  $K$  个模型, 用此  $K$  个模型的分正确率的平均数作为该分类器的性能指标。K-CV 可以有效地避免过学习以及欠学习状态的发生, 最后得到的结果也比较具有说服力。数据于上海 A20 高速公路上采集, 并从中挑选了 150 组数据, 其中有 75 个小车样本(轿车), 39 个中车样本(面包车), 36 个大车样本(大客车或货车), 最终得到的三类车型划分的测试正确率如表 1 (括号内为对应算法处理后每个样本的维数)。

表 1 各种不同算法的识别结果

组号	样本特征统一	特征降维提取	分类器	测试正确率(%)
1	每个样本		KNN	83.91
2	PCA	LDA(3)	SVM	90.10
3	(11)		KNN	62.00
4	每个样本	LDA(3)	SVM	67.94
5	本求和		KNN	84.20
6	(120)	PCA(3)+Fisher(1) SUM(3)+Fisher(1)	KNN	80.27

由以上实验结果可知对同样的特征 SVM 分类器的性能大体上优于 KNN 分类器, 并且模型训练

出来之后计算量比 KNN 小, 收敛于全局解。

同时看到用统一样本方法 2 中, 对 120 维的初始样本直接做 LDA 的效果很差, 如组 3-4, 原因在于样本维度过高, 样本不够多, 不能保证类内散布矩阵的奇异性, 而用伪逆代替改善效果不明显。因此采取先降维再 LDA 的方法, 当特征维数降到 1 时, 用 SVM 分类器不合适, 故此时用 KNN 分类器。可见用 PCA 降维的方法(组 5)优于直接求和降维(组 6), 因 PCA 算法实现降维, 更能有效地考虑信号的统计特征。

统一样本方法 1 中, 先对每个目标的数据用主成分分析降维, 能有效地消除样本的冗余度, 再使用 LDA 将样本数据映射到最佳分类方向(组 1, 2), 结合 SVM 分类器可达 90.1% 的测试正确率(组 2)。车型识别方面用户需要的正确率一般为 85% 左右, 因此基于 PCA-LDA-SVM 算法的识别正确率已满足实际用户的需求。

## 5 结束语

本文就多普勒雷达进行车型识别的原理做了介绍, 具体阐述了车型识别的算法, 通过一系列比较提出 PCA-LDA-SVM 算法, 较适合于多普勒雷达目标识别系统, 其识别效果已符合用户的一般需求。接下来的研究将注重于更多类型的车型划分, 并进一步提高车型识别的正确率。

### 参考文献:

- [1] Hu Hui, Hu Dawei. Study on the application of ITS and key technologies in logistics operation[C]//Proc of IEEE Conference on Information Management, Innovation Management and Industrial Engineering Xi'an, China: IEEE Computer Society, 2009: 263-266.
- [2] Fang Jianxin, Meng Huadong, Zhang Hao, et al, A Low-cost Vehicle detection and classification system based on unmodulated continuous-wave radar[C]//Proc of 10<sup>th</sup> IEEE Conference on Intelligent Transportation Systems. Bellevue, WA, USA: Institute of Electrical and Electronics Engineers Inc, 2007: 715-720.
- [3] 李明. 雷达目标识别技术研究进展及发展趋势分析[J], 现代雷达, 2010, 32(10): 1-8.
- [4] Jellgren J, Gadd S, Jonsson N, et al, Analysis of Doppler measurements of ground vehicles[C]//Proc of IEEE International Radar conference. Arlington, VA: IEEE Press, 2005: 284-289.
- [5] Xuan Yiguang, Meng Huadong, Wang Xiqin, et al,

- A high-range-resolution microwave radar system for traffic flow rate measurement [C]//Proc of IEEE conference on Intelligent Transportation Systems. Vienna, Austria; Institute of Electrical and Electronics Engineers Inc, 2005; 880-885.
- [6] Isamu M, Youichiro N K O, Akihiro K. Power delay profile matching for vehicle target Recognition [C]//Proc of 70<sup>th</sup> IEEE Vehicular Technology conference Fall. Anchorage, AK, USA; Institute of Electrical and Electronics Engineers Inc, 2009; 1-6.
- [7] Raja Abdullah R S A. , Saripan M I, Cherniakov M. Neural network based for automatic vehicle classification in forward scattering radar[C]//Proc of IET International Conference on Radar System. Edinburgh, United Kingdom;Institution of Engineer And Technology, 2007;1-5.
- [8] Roe H, Hobson G S. Improved discrimination of microwave vehicle profiles[C]//IEEE MTT-S International Microwave Symposium Digest. Albuquerque, NM, USA;IEEE Press, 1992; 717-720.
- [9] Richard O D, Peter E H, David G S. 模式识别[M], 北京:机械工业出版社,2007;94-99.
- [10] Gao Xiangming, Yang Shifeng, Hu Yu. Leakage forecasting for water supply network based on GA-SVM model[C]//Proc of the 2010 Symposium on Piezoelectricity, Acoustic Waves and Device Applications. XiaMen China: IEEE Computer Society, 2010; 206-209.
- [11] 朱明早,罗大庸,易励群. 一种序列的加权KNN 分类方法[J],电子学报, 2009,37(11):2584-2588.

**作者简介:**方菲菲(1987-),女,硕士研究生,研究方向:雷达信号处理,模式识别,E-mail:fff@mail.sim.ac.cn;余稳(1966-),男,副研究员,硕士生导师,研究方向:微波雷达技术。