

参赛密码 _____
(由组委会填写)



第十二届“中关村青联杯”全国研究生 数学建模竞赛

题 目 移动通信无线信道指纹建模及识别

摘 要：

无线信道中多径系数、多径时延、时延扩展、多普勒扩展等参数可作为“指纹”特征对场景进行识别，对改善无线信号传输效率具有重要意义。

本文主要包含五部分内容：无线信道建模、“指纹”特征提取、场景识别、时间序列分段、分段序列聚类。1) 无线信道建模中，从直达径物理模型和双径物理模型出发，建立了多基复包络线性叠加信道模型，体现指纹特征。2) “指纹”特征提取中，提出基于 CLEAN 超分辨谱估计的指纹特征提取算法，通过 Sinc 插值利用 CLEAN 算法逐步提取“指纹”特征。3) 场景识别中，首先使用成像算法获得接收信号距离图像和多普勒域图像，实现了场景预识别；利用超分辨信道路径数目及变化识别出不同场景；采用 Kalman 滤波获得信道主径，并提出利用主径衰落特征区分不同场景的“指纹椭圆”方法，根据接收信号指纹椭圆位置完成场景识别。采用上述方法判断出问题 2 中 Test1 为场景 1 的概率接近 100%，Test2 有 91.9% 的概率为场景 2，有 8.1% 的概率为场景 3。4) 时间序列分段中提出变尺度分段法，以多普勒频率为指纹特征，首先根据等宽度窗口内噪声统计量进行大尺度分段；然后利用变宽度滑窗法增加噪声个数至噪声数阈值，完成强噪声环境小尺度分段；利用时态边缘检测 Sobel 算子获取弱噪声环境下斜率变化较大的边缘点作为分段点，实现弱噪声环境下小尺度分段。该方法将问题 3 中时间序列分为 43 段，拟合误差仅为 0.32，压缩率高达 99.7%。5) 分段序列聚类中，提出基于支持向量机 SVM 的邻段区域融合聚类法，以相邻三个分段中两分段为向量机，获得另一分段分类并与该段融合，以此类推，将问题 3 的分段数据聚为 12 类。采用 SVM 作为评价函数对问题 4 中数据进行判断，并利用 Bayes 假设检验给出问题 4 中两组数据的分类结果及 Bayes 后验概率。其中 Test1 不属于问题 3 中时间序列，Test2 有 87% 概率属于问题 3 中时间序列的 75.4m~77m 之间。

目录

第一章	绪论.....	1
1.1	问题重述及问题分析.....	1
1.2	基本假设.....	2
1.3	文章结构脉络.....	2
第二章	信道建模与指纹特征提取.....	4
2.1	无线信道物理建模.....	4
2.1.1	直达径传输模型.....	4
2.1.2	双径传输模型.....	4
2.2	多基复包络线性叠加信道模型.....	6
2.3	基于成像算法的信道特征预提取.....	6
2.3.1	信道距离像及距离像序列.....	6
2.3.2	距离像序列的走动和校正.....	7
2.3.3	距离多普勒域的联合成像.....	8
2.4	基于 CLEAN 的超分辨谱估计指纹特征提取.....	10
2.4.1	超分辨路径提取.....	10
2.4.2	路径滤波-关联的主径获取.....	11
2.4.3	问题 2 测量信号的场景识别.....	12
2.5	信道指纹特征综合分析及深度挖掘.....	13
2.5.1	指纹特征深度挖掘.....	13
2.5.2	基于“指纹椭圆”的场景识别.....	14
第三章	基于时态边缘算子的时间序列分段方法.....	17
3.1	指纹特征初步分析.....	17
3.1.1	指纹特征种类随时间序号的变化.....	17
3.1.2	基于“指纹特征”的直观场景分析.....	19
3.1.3	指纹特征噪声分析.....	20
3.1.4	变宽度滑窗噪声统计.....	21
3.2	基于时态边缘算子的时间序列分段方法.....	22
3.2.1	时间序列的边缘点及 Sobel 算子.....	22
3.2.2	边缘幅度变化极值点.....	23
3.2.3	时间序列分段线性表示的拟合误差和压缩率.....	23
3.3	基于时态边缘算子的尺度收缩时间序列分段算法.....	24
第四章	基于支持向量机的邻段聚类方法.....	25
4.1	基于支持向量机的场景识别.....	25
4.1.1	基于“指纹特征”的场景匹配.....	25
4.1.2	基于 Bayes 假设检验的概率分析.....	25
4.2	基于 SVM 的邻段聚类区域划分.....	27
4.2.1	原理示意.....	27
4.2.2	问题 4 实测数据的分类判断.....	28
第五章	总结.....	31
参考文献	32
附录: Matlab 程序代码	33

第一章 绪论

1.1 问题重述及问题分析

无线信道“指纹”特征就是无线信道之间的具有差异化的特征，利用这些特征能够有效区分不同信道，从而应用于无线网络优化。无线信道“指纹”特征建模，就是基于先验模型和实测数据提取不同信道的差异性，归纳分析出“指纹”特征的数学模型[1]。实际移动通信中往往涉及信号的多径传输，不同信道传播距离和物体对信号的调制不同，引起多径时延和多径系数存在差异。且对于移动通信而言，接收端的运动以及信道中物体的运动，都会引起多径时延、时延扩展和多径系数的时变性，同时发射端与接收端之间相对速度会引起多普勒频移及多普勒扩展。不同物体对信号的反射性、散射性、衍射性，以及空间直射传播也大大增加了信道的复杂度[2]。为此，移动通信中无线信道建模、“指纹”特征提取、“指纹”特征聚类对于掌握无线信道特征具有重要意义。

因为接收端接收到的信号为多径传输信号的叠加，其功率在信号波长量级上会出现小尺度快衰落，而其平均值会因为环境变化和气象变化在时间上出现大尺度慢衰落。不同的衰落情况能够从一定程度上反应不同的信道特征。为改善信号传输特性，信号要经过发射/接收端滤波器函数进行调制，同时也受到环境背景高斯白噪声的影响。

从噪声背景下，获得信号经过多径传输后的接收信号，利用先验知识和实测数据对无线信道进行建模，获取信道的多径数目、多径时延、时延扩展、多普勒扩展、是否有直达经等参数，建立评价函数对挑选出的特征参数进行聚类以体现不同场景“指纹”特征，有效区分测量数据所处的不同场景和区域。

问题 1 重述：建立具有物理意义的信道指纹模型；对提供的三种场景的信道进行指纹特征提取；提出可以区分三种场景的评价指标；提供程序代码。

问题 1 分析：从信号的叠加性入手，将接收信号表示为多径传输信号之和，为此可以建立以多径传输信道为基的多基复包络线性叠加模型。各个基的参数可以通过多径传输信号的幅度、时延、相位、多普勒频率等参数入手。在含有噪声的信号中提取信号的主要成分，因为在信号叠加过程会引起脉冲变形，为此利用单个脉冲的形状作为递减项，逐步对接收信号的峰值自大而小进行提取。在峰值提取过程中，需要确定峰值提取的停止时刻。考虑到若信号和噪声之间能量抖动的差异，认为当剩余信号能量的方差基本保持不变，则认为剩余信号均为噪声，不再提取峰值。

以提取峰值信号的各个基以及相应的参数进行叠加可以反映原始信号的特征，将这些基的参数作为指纹区分三种场景。例如可以类似于能量大尺度衰减和小尺度抖动将场景区分为平坦-慢衰落、平坦-快衰落、频选-慢衰落、频选-快衰落四种场景，任意选择两个指纹特征判断其均值方差情况，将场景分为四部分。

另一方面，可以类比于雷达成像信息，在二维空间利用两个指纹特征对数据进行成像，使用模式识别相关算法对图像进行区分和识别。

问题 2 重述：根据问题 1 中建立的信道指纹模型，提取所提供的 2 组数据指纹，并根据提出的评价指标判断对应哪一种场景。

问题 2 分析：该问题为问题 1 中提出的场景识别方法的应用。从实测数据中提取问题 1 中进行场景划分需要的指纹特征，判断其属于哪个场景即可。

问题 3 重述：根据展示的步行 150m 的长时间序列，提取指纹特征，采取方法对该时间序列进行分段，根据每段的指纹特征，对分段结果进行分类，并选取评价标准衡量分段和分类的合理性。

问题 3 分析：该问题考察的是测试时间较长的时间序列分段及聚类。首先利用指纹特征的变化情况选择一种较为合适的分段方法，其中比较常用的方法有时间序列相似性聚集 PAA 方法、滑动窗口技术、基于重点点的分段方法等。在选择方法时，可能涉及到不同方法的组合，比如将序列首先进行粗略分段，而后针对粗略分段后每一段进行细分。分段的评价主要涉及分段的有效性和全局性，需要既能够忽略次要特征，又能反映全局特征。因此需要在细节和全局两个维度建立评价指标对分段结果进行评价，最简单的评价方式就是衡量分段后每段对原始数据的拟合效果，以及因为分段引起的数据压缩。

对于分段后的时间序列，提取每一段的指纹特征，建立指纹特征之间相似性的评价指标，对指纹特征相似的分段进行合并。可以从小到大进行逐步分类，每次以三个分段为处理对象，考察能否聚为两类，沿时间轴一个一个分段增加，使分段进行融合和分裂，获得分类情况。

问题 4 重述：根据问题 3 中的分类结果和问题 1 的评价指标，选择某种方法判断问题 4 中的两组数据是否采自问题 3 的时间序列，并给出 Bayes 后验概率。

问题 4 分析：提取实测数据的指纹特征，并利用指纹特征和不同分类之间的相似性评价指标进行分类，考察其属于各个分类的可能性。这一过程可以采取两两分裂的方式，首先将所有分类分成两大类，考察属于大类的可能性，而后在概率较大的大类中将分类再缩小一个层次分为两类，如此判断实测数据属于哪一类。

另一方面，可以直接利用问题 3 中的指纹和匹配原则，对类别进行筛选。

1.2 基本假设

- 1) 假设所有场景为单发射端、单接收端，不存在多发射端和多接收端干扰；
- 2) 假设发射端以球面波向全空间发射无线信号；
- 3) 假设发射信号稳定，即每个发射信号完全相同；
- 4) 假设接收端各个方向具有相同单位增益，能够接收来自全空间的无线信号；
- 5) 假设所有信道为标量信道，不具有方向上的统计特性；

1.3 文章结构脉络

为清晰展示文章的思维逻辑、模型与方法以及与各个问题之间的关系，将文章结构脉络展示为如图 1 所示。

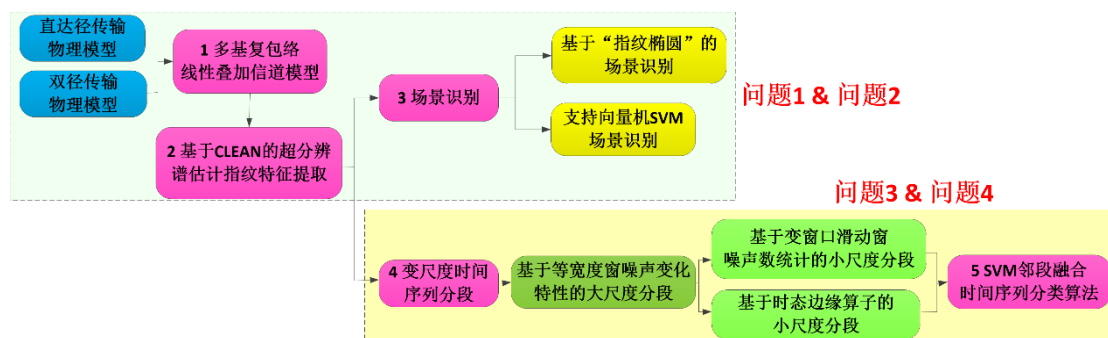


图 1 文章结构脉络

第二章 信道建模与指纹特征提取

2.1 无线信道物理建模

2.1.1 直达径传输模型

直达径是指无线信号从发射端经自由空间直接传输到接收端的传输模型，如图 2 所示，可认为是一种理想、均匀、各向同性的介质空间，当无线信号在自由空间传输时不发生反射、折射、散射和吸收现象，只存在能量在全空间扩散而引起的传输损耗。

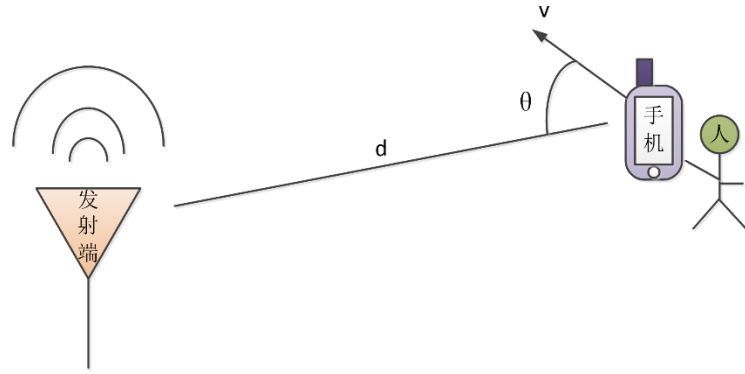


图 2 直达径传输示意

若发射点处以球面波辐射时接收的功率为

$$P_r = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d^2} \quad (1)$$

其中 P_t 为发射功率， G_t 和 G_r 为发射端和接收端天线增益， λ 为波长， d 为发射端与接收端距离。其中天线增益为

$$G_r = \frac{4\pi A_{er}}{\lambda^2}, G_t = \frac{4\pi A_{et}}{\lambda^2} \quad (2)$$

式中 A_{et} 、 A_{er} 分别为发射端和接收端天线有效截面积。

当接收端处于运动时，因为接收端与发射端之间存在相对速度引起多普勒频移

$$f_d = \frac{v \cos \theta}{\lambda} \quad (3)$$

式中 v 为接收端移动速度， θ 为接收端速度与发射接收端路径夹角。

2.1.2 双径传输模型

因为无线通信环境的复杂性，无线通信中很少有单一路径传输的情况，往往会出现多径传输现象[3]。下面以双径传输模型为例，研究多径传输带来的影响。简单起见，考虑自由空间直达径和存在反射体的反射径的双径情况，如图 3 所示。

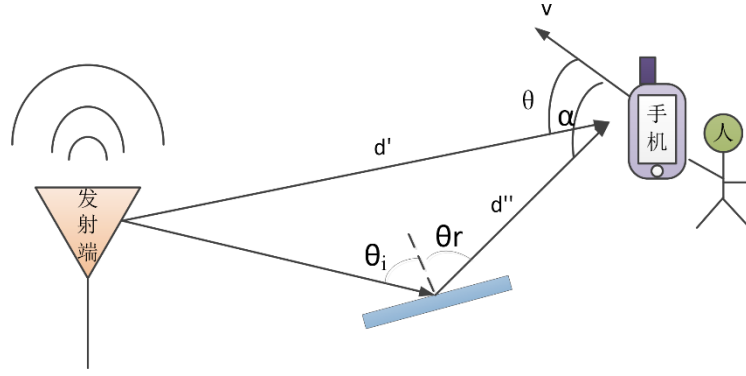


图 3 双径传输示意

由接收信号的叠加特性，接收端场强为双径场强之和

$$E_{tot} = E_d + E_r \quad (4)$$

式中 E_d 为直达径场强， E_r 为反射径场强。

接收端与发射端直达距离为 d' ，经过反射径的传输距离为 d'' ，其中反射径应满足菲涅尔公式，即入射角与反射角相等， $\theta_i = \theta_r$ 。那么接收端接收到的直达径电场为

$$E_d(d', t) = \frac{E_0 d_0}{d'} \cos \left(w_c \left(t - \frac{d'}{c} \right) \right) \quad (5)$$

其中 E_0 为在 d_0 处的场强， w_c 为信号频率。而由反射镜产生的场强为

$$E_r(d'', t) = \gamma \frac{E_0 d_0}{d''} \cos \left(w_c \left(t - \frac{d''}{c} \right) \right) \quad (6)$$

其中 γ 为反射面反射率。假设反射面反射率为-1（负号因为半波损失），那么接收端的总场强为

$$E_{tot}(d, t) = \frac{E_0 d_0}{d'} \cos \left(w_c \left(t - \frac{d'}{c} \right) \right) - \frac{E_0 d_0}{d''} \cos \left(w_c \left(t - \frac{d''}{c} \right) \right) \quad (7)$$

双径之间距离差为

$$\Delta = d'' - d' \quad (8)$$

那么双径场强间的相位差和时延扩展分别为

$$\begin{aligned} \theta_\Delta &= \frac{2\pi\Delta}{\lambda} \\ \tau_d &= \frac{\Delta}{c} \end{aligned} \quad (9)$$

反射径因为接收端运动产生的多普勒频率为

$$f_{dr} = \frac{v \cos \alpha}{\lambda} \quad (10)$$

一般而言，接收端与发射端之间距离 d 很大，因此 d'' 与 d' 之间差异较小，叠加场强可以简化为

$$E_{tot}(d, t) = \frac{E_0 d_0}{d} [\cos \theta_\Delta - 1] \quad (11)$$

可以发现随着测量时间的增大， E_{tot} 会因为直达径和反射径场强之间相位差变化而

发生强弱抖动，表现为接收功率的抖动，可认为是无线电波双径传输的“干涉”效应。

2.2 多基复包络线性叠加信道模型

信号处理的目的在于从信号中提取相应的特征，为分析调制信号的系统提供必要的素材。合理建立信道模型能够体现信道特征，且与实测数据有较好的吻合度[4]。分析可知，无线通信的信道模型应具有如下基本参数：

- (1) 路径个数 $L(m) \in \mathbb{N}$ ，是时变量，简记为 L 。
- (2) 每个路径的幅度调制系数 $a_l(m) \in \mathbb{C}$ ， $l=1,2,\dots,L$ ，是时变量，简记为 a_l 。
- (3) 每个路径的长度变化率 $v_l(m) \in \mathbb{R}$ ， $l=1,2,\dots,L$ ，是时变量，简记为 v_l 。
- (4) 每个路径的时延 $\tau_l(m) \in \mathbb{R}$ ， $l=1,2,\dots,L$ ，是时变量，简记为 τ_l 。

若发射信号的载频为 λ ，则根据条件 (3) 可以求出该信道的多普勒相位调制因子

$$\phi_{\text{doppler}} = \exp(j2\pi v_l / \lambda) \quad (12)$$

结合条件 (2) 中的幅度调制因子可知，每个路径的信道系数为

$$h_l = a_l \phi_{\text{doppler}} = a_l \exp(j2\pi v_l / \lambda) \quad (13)$$

又因为每个路径具有时延 τ_l ，则 L 个路径总的系统函数可以表示为

$$h(t) = \sum_{l=1}^L h_l \cdot \delta(t - \tau_l) \quad (14)$$

若发射单脉冲信号 $\delta(k)$ ，则接收到的信号为

$$\begin{aligned} s_1(t) &= h(t) * g(t) \\ &= \sum_{l=1}^L (h_l \delta(t - \tau_l) * g(t)) \\ &= \sum_{l=1}^L h_l \cdot g(t - \tau_l) \\ &= \sum_{l=1}^L a_l \exp(j2\pi v_l / \lambda) g(t - \tau_l) \end{aligned} \quad (15)$$

上式的物理意义是：信号 $g(t)$ 通过每条路径 $h_l \delta(t - \tau_l)$ 后得到的信号的线性求和。

2.3 基于成像算法的信道特征预提取

2.3.1 信道距离像及距离像序列

距离多普勒 (Range-Doppler, RD) 成像算法是合成孔径雷达成像中的经典算法之一，它能够在时域 (距离) 和频域 (多普勒) 区分出信号的不同分量。类比于雷达成像，我们提出了基于成像算法的信道图像特征提取方法。根据本章第一节分析，接收信号为

$$s_1(t) = \sum_{l=1}^L a_l \exp(j2\pi v_l / \lambda) g(t - \tau_l) \quad (16)$$

由于不同路径信号传播的距离不同，具有不同的时延 τ_l ，因此能够在根据波形上的峰值加以区分，所以我们可以把每个接收信号在快时间上的时域波形称为一个距离像(Range-Profile, RP)；在慢时间上一串连续的距离像可以组合成一个二维的图像，我们称之为距离像序列(RP sequence)。如图 4 所示是场景 3 中 Test2 的单帧距离像和全部慢时间上的距离像序列，从峰值分布可以看出信道具有多个路径，且这些路径的强度和距离在不断变化。

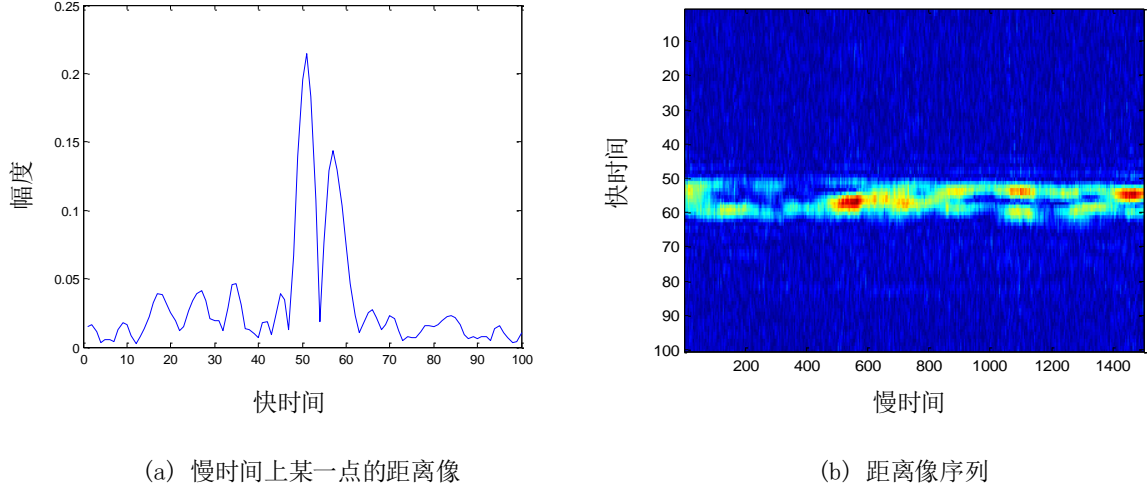


图 4 场景 3 Test2 的距离像（一维）及距离像序列（二维）

2.3.2 距离像序列的走动和校正

由于目标或者环境的运动，路径的距离会随时间变化，会导致距离像序列中包络峰值的走动。如图 5（a）所示是场景 2Test5 的距离像序列（纵向经过放大），由于距离的变化导致的峰值位置的走动。这种走动会增加我们后续分析的难度，需要通过滑动自相关进行校正。如下是距离像走动校正的具体流程：

Step1: 从第二帧距离像开始，求每一帧距离像的实包络 $|RP_m|$ 与上一帧距离像的实包络 $|RP_{m-1}|$ 进行滑动自相关，即求卷积

$$I(k) = \sum_{p=0}^{K-1} |RP_m(p-k)| \cdot |RP_{m-1}(p)| \quad (17)$$

其中 $k=-K+1, -K+2, \dots, K-1, RP_m(p)$ 表示单个距离像 $RP_m(p)$ 的第 $p+1$ 个采样点。

Step2: 找到令 I 取最大值的 k_m ，我们认为在相关系数最大时相邻两帧的距离像拥有最好的相似性，因此将两个距离像对距离像 RP_m 在快时间上平移 k_m 个单位。

Step3: 重复上述 Step，直到平移完所有的距离像

在进行上述对齐的步骤之前，可以对每一帧距离像进行插值，以提高对齐的准确程度。在得到每一个距离像的如图 5(b) 就是对图 5(a) 中的距离像序列进行相关对齐后的结果，可以看出同一个路径上的信号被对齐到了相近的时延位置。可以用公式表达为

$$s_1(t) = \sum_{l=1}^L a_l \exp(j2\pi v_l / \lambda) g(t - \bar{\tau}_l) \quad (18)$$

这里的每个路径的时延 $\bar{\tau}_l$ 与最初的模型表达式(16)中 τ_l 的区别是，它具有更小的时变性。对齐处理使得同一个信号在距离域和多普勒域上正交，其好处在于可以更方便的分析同一个信道中的多普勒频率，而不用担心它的多普勒频谱在距离域上的展宽。

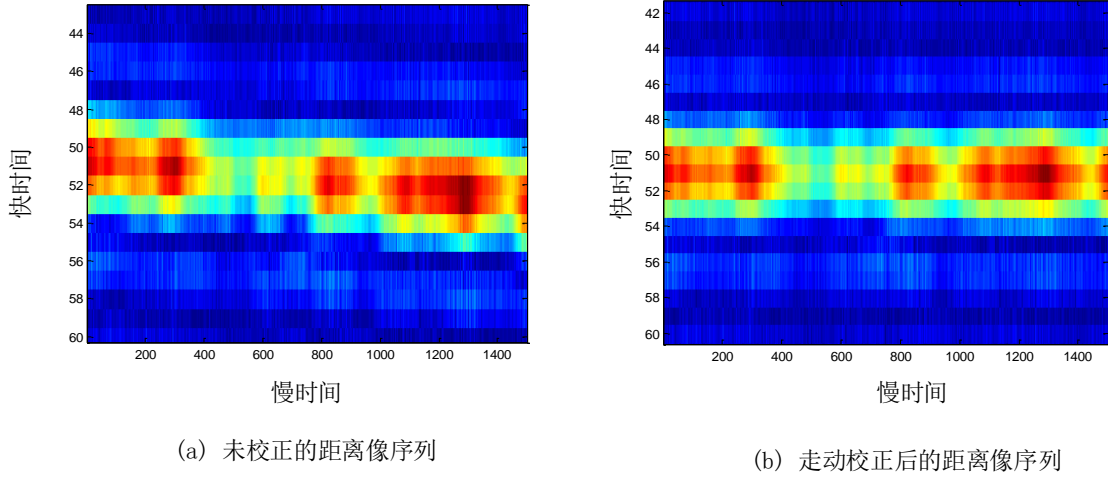


图 5 场景 2Test5 距离像的包络走动校正

利用 Step2 中估计出的 k_m 进行曲线拟合，得到斜率 β ，进而可以对信道的距离变化率进行粗略的估计

$$v = \beta \times c \times 1500 \times 65 \times 10^{-9} \quad (19)$$

下表列出了目标速度的估计结果，

表 1 利用包络走动对场景 1,2,3 的速度估计结果

	Test1	Test2	Test3	Test4	Test5
场景 1	45.77708	63.5511	69.38296	46.23856	61.22637
场景 2	49.20634	41.78398	38.23389	12.01886	32.24009
场景 3	32.3586	22.0479	21.72569	59.76383	63.08838

2.3.3 距离多普勒域的联合成像

经过距离走动校正后，可以近似认为同一路径中信号在整个距离像序列中都处于相近的时延位置，而不会因为目标运动而发生较大改变。因此我们可以在慢时间方向进行离散傅里叶变换(FFT)以得到每个路径的多普勒频移。由于得到的结果是一个二维的图像，其在快时间上体现的是路径的时延，在慢时间频域上体现的是每个路径的多普勒频率，因此，在某种意义上该结果可以看成是对无线信道的一种成像分析，该图像从距离和多普勒频率（距离变化速度）上对每个信道进行区分。相比距离像来说具有更高一个维度的分辨能力。

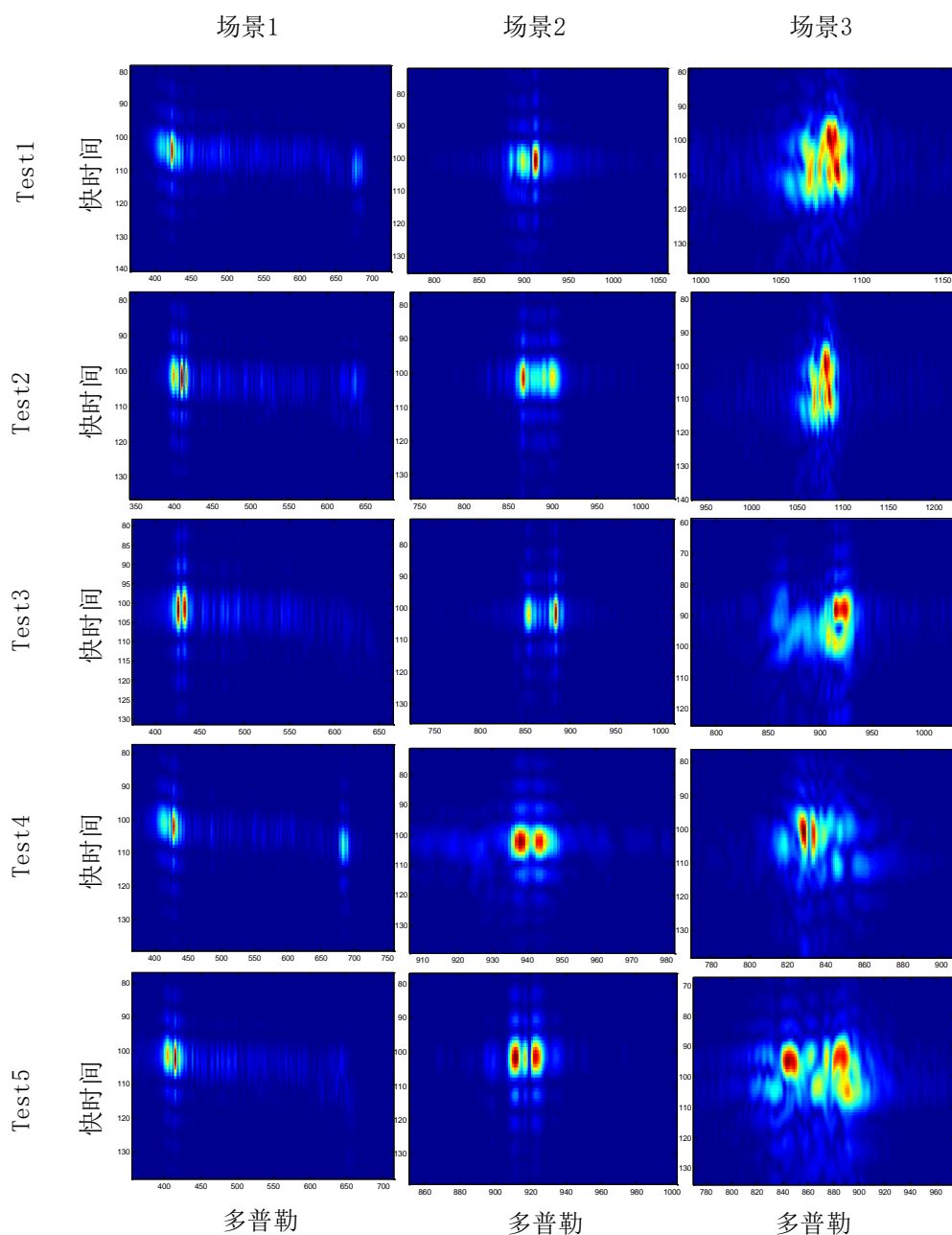


图 6 场景 1, 2, 3 的信道距离多普勒成像结果

图 6 所示是对每个场景的距离多普勒成像，图中的每个峰值都对应了一个真实存在的信号，可以看出场景 1,2,3 之间存在较大的差别。图 7 是对测试数据包中的两组数据进行距离多普勒域的成像，二者对比可以较明显的发现

- (1). Test1ForScrene 与场景 1 较为接近，他们共同的特点是存在较大的多普勒扩展和较小的时延扩展。
 - (2). Test2ForScrene 与场景 2 较为接近，时延扩展和多普勒扩展水平均较小。
- 因此基于成像算法可以较为有效的实现信道指纹特征提取和场景识别。

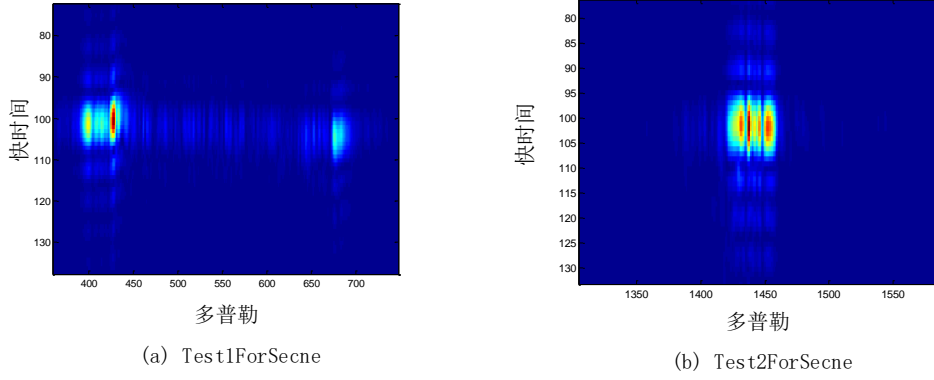


图 7 数据包 2 中数据距离多普勒成像

2.4 基于 CLEAN 的超分辨谱估计指纹特征提取

实际情况下，由于发射信号的带宽有限，很难真正等效为一个理想的冲击函数，因此脉冲的峰值在时域上会存在一定的宽度。我们通常把峰值主瓣的宽度称为时间分辨率，当两个脉冲信号的距离小于分辨率时，二者会互相叠加，导致在时间上难以区分。这里我们利用 CLEAN 算法实现对信道中各路径的超分辨提取[5]。CLEAN 算法是射电天文学家 Hogbom 提出的一种解卷积的处理方法，该过程被广泛的运用于信号处理领域，具有较高的信噪比[6, 7]。

2.4.1 超分辨路径提取

将接收信号的表达式写为如下离散形式

$$s(n) = \sum_{l=1}^L a_l \exp(j2\pi v_l / \lambda) g(n - \tau_l) \quad (20)$$

Step1: 对信号 $s(n), n=0, 1, \dots, N-1$ 进行 *sinc* 插值得到 $s_q(n), n=0, 1, \dots, qN-1$ ，以提高对路径时延的分辨能力，经过尝试插值倍数设为 8 倍较为合适，既能获得较高的分辨率，又能保证计算量不会太大。

Step2: 对插值后的信号 $s_q(n)$ 寻找最大值的位置 $n=n_l$ ，计算 $s_q(n)$ 点的幅度 a_l 和相位 ϕ_l ，从而生成对应的点扩散函数

$$p_l(n) = a_l \exp(j2\pi\phi_l) g_q(n - q\tau_l) \quad (21)$$

其中 $g_q(n - q\tau_l)$ 是等效滤波器 $g(n - q\tau_l)$ 中 q 被 *sinc* 插值后的结果。

Step3: 从信号 $s_q(n)$ 中减去 Step2 中生成的点扩散函数，用剩余信号作为 $s_q(n)$ ，计算剩余信号的方差 σ_l^2 。

Step4: 重复 Step2 和 3，直到方差收敛。

在实际处理中我们发现将相邻两步方差变化率小于 50% 作为该过程的停止条件，可以较好的完成原信号分量的提取，路径数的虚警和漏警率都在许可范围之内。图 8 以场景 1Test1 中的第一帧信号为例，演示了利用 *sinc* 插值和 CLEAN 技术实现超分辨信道路径提取的过程，图中第一行为信号的变化过程，第二行为点扩散函数的变化过程。从第 1 幅图中可以看出，原信号中有两个主要的信号重叠在了一起难以直接区分。在 *sinc* 插值后用 CLEAN 方法对路径分量进行一一提取，能够很好的对信道中每个路径的分量进

行精确而有效的分析。

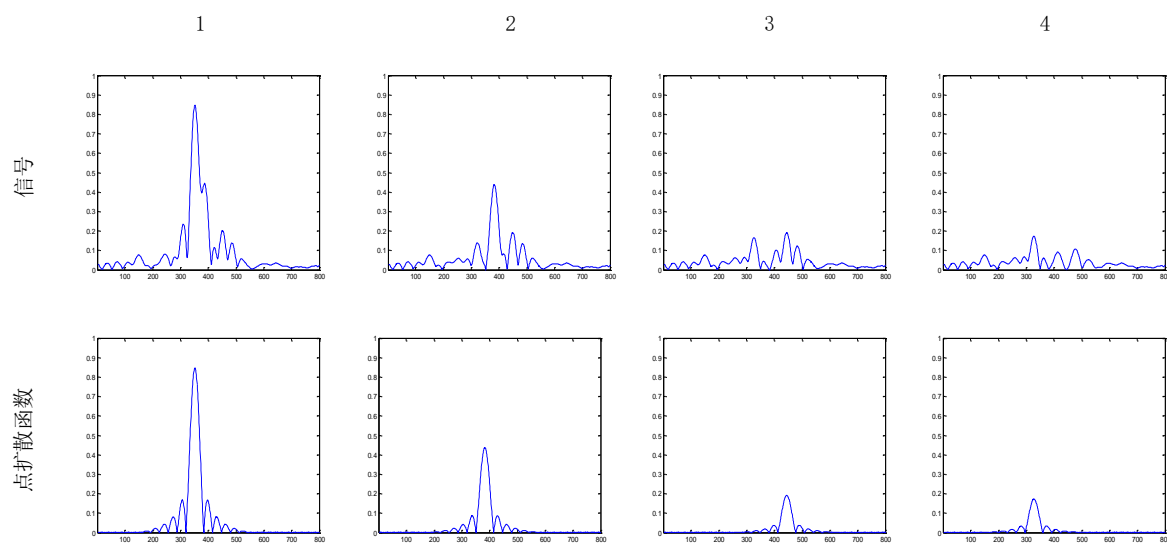


图 8 对场景 1Test1 第一帧数据进行 CLEAN 超分辨的过程

对于每一帧距离像，我们都用该算法进行处理，可以得到信道中各个路径的时延分布，图 9 我们分别列出了场景 1,2,3 中 Test1 的分析结果，可以看出：

场景 1：信道中有两条稳定的路径，这两条路径的距离变化率稳定，且连续性较好。

场景 2：由一条路径构成，距离变化率较稳定，但是有时会出现断裂。

场景 3：由 2~3 条路径构成，距离变化无规律，连续性也较差。

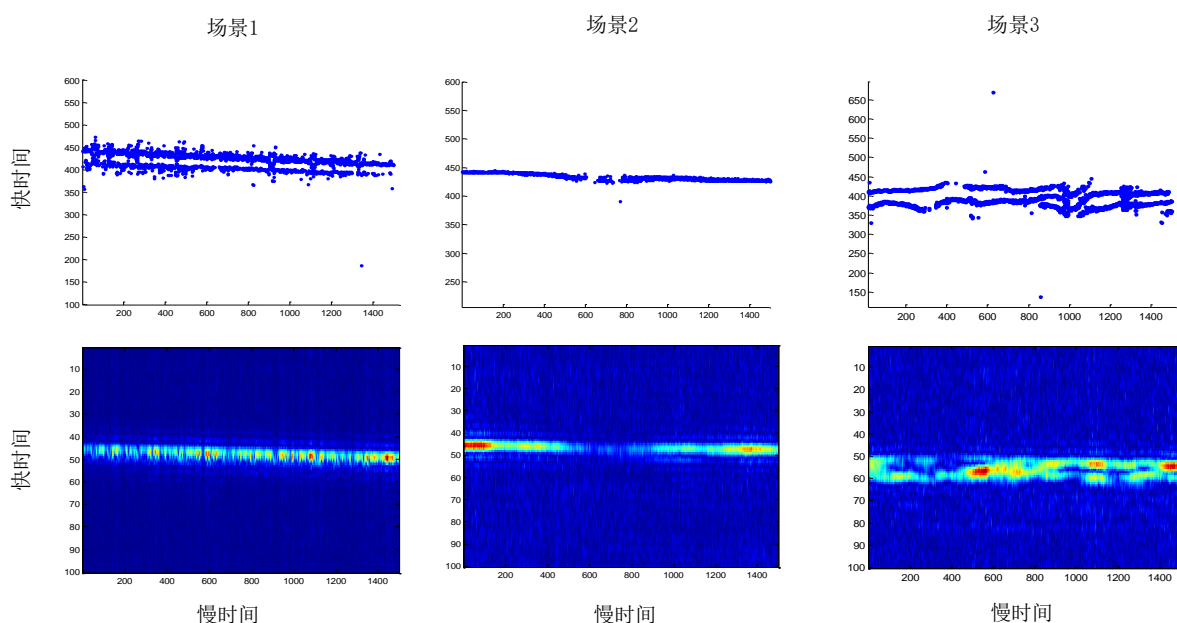


图 9 基于 CLEAN 算法的超分辨信道路径分析（第一行为超分辨算法的分析结果，第二行是对应的距离像序列）

2.4.2 路径滤波-关联的主径获取

利用 CLEAN 技术和超分辨技术[8, 9]我们得到每一帧中信道的不同路径时延，从两组待测样本超分辨路径分析结果散点图中我们也能看出一段时间内的时延量之间是可

以互相关联并形成连续的曲线的。因此，我们可以利用 Klamman 滤波和航迹关联技术对其中的一条或者多条路径进行关联和跟踪[10, 11]，方便后续对每条路径进行独立的分析。

利用超分辨提取出来的时延作为 Kalman 滤波过程中的观测量，而由于同一时刻存在多个时延观测量，所以需要航迹关联技术来选择正确的观测作为每一步滤波的输入。如图 10 所示，是我们利用对场景 1 中的 Test1 的主径进行滤波和航迹关联的结果，配合上节中的超分辨算法，滤波结果很好的反映出了这条路径的时延变化情况。

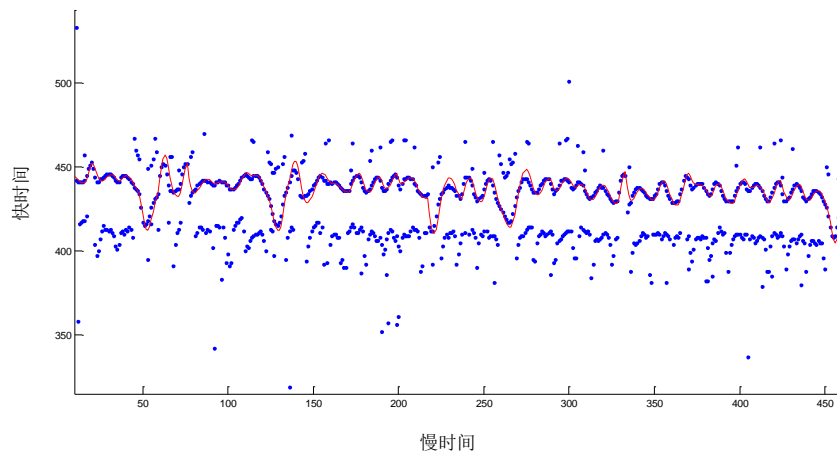


图 10 场景 1 Test1 前 450 帧信号利用 Klamman 滤波和航迹关联进行主径提取

2.4.3 问题 2 测量信号的场景识别

图 11 所示是数据包 2 中两组测试样本的超分辨分析结果，与问题 1 中提供的 3 个场景的超分辨信道路径进行对比分析，可以看出 Test1ForScene 与场景 1 的超分辨信道路径特性较为接近，具有两条稳定的路径；Test2ForScene 与场景 2 的样本较为接近，只有一条速度稳定的路径。

因此，基于 CLEAN 的超分辨谱估计能够有效对信道指纹特征予以提取和场景识别，是一种较为准确的方法。

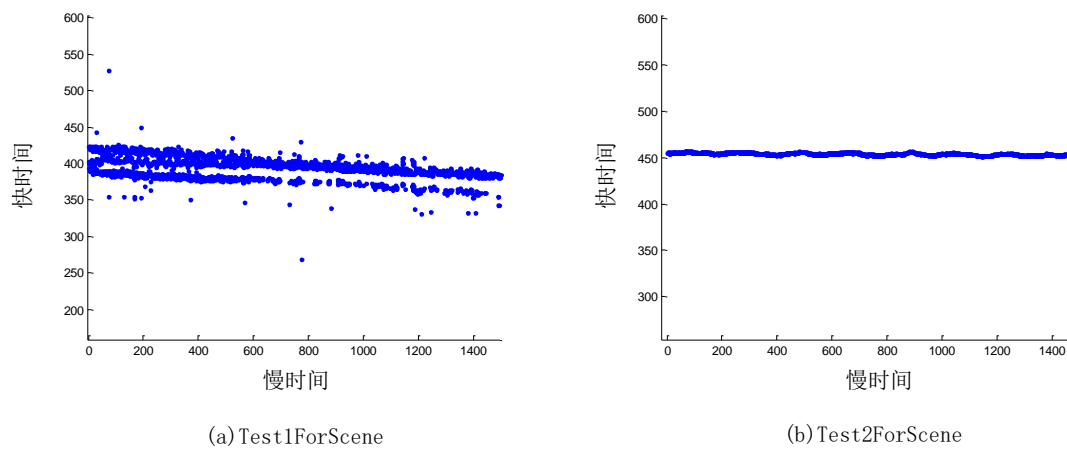


图 11 两组待测试样本的超分辨路径分析结果

2.5 信道指纹特征综合分析及深度挖掘

由于受噪声和快速变化的小尺度衰落的影响,无线信道的特征存在一定的统计意义,可能出现在相邻两帧信号出现急剧的变化,但是从长的时间上来说又符合这个特征的情况。因此,在求部分指纹时,我们需要将相邻的数帧或者至数十帧的信号进行合并,成为一个新的样本,在这个新的样本中,如时延、多普勒、幅度会有更稳定的特性。通过比较我们发现,每个新样本包含 50 帧信号较为合适。因此数据包 1、2 中,每个 Test (包含 1500 帧信号) 都拆分为 30 个样本。

2.5.1 指纹特征深度挖掘

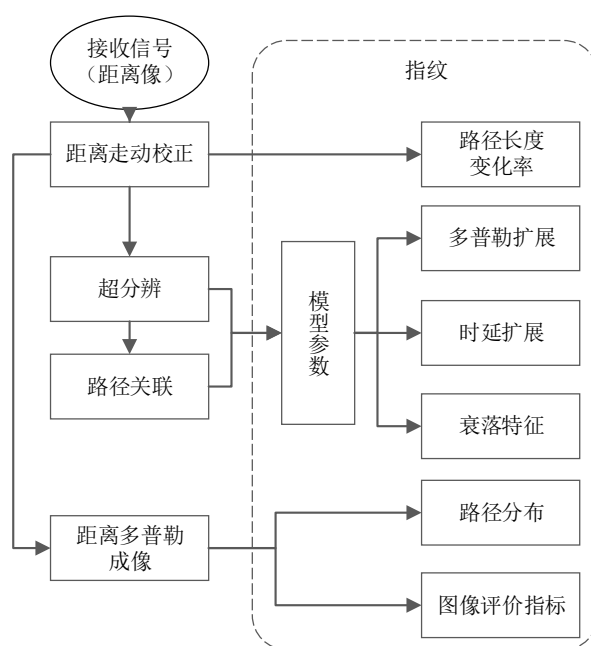


图 12 信号处理及特征提取流程图

图 12 所示是本文中进行参数估计和特征提取的具体流程。其中,距离走动校正是对接收信号进行相关对齐,能够在此基础上分析得到信道的路径变化率。利用超分辨技术我们可以对信道中各条路径的具体参数进行估计,从而提取出信道的多普勒扩展、时延扩展、衰落特征等等。另外,我们借鉴雷达成像中的距离多普勒成像算法对信道在距离多普勒域进行成像分析,从而在距离和多普勒域上得不同路径的分布,另外我们还能够利用成像结果得到对比度和熵等图像评价的指标作为信道的指纹。

1) 时延扩展

时延扩展是场景的基本特征之一,它由信道间的长度差决定。利用 CLEAN 超分辨技术提取出了每个时刻信道的路径时延分布 $\tau_l(n)$,对于每一个样本我们通过如下步骤求其时延扩展量 $\Delta\tau$:

Step1: 对样本中信号的距离像,利用相关对齐的方法进行走动校正,消除时延斜率。

Step2: 对样本中所有的时延量总和,采用 K-means 方法进行聚类。

Step3: 计算出最大的类间距离 $\Delta\tau$,即为该样本的时延扩展量。

该方法相比于直接从时域的两个峰值之间求时延扩展的方法来说,其超分辨能力可

以使时延扩展的计算精度达到原来精度的 1/8，而在原有的分辨率上因为两个峰值常常是叠加在一起甚至提取不出时延扩展量。

2) 多普勒扩展

多普勒扩展的求取方法较为简单：

Step1: 对于一个由 N 帧信号构成的样本，将每帧信号在快时间域上带相位相加得到一个长度为 N 的复时间序列 $[s_t(0), s_t(1), \dots, s_t(N-1)]$ 。

Step2: 对该序列做 FFT 得到频谱 $[s_f(0), s_f(1), \dots, s_f(N-1)]$ ，从 0 到 N 寻找第一个峰值位置 f_0 ，从 N 到 0 寻找第一个峰值 f_{end} 。

Step3: 多普勒扩展为 $\Delta f_d = f_{end} - f_0$ 。

3) 衰落特征

由于信道中可能包含多个路径，每个路径的幅度调制因子（多径系数） a_l 可能具有不同的特性，从而增加对于衰落特征分析的难度，因此，我们只取主径的幅度调制因子进行分析。对于每个样本，其步骤为：

Step1: 用 2.3 节中 Kalman 滤波和航迹关联技术提取每个信道的主径的幅度调制因子 $a_{main}(n), n=0, 1, \dots, N-1$ 。

Step2: 对 $a_{main}(n)$ 进行三次曲线拟合，记录各阶系数 c 拟合残差 r 。

在其中各阶系数反映了该样本中幅度调制因子 $a_{main}(n)$ 的变化趋势，能够在一定程度上反映信道大尺度衰落的特性；拟合残差 r 的意义在于，残差越大说明 $a_{main}(n)$ 在短时间内一次以上的高阶变化越剧烈，也就是小尺度衰落越明显。

2.5.2 基于“指纹椭圆”的场景识别

指纹特征可作为接收信号与场景之间匹配的衡量因素。对于问题 1 中提供的 3 个场景，两两作为一组，获得各自取值的均值和方差，在二维空间可获得相应的“指纹椭圆”。选择可让“指纹椭圆”之间交集较小的两个指纹特征，对于问题 1 中 3 个场景可选择其多径个数 L 、多径系数 a_l 作为“指纹椭圆”的决定量，得到不同场景的指纹椭圆分布情况，“指纹椭圆”方程为

$$\begin{aligned} \frac{(L - \mu_L)^2}{\sigma_L^2} + \frac{(a_l - \mu_{a_l})^2}{\sigma_{a_l}^2} &= 1 \\ L &= \mu_L + \sigma_L \cos(t) \\ a_l &= \mu_{a_l} + \sigma_{a_l} \sin(t) \end{aligned} \quad (22)$$

其中 μ_L 为多径个数均值， σ_L 为多径个数方差， μ_{a_l} 为多径系数均值， σ_{a_l} 为多径系数方差。再对问题 2 中两个测试数据的相应的指纹特征椭圆进行提取，判断其处于哪一种场景的指纹椭圆分布区，可判断测试数据的场景种类。

如图 13 所示，红色虚线区域为场景 1 的“指纹椭圆”分布区域，其中不同红色实线为场景 1 的 5 组测试数据“指纹椭圆”；紫色虚线区域为场景 2 的“指纹椭圆”分布区域，其中不同紫实线为场景 2 的 5 组测试数据“指纹椭圆”；绿色虚线区域为场景 3 的“指纹椭圆分布区域”，其中不同绿色实线为场景 3 的 5 组测试数据“指纹椭圆”。由图可知，根据“指纹椭圆”可将不同场景有效划分为二维特征中不同区域。

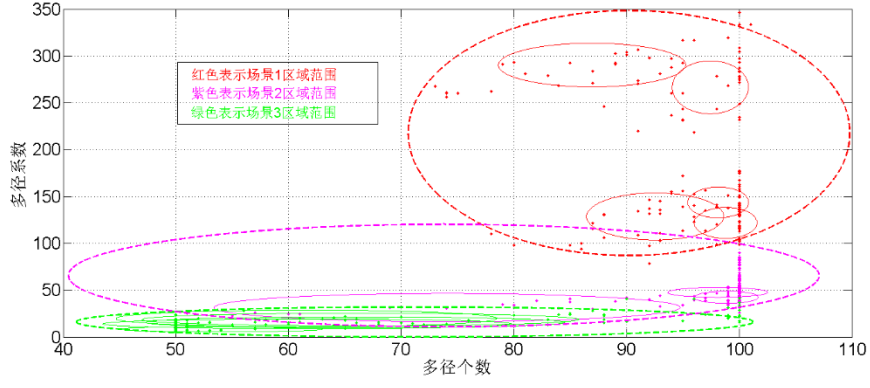


图 13 问题 1 中三个场景的指纹椭圆区域

其中场景 2 和场景 3 “指纹椭圆”分布区域具有一定的交叉，假设两区域的交集面积为 S_x ，场景 2 分布区域面积为 S_1 ，场景 3 分布区域面积为 S_2 ，那么交集区域处于各个场景的概率可通过下式估计

$$\begin{aligned} P_2 &= \frac{S_3 - S_x}{S_2 - S_x + S_3 - S_x} = \frac{S_3 - S_x}{S_2 + S_3 - 2S_x} \\ P_3 &= \frac{S_2 - S_x}{S_2 - S_x + S_3 - S_x} = \frac{S_2 - S_x}{S_2 + S_3 - 2S_x} \end{aligned} \quad (23)$$

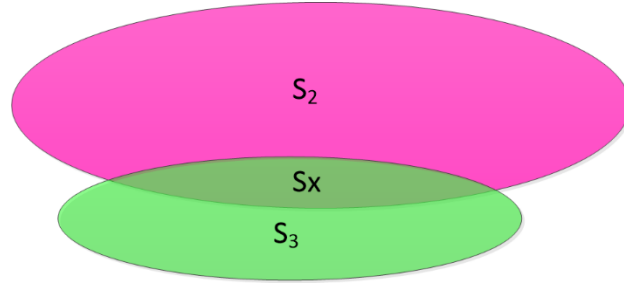


图 14 指纹椭圆区域相交时的置信概率估计

对问题 2 中的多径个数、多径系数为参考量绘制其“指纹椭圆”，并考察其与不同场景的位置关系如图 15 所示。黑色表示问题 2 中 Test1ForScene 测试数据，可以发现其处于场景 1 所处区域中；而对于问题 2 中 Test2ForScene 测试数据，其处于场景 2 所处区域。根据上述不同场景和不同测试数据的“指纹椭圆”，可以有效判断测试数据所属场景环境。

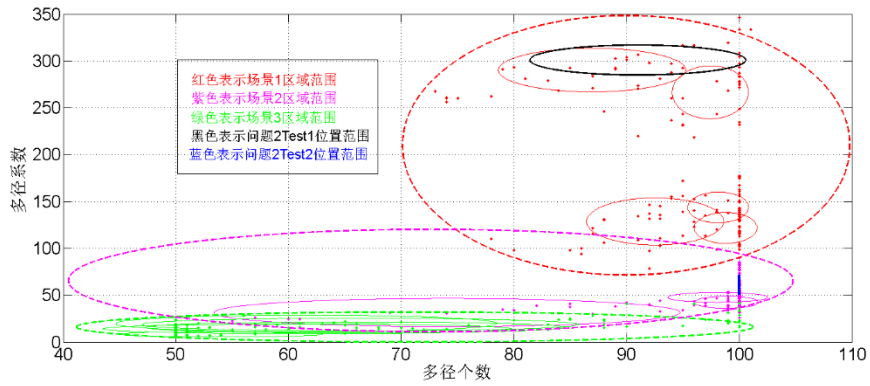


图 15 问题 2 中两组数据的场景识别结果

第三章 基于时态边缘算子的时间序列分段方法

对于经过真实信道接收到的无线信号提取的“指纹”特征，其实质是一组时间序列。对于时间较长的序列，如问题 3 中时间长度为 177s，在不同时间段其“指纹”特征有一定的变化规律，有助于忽略一些细节变化而把握局部特征。对“指纹”特征时间序列分段是对信道环境、场景进行深度挖掘的重要步骤。

时间序列是由记录值和记录时间组成的有序集合

$$X = \langle x_1 = (v_1, t_1), x_2 = (v_2, t_2), \dots, x_n = (v_n, t_n) \rangle \quad (24)$$

其中元素 $x_i = (v_i, t_i)$ 表示时间序列在 t_i 时刻的记录值 v_i ，且 t_i 为由小到大变化。时间序列记录了信号的某个观测量在每个时刻的值，也包含部分受周围环境影响出现的噪声、异常等等。

假设时间序列的分段点集合为 $\langle x_{t1}, x_{t2}, \dots, x_{tm} \rangle$ ，则时间序列的分段线性表示为

$$X = \langle x_1, \dots, x_{t1}, x_{t1+1}, x_{t2}, \dots, x_{tm}, x_n \rangle \quad (25)$$

3.1 指纹特征初步分析

3.1.1 指纹特征种类随时间序号的变化

根据问题 1 中建立的无线衰落模型，可以提取出问题 3 中测量信号的多径个数、多径系数、多径时延、时延扩展、多普勒扩展等参数随时间序号的变化关系，这些物理量作为指纹特征可以表征信号随着移动台步行 150m 的场景的变化过程。图 16 分别展示了第 1~3 条多径系数随时间的变化关系，能够明显看到时间序列首尾部分多径系数存在明显抖动，受到噪声影响严重，其中主径能够看到在噪声较弱区域，系数也有小尺度的缓慢变化。图 17 展示了多径个数随时间的变化关系，多径个数从 1 条至 4 条不等，且随时间发生变化，不同时间一般多径个数为 2 条。图 18 展示了不同路径相位随时间的变化关系，明显的相位变化意味着该路径环境变化明显。图 19 展示了不同路径的时延变化，可以明显看出路径 1 时延随时间序号的变化最为明显，表明该路径上可能因为场景中物体移动或者接收端的移动引起较大的路径时延变化。图 20 展示了多普勒频率随时间的变化。

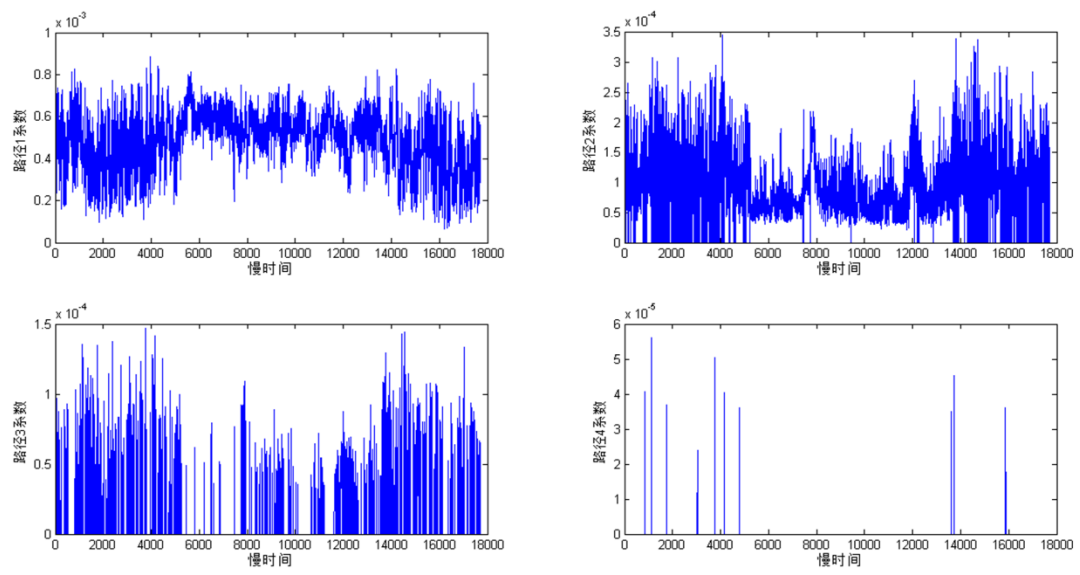


图 16 问题 3 不同路径多径系数随时间变化关系

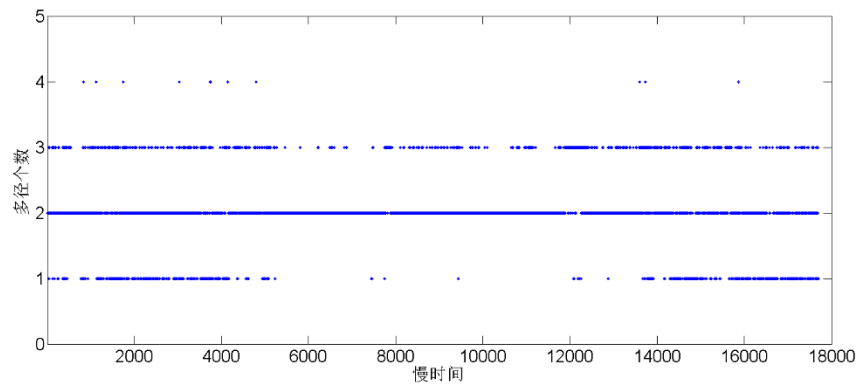


图 17 问题 3 多径个数随时间变化关系

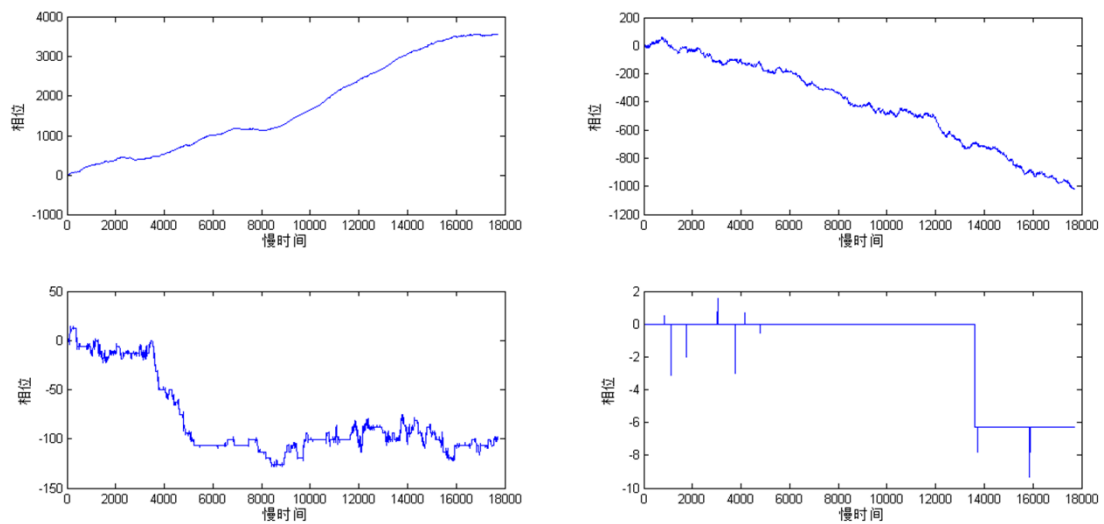


图 18 问题 3 不同路径相位随时间变化关系

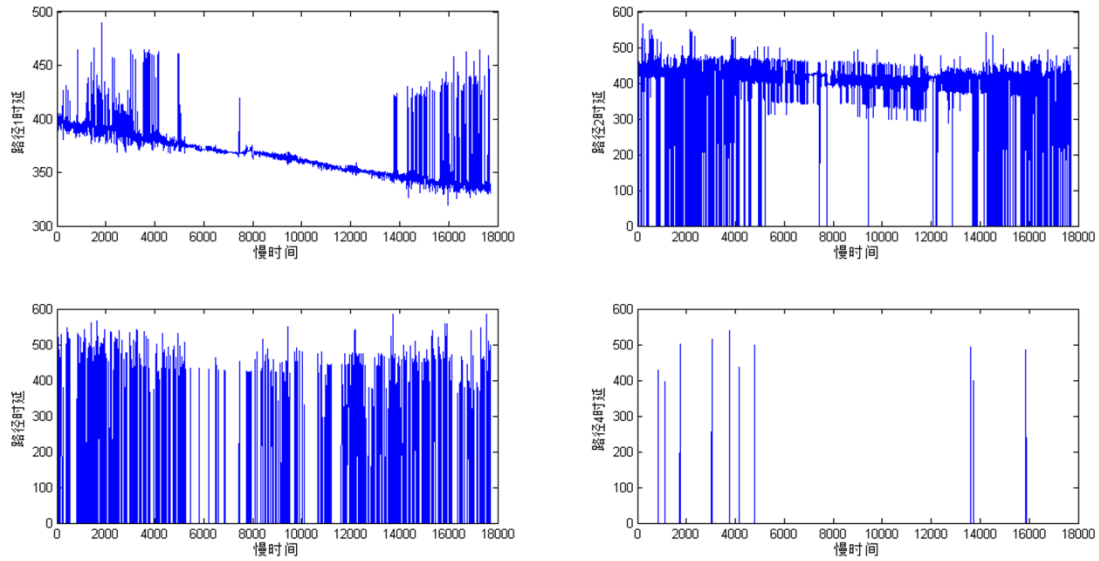


图 19 问题 3 不同路径时延随时间变化关系

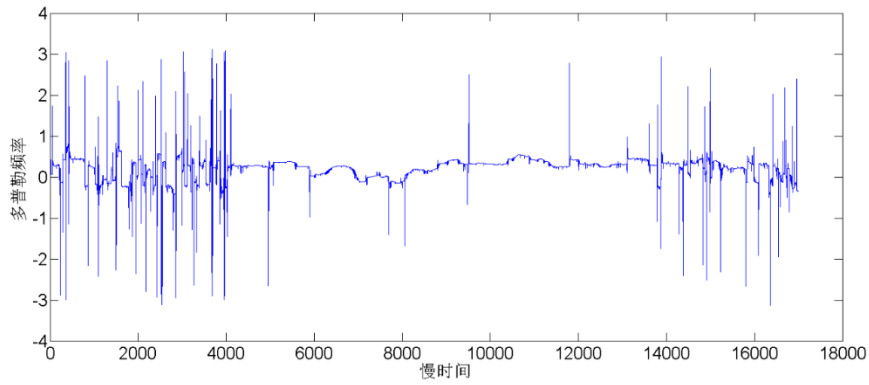


图 20 问题 3 主径多普勒相位随时间变化关系

3.1.2 基于“指纹特征”的直观场景分析

问题 3 中接收信号随时间序号的强度变化如图 21 所示，可以直观发现时间序号在 1~4000，13500~17700 范围内，接收信号在小尺度上强度迅速涨落，而 4000~13500 范围内小尺度涨落比较不明显，且整体强度较强。因此，直观上仅能将接该时间序列分为三部分。

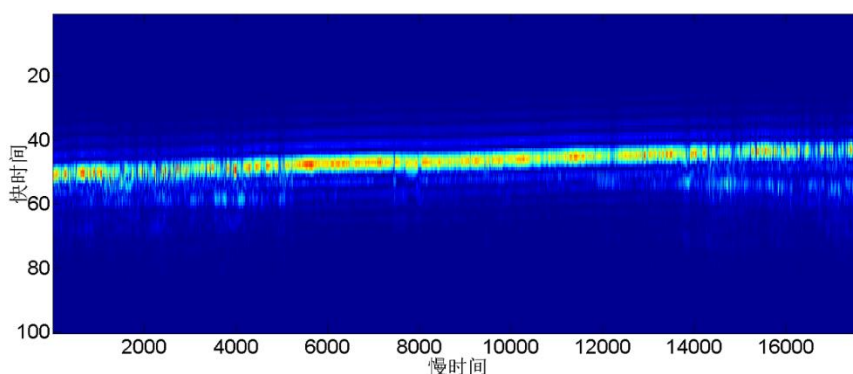


图 21 问题 3 中接收信号强度随时间变化关系

为将时间序列进行更精细的分段,选择大小尺度变化较为合适的主径多普勒频率变化作为特征点对该段测量数据进行分段。直观上可以验证在 0~4000, 13500~17700 范围内噪声较多,引起多普勒频率瞬时跳变明显,而 4000~13500 范围内噪声影响很小,频率变化较为和缓。

3.1.3 指纹特征噪声分析

类比于时间序列分段聚集近似方法 (Precise aggregate approximation, PAA), 将时间序列进行等间隔分段,可获得每个窗口内方差随时间序号的变化关系,取窗口宽度为 400 获得如图 22 所示方差随时间序号的变化关系。可以发现以 4000、13500 两个时间为分割点,可将时间序列分为多普勒频率变化较大、多普勒频率变化较小、多普勒频率变化较大三部分。对于多普勒频率较大的部分,是由于噪声较多而引起,因此其为强噪声场景,我们采用对噪声的统计特性对其进行分段。对于低噪声场景,场景之间的不同更容易凸现出来,所以对其进行尺度收缩的进一步时间序列分割。

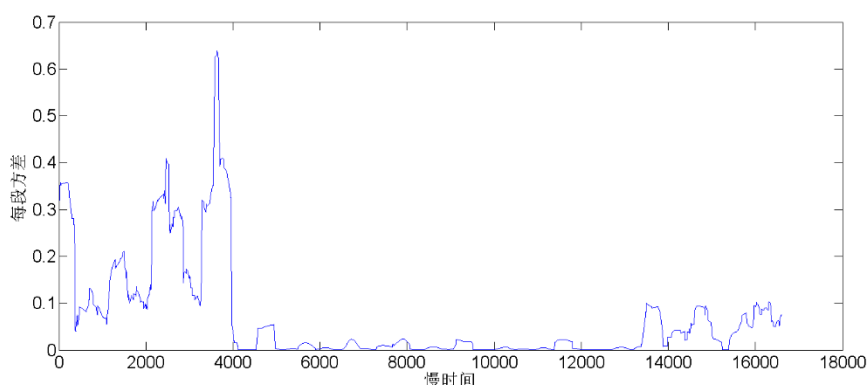


图 22 问题 3 等宽度窗口内多普勒频率方差随时间变化

与上述等窗口方差变化获得噪声情况的方法相似,可以借助于一定窗口内频率超过某一阈值的点认为是由噪声引起,统计窗口内的噪声个数,获得窗口内噪声个数随时间序号的变化关系,亦即

$$n_{fnoise_i} = \sum_{m=1}^{wid} f_m > f_t, \quad (26)$$

$$n_{fnoise} = \{n_{fnoise_1}, n_{fnoise_2}, \dots, n_{fnoise_wid}\} = n_{fnoise}(t_{wid})$$

其中 f_t 为区分信号和噪声的频率阈值, n_{fnoise} 随时间变化。将 n_{fnoise} 与阈值噪声数 $n_{threshold_large}$ 进行比较, 认为 $n_{fnoise} > n_{threshold_large}$ 时为强噪声环境, 认为 $n_{fnoise} < n_{threshold_large}$ 时为弱噪声环境。对问题 3 中频率进行噪声分析得到宽度为 400 的窗口内噪声数变化关系如图 23 所示。可以获得强噪声为 1~11, 34~42 窗口, 相应的时间序号为 1~4400, 13600~17700 范围内。

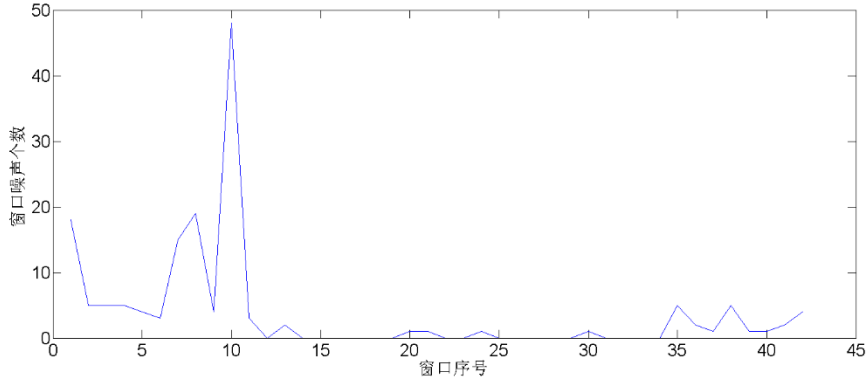


图 23 问题 3 等宽度窗口内噪声个数

根据设置的频率阈值, 可获得如图 24 所示接收信号去噪之后的频率随时间变化序列。可以发现即使去噪之后, 在强噪声环境(1~4400, 13600~17700 范围)中频率仍然有剧烈的涨落。

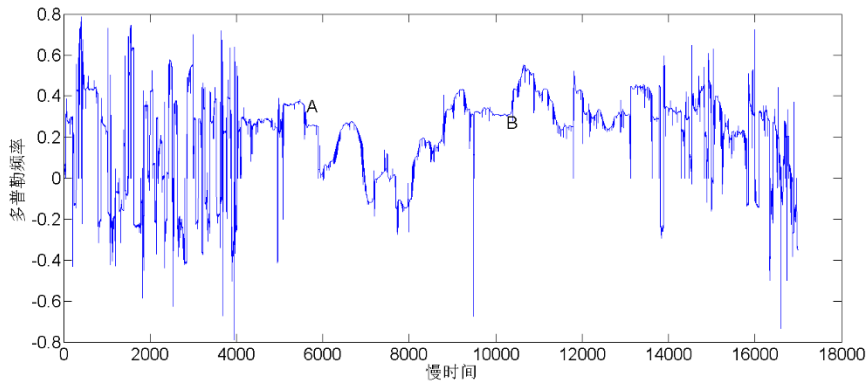


图 24 问题 3 去噪后的多普勒频率变化

3.1.4 变宽度滑窗噪声统计

对于强噪声环境, 可通过噪声统计特性进行分类: 选择更小尺度噪声阈值 $n_{threshold_small}$, 采用变宽度滑动窗口算法, 当由小增大的窗口长度内的噪声个数达到 $n_{threshold_small}$ 时, 该段分段完成, 从下一时刻继续滑窗分段。对两段强噪声环境的时间序列分段结果如图 25 所示, 红三角为分段点。

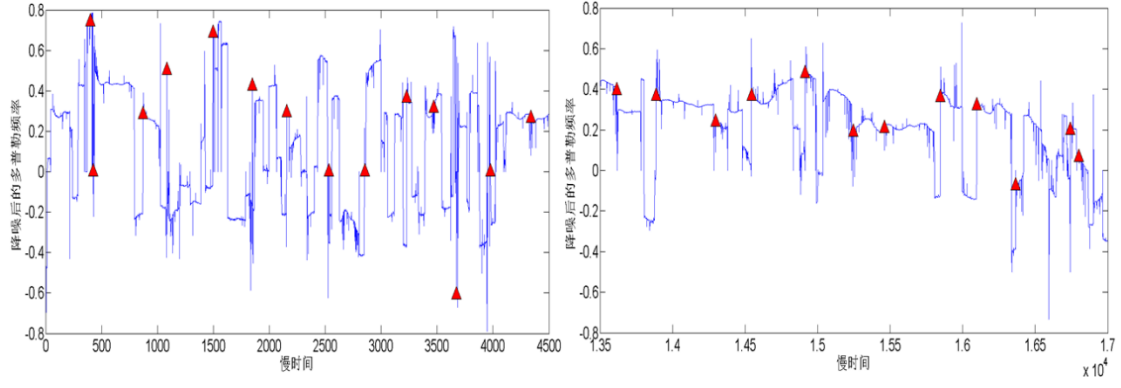


图 25 强噪声环境分段情况

3.2 基于时态边缘算子的时间序列分段方法

对于上述大尺度分段后的弱噪声区域,采用时态边缘算子对其进行小尺度再分段。边缘算子常用于检测图像边缘,以原始图像为基础,对图像各个像素考察其某个邻域内灰度阶跃变化,利用边缘邻近一阶或二阶方向导数变化规律检测边缘。与此相似的,对于时间序列而言,其也存在因为不同场景变化导致的时间序列在时间轴上的边缘,边缘点所在位置一阶或二阶方向导数具有明显变化规律,由此确定的边缘点可作为线段分割点依次连接从而获得时间序列的分段表示,此即基于时态边缘算子的时间序列分段方法。

3.2.1 时间序列的边缘点及 Sobel 算子

图像中边缘即像素灰度明显发生阶跃变化的点。因此,时间序列的边缘点即时间轴邻域内时间序列趋势发生明显改变的点。对于时间序列 $X = \langle x_1, x_2, \dots, x_n \rangle$,考虑其非稳态性和相关性特点,可选择局部范围内边缘幅值极值点作为边缘点。若存在 $i < k < j$,使得 x_k 是 X_{i-j} 范围内边缘幅值极大点,那么 x_k 为极大边缘点;相应的,若存在 $i < k < j$,使得 x_k 是 X_{i-j} 范围内边缘幅值极小点,那么 x_k 为极小边缘点。比如对于图 25 而言,在 A 点处左方其变化很和缓,而其右端则发生剧烈下降,同样对 B 点而言,其左端变化平坦而右端剧烈增加,诸如这种类型的点可以作为边缘点进行时间序列分段。

在经典边缘算子 Sobel 算子中[12,13],其根据像素点上下左右邻域点灰度加权差,在边缘处达到极值进行边缘检测。与此类似的可设计与其相似的时间序时态边缘算子:

$$Sobel(t, u) = \left[w(i) * (x_{t+i} - x_t) / i = -1, -2, \dots, -u, 0, u, \dots, 2, 1 \right] \quad (27)$$

其中 u 表示时态边缘算子的检测窗口长度, $w(i)$ 为检测窗口的位置 i 上的权重。采用时态边缘算子与时间序列作卷积运算,可得到时间序列上各点边缘幅度,幅度越高表示该点邻域时间序列趋势变化越大;边缘幅度越低则说明该点与邻域内时间序列各点处于同一趋势。

3.2.2 边缘幅度变化极值点

使用 Sobel 算子对问题 3 中的多普勒频率进行边缘提取，并使图 26 中斜率变化量超过某阈值点作为分割点进行时间序列分段，对弱噪声部分进行的小尺度再分段得到的分段点如图 27 中红三角所示。

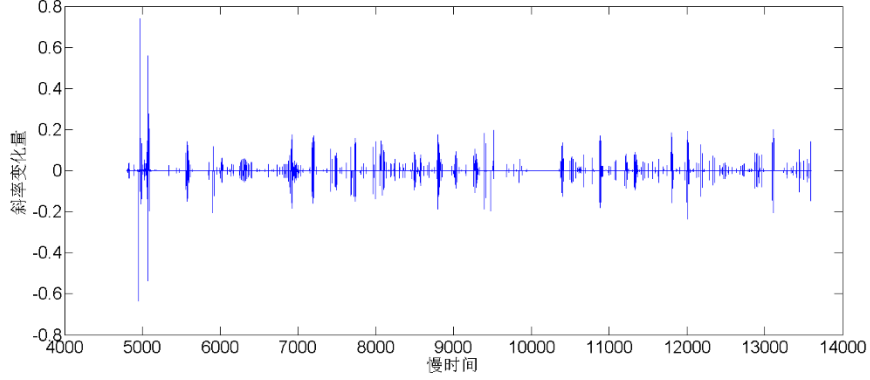


图 26 多普勒频率斜率变化量

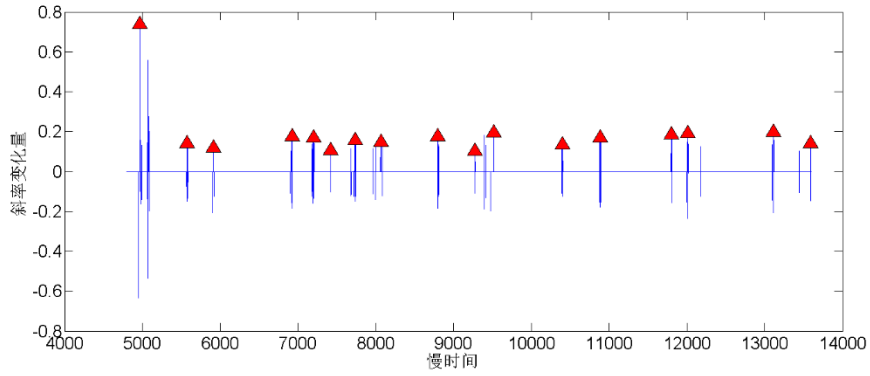


图 27 弱噪声环境的小尺度分段结果

3.2.3 时间序列分段线性表示的拟合误差和压缩率

对时间序列 $X = \langle x_1, x_2, \dots, x_n \rangle$ 的分段线性表示为[14]

$$X_L = \langle f_1(x_{t_1}, x_{t_1}), f_2(x_{t_{l+1}}, x_{t_2}), \dots, f_3(x_{t_m}, x_{t_n}) \rangle \quad (28)$$

由此可计算线性拟合值为

$$X_L^e = \langle x_1^e, x_2^e, \dots, x_n^e \rangle \quad (29)$$

其拟合误差为

$$E = \sqrt{\sum_{i=1}^n (x_i - x_i^e)^2} \quad (30)$$

而分段段数体现了对时间序列的压缩程度，压缩率可以表示为

$$C = [1 - (m + 1) / n] \cdot 100\% \quad (31)$$

分段段数越多，则拟合误差相应越低，然而压缩率增大，则对于场景的误判概率也会增大。

3.3 基于时态边缘算子的尺度收缩时间序列分段算法

将对整条时间序列进行分段的方法可总结为基于时态边缘算子的尺度收缩时间序列分段算法，具体操作过程如下：首先利用等窗口数据分布方差变化或等窗口噪声个数变化对强噪声环境和弱噪声环境进行大尺度分段；然后采用时态边缘算子检测趋势有明显变化的分割点，对弱噪声时间段进行小尺度再分段，整个分段流程如图 28 所示。

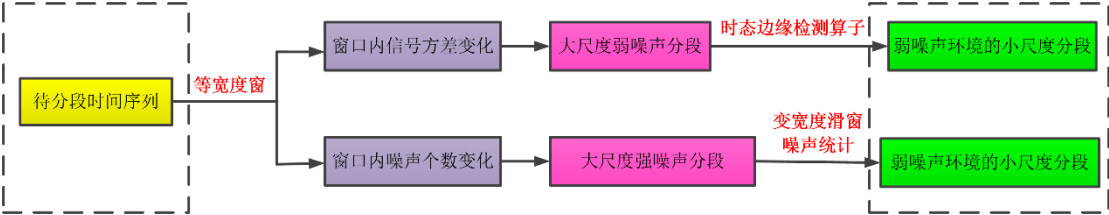


图 28 变尺度时间序列分类流程

采用上述过程，最终得到对数据 3 中以多普勒频率为特征进行的分段结果如图 29 所示，其中红三角所对应位置为分段位置，并展示了原始测量数据的幅值图。对比可以发现，对“指纹”特征采用我们提出的基于时态边缘算子的尺度收缩时间序列分段算法，能够获得直观上无法获得的分段信息。

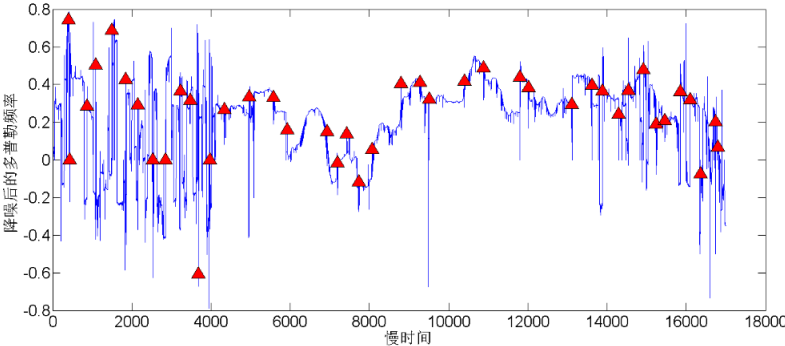


图 29 问题 3 中测试数据分段结果

利用分段点对时间序列进行分段线性表示的结果如图 30 所示，得到的拟合误差仅为 0.32，压缩率为 99.7%，既减少了无关细节的影响，降低了存储容量，又与原始时间序列具有较高的相似度。

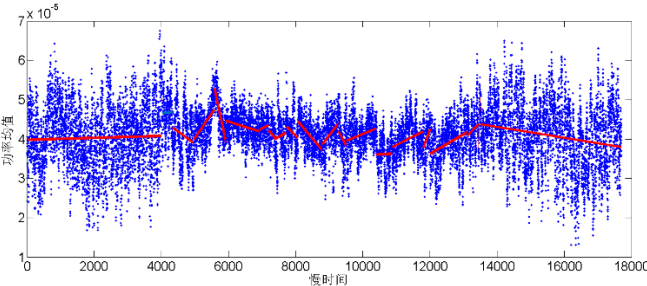


图 30 分段序列的线性拟合结果

第四章 基于支持向量机的邻段聚类方法

支持向量机（Support Vector Machine, SVM）理论源于 Vapnik 在 1963 年提出的用于解决模式识别问题的支持向量方法。这种方法主要是从训练集中选取一组特征子集，使得对特征子集的线性划分等价于对整个数据集的分割[15]。这组特征子集称为支持向量（Support Vector, SV）。

4.1 基于支持向量机的场景识别

4.1.1 基于“指纹特征”的场景匹配

问题 2 要求将两组测试数据 Test1ForScene 和 Test2ForScene 分别与问题 1 中给出的 3 个场景进行匹配。为了识别测试数据所属场景，本文采用 SVM 方法[16]对测试数据做如下处理：

（1）根据问题 1 中建立的无线衰落模型，提取出问题 1 中 3 个场景信号的多径个数、多径系数、多径时延、时延扩展、多普勒扩展等“指纹”特征向量序列 $\{\hat{x}_n\}$ ，利用同样的方法提取出测试数据中的“指纹”特征向量序列 $\{\hat{T}_k\}$ ；

（2）将序列 $\{\hat{x}_n, y_n\}$ 作为训练数据，其中 $\{\hat{x}_n\}$ 为场景 1~场景 3 中的“指纹”特征向量序列， $y_i \in \{1, 2, 3\}$ 为 $\{\hat{x}_n\}$ 所属场景的标签，构造一个最优超平面如下：

$$x(w \cdot \hat{x}) + b = 0 \quad (32)$$

这个最优超平面既要满足下面的约束条件

$$y_i [x(w \cdot \hat{x}) + b = 0] \geq 1, i = 1, 2, \dots, N \quad (33)$$

同时还要使下面的函数取得最小值

$$\phi(w) = \frac{1}{2} \|w\|^2 = \frac{1}{2} (w \cdot w) \quad (34)$$

通过求解最优化问题可得最优超平面的形式如下：

$$\sum_{SV} y_i a_i^0 (x \cdot x_i) + b_0 = 0 \quad (35)$$

其中 SV 表示支持向量， a_i^0 是拉格朗日乘子。

（3）将从测试数据中提取的“指纹”特征向量序列 $\{\hat{T}_k\}$ 代入最优超平面表达式中，可以得到测试数据所属场景的标签 $\{\hat{L}_k\}$ ，即可以识别测试数据所属场景。

4.1.2 基于 Bayes 假设检验的概率分析

对测试数据中提取的“指纹”特征向量序列的匹配结果如图 31、图 32 和表

2、表 3 所示。

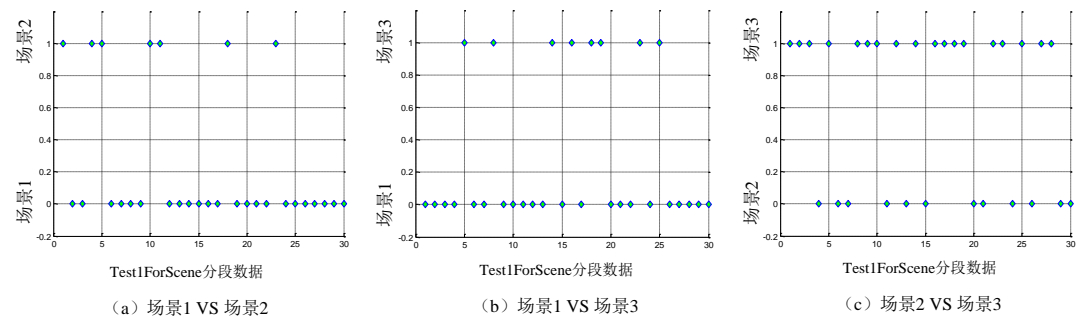


图 31 基于 SVM 的问题 2Test1 与问题 1 中场景匹配结果

表 2 基于 SVM 的问题 2Test1 场景匹配结果及 Bayes 后验概率

测试数据	Test1ForScene.mat					
分类器	场景 1 VS 场景 2		场景 1 VS 场景 3		场景 2 VS 场景 3	
分类结果	23	7	22	8	21	9
Bayes 后验概率	1	0	1	0	1	0

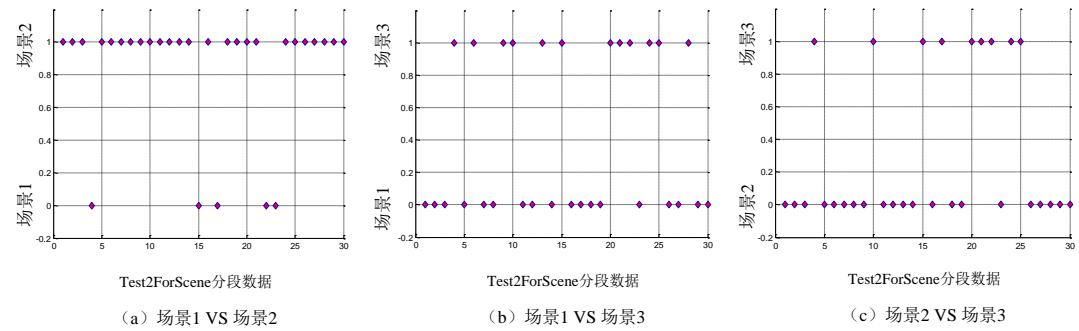


图 32 基于 SVM 的问题 2Test2 与问题 1 中场景匹配结果

表 3 基于 SVM 的问题 2Test2 场景匹配结果及 Bayes 后验概率

测试数据	Test2ForScene.mat					
分类器	场景 1 VS 场景 2		场景 1 VS 场景 3		场景 2 VS 场景 3	
分类结果	5	25	18	12	21	9
Bayes 后验概率	0	1	91.9%	8.1%	1	0

根据支撑向量机 SVM 的输出，利用贝叶斯公式可得到分类的后验概率。样本通过 SVM 后得到输出 1 或 0 的过程可以看成是一个二项分布，并且假设有两类总体 A 和 B，它们中的样本取 1 和 0 的先验概率分别为 P_A 和 $1-P_A$ 以及 P_B 和 $1-P_B$ 。这里的先验概率 P_A 和 P_B 可以用训练后的 SVM 对于训练样本 A 和 B 进行测试从而获得[17,18]。

假设新的数据为 Data 中有 N 个样本，用训练的好的 SVM 对 Data 进行测试，得到 1 的个数为 n 。根据贝叶斯公式有

$$\begin{aligned}
P(A|n) &= \frac{P(A)P(n|A)}{P(A)P(n|A) + P(B)P(n|B)} \\
&= \frac{C_N^n p_a^n (1-p_a)^{N-n}}{C_N^n p_a^n (1-p_a)^{N-n} + C_N^n p_b^n (1-p_b)^{N-n}}
\end{aligned} \tag{1.1}$$

在已知数据 *Data* 中 N 个样本经过 SVM 后出现 1 的个数为 n ，上式给出了数据 *Data* 属于总体 A 的后验概率计算公式。

通过 SVM 对场景进行两两匹配后，可以推断问题 2 中两组数据的场景。

对 Test1ForScene 进行判断时，1) 分别利用场景 1 和场景 2 与 Test1 进行匹配，发现属于场景 1 的概率高达 100%；2) 分别利用挑选出的场景 1 和剩下的场景 3 对 Test1ForScene 进行匹配，发现属于场景 1 的概率为 100%；3) 可见场景 1 和场景 2/场景 3 分别进行对比，场景 1 的匹配概率均为 100%，为此，可判断出问题 2 中 Test1ForScene 属于场景 1。

对 Test2ForScene 进行判断时，1) 分别利用场景 1 和场景 3 与 Test2 进行匹配，发现属于场景 1 的概率高达 91.9%；2) 分别利用挑选出的场景 1 和剩下的场景 2 对 Test2ForScene 进行匹配，发现属于场景 2 的概率为 100%；3) 可判断出问题 2 中 Test2ForScene 属于场景 2 的概率为 91.9%，属于场景 3 的概率为 8.1%。

综上所述，基于“指纹”特征向量序列的 SVM 分类结果，我们判断测试数据 Test1ForScene 属于场景 1，测试数据 Test2ForScene 属于场景 2。

4.2 基于 SVM 的邻段聚类区域划分

4.2.1 原理示意

我们对数据 3 中提取的“指纹”特征向量运用基于时态边缘算子的尺度收缩时间序列分段算法进行分段，得到数据 3 中时间序列的一组细分结果，各个分段具有各自的指纹，可以利用指纹的相似程度将分段结果进行聚类，从而得到比较合理的“区域划分”。

为了对细分结果进行聚类，我们提出基于 SVM 的邻段聚类方法，该方法的步骤如下：

(1) 将 3.1 中得到的每一段细分区域作为一个样本集合，依次标记为 $\{S_1, S_2, \dots, S_n\}$ ，如图 33 所示：

S_1	S_2	S_3	S_{i-1}	S_i	S_{i+1}	S_{i+2}	S_n
-------	-------	-------	-------	-----------	-------	-----------	-----------	-------	-------

图 33 基于 SVM 的邻段聚类方法步骤 1

(2) 每次取 3 个样本集合 $\{S_{i-1}, S_i, S_{i+1}\}$ 作为 SVM 的聚类输入，其中样本集合 $\{S_{i-1}, S_{i+1}\}$ 为训练样本，样本集合 $\{S_i\}$ 为测试样本进行聚类，如图 34 所示：

S_1	S_2	S_3	S_{i-1}	S_i	S_{i+1}	S_{i+2}	S_n
				训练	测试	训练			

图 34 基于 SVM 的邻段聚类方法步骤 2

(3) 假如样本集合 $\{S_i\}$ 与 $\{S_{i-1}\}$ 的“指纹”特征向量序列更为相似[19], 则将样本集合 $\{S_i\}$ 与 $\{S_{i-1}\}$ 合并为 $\{S_i\}$, 如图 35 所示;



图 35 基于 SVM 的邻段聚类方法步骤 3

(4) 重复步骤 2 和 3, 依次将为进行聚类的样本集合 $\{S_k\}$ 作为测试样本, 左右相邻的样本集合 $\{S_{k-1}, S_{k+1}\}$ 作为训练样本, 输入 SVM 中进行聚类, 并进行合并, 最后得到的聚类结果示意如图 36 所示;

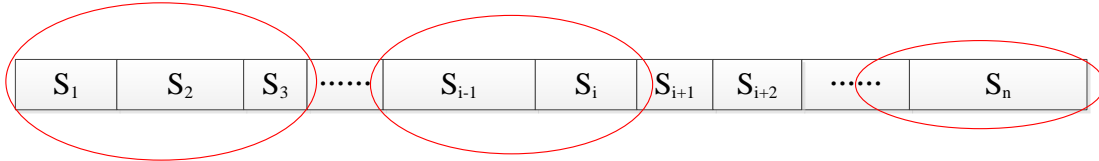


图 36 基于 SVM 的邻段聚类方法步骤 4

通过基于 SVM 的邻段聚类进行区域划分, 我们得到的区域划分结果如表 4 所示:

表 4 问题 3 中时间序列最终聚类结果

区域序号	1	2	3	4
区域范围	(1, 426)	(427, 2155)	(2156, 3471)	(3472, 5578)
区域序号	5	6	7	8
区域范围	(5579, 7199)	(7200, 9279)	(9280, 11798)	(11799, 14544)
区域序号	9	10	11	12
区域范围	(14545, 15245)	(15246, 16097)	(16098, 16742)	(16743, 17700)

4.2.2 问题 4 实测数据的分类判断

通过上述的时间序列分割和聚类步骤, 我们得到了对数据包 3 中数据在慢时间上的一个划分总共 12 段数据, 记为 $Seg(i), i = 1, 2, \dots, 12$ 。这里我们对每一个 i 构造一个用于识别 $Seg(i)$ 和其余所有数据的 SVM, 利用这个 SVM 可以对样本 A 进行测试, 根据上节中的公式计算出样本 A 和 $Seg(i)$ 为同一类的概率为 $P(i)$ 。表 5 列出了两个测试样本的 $P(i)$ 的数值。

表 5 问题 4 中实测数据的分类匹配结果

i	1	2	3	4	5	6
SampleForTest1	29%	26%	30%	37%	21%	15%
SampleForTest2	31%	37%	47%	34%	55%	87%
i	7	8	9	10	11	12
SampleForTest1	24%	33%	42%	34%	29%	30%
SampleForTest2	51%	42%	39%	29%	27%	33%

(1)对于样本 **SampleForTest1**，贝叶斯后验概率存在几个峰值，我们对其贝叶斯后延概率存在明显峰值的两段 **Sega(4)**和 **Sega(9)**进行对比，选取主径系数、距离多普勒成像和时延拓展为“指纹”特征，对比结果如图 37、图 38 和表 6 所示。

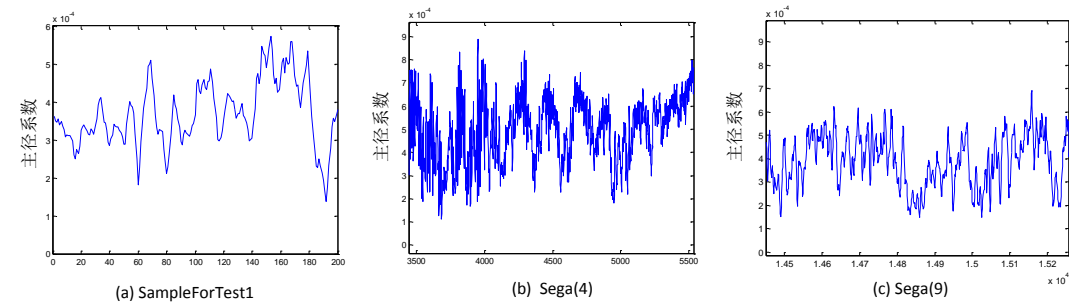


图 37 问题 4Test1 与不同类的主径系数对比

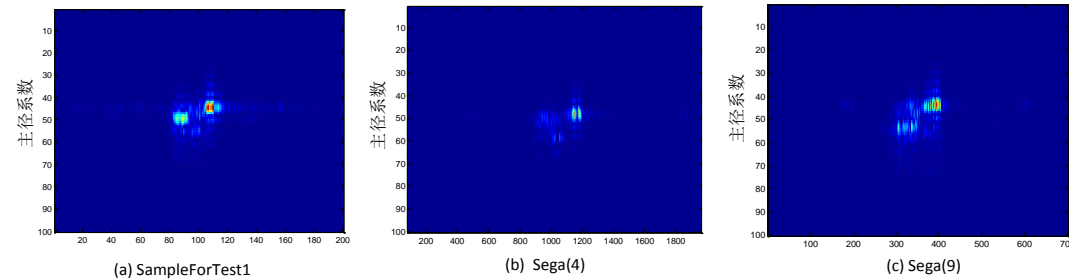


图 38 问题 4 中 Test1 与不同类的距离多普勒成像对比

表 6 问题 4 中 Test1 与不同类的时延扩展对比

数据段	SampleForTest1	Sega(4)	Sega(9)
时延拓展	1030.8	1312.9	1794.1

通过上述对比我们可以发现：(1)在主径系数方面，**SampleForTest1** 与 **Sega(4)** 比 **Sega(9)**变化更为剧烈，即在主径中的信号衰落特性变化更为剧烈；(2)在距离多普勒成像方面，由图中可以明显看出，**SampleForTest1** 中存在两个信号较强的路径，而在 **Sega(4)**和 **Sega(9)**中的路径信号则没有这么明显，且 **SampleForTest1** 中的多普勒拓展较宽；(3)在时延拓展方面，**SampleForTest1** 明显要比 **Sega(4)**和 **Sega(9)**中的数据要小。

(2)对于第二个样本 **SampleForTest2**，在第 6 段存在一个明显较大的概率，而且接近 90%，因此我们可以认为，该样本是属于这一段中。为了验证检验是否正确，我们对 **Seg(6)**和 **SampleForTest2** 的信道中主路径的幅度调制因子求出了他们的归一化滑动自相关系数（如图 39）。从相关的结果中可以发现，两个数据拥有较高的相关系数，并且在 8901 个点处出现了接近于 1 的峰值，其位置大约在 75.4~77 米处。

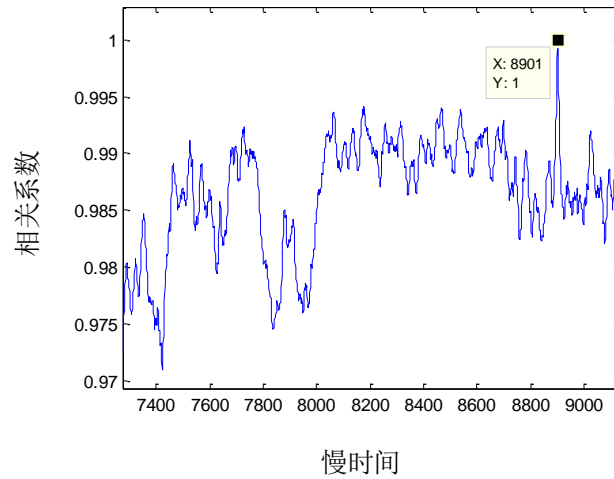


图 39 问题 4 中 Test2 主径幅度调制因子归一化滑动相关系数

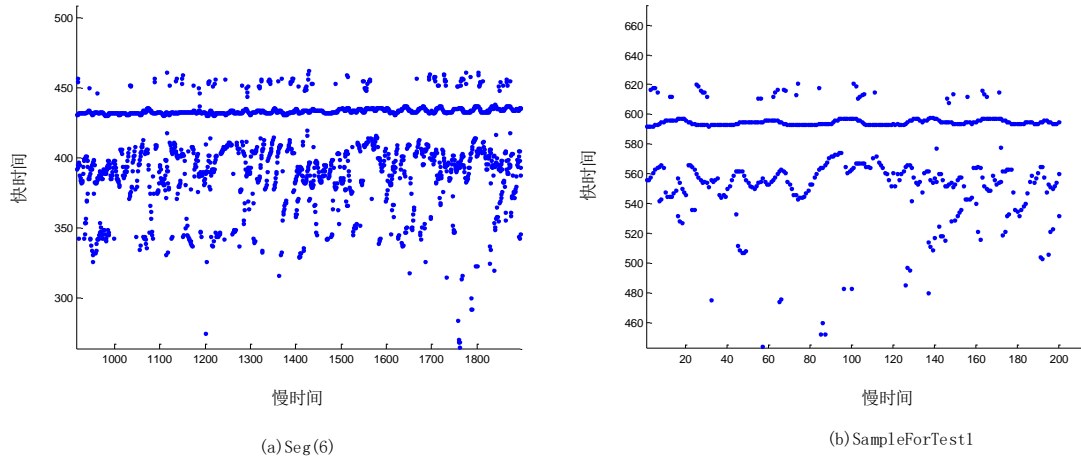


图 40 问题 4 中 Test2 超分辨信道路径提取结果

同时我们利用超分辨提取每个时刻信道中不同路径的时延，如图 40 所示，通过比较可以看出，二个信道中均有一条稳定的主径，主径下方还有一条稳定性差，时延波动较大的路径。同时，我们可以发现两个数据中的主径起伏变化的规律也较为接近。经过计算，Seg(6)的平均时延扩展和多普勒扩展为 334ns 和 14.2Hz 而 SampleForTest2 的为 319ns 和 10.5Hz，二者较为接近。

综合上述“指纹”特征的对比，我们认为 SampleForTest1 不属于问题 3 中的任何区域；SampleForTest2 有 87% 的概率属于数据中的第六段，且位置在 75.4~77m 附近。

第五章 总结

在本文中，我们详细分析了移动通信中电磁波在多径信道中的传播过程，推导了接收信号能量和多普勒频率的物理量随接收端移动而产生的变化。进而建立了无线信道的多基复包络线性叠加模型，实测数据处理结果说明，该模型可以对真实场景下的无线信道进行较好的拟合和描述。

针对特征提取，(1) 我们借鉴雷达信号处理技术中的成像方法对信道进行距离多普勒域的联合成像，从而可以在时延和多普勒的二维平面上将同一信道中的不同路径分开，实现对不同场景的区分。(2) 对于模型中的参数，我们利用基于 CLEAN 的超分辨谱估计的方法对每条路径进行估计，并利用 Kalman 滤波和航迹关联技术对信道中存在的主要路径进行提取。处理结果表明，这种方法可以获得精确的参数估计结果，并且从估计出来的参数中也能够较好的对各个场景进行区分。(3) 在成像和参数估计的基础上，我们构造了无线信道的特征向量，即“指纹”。它包括时延扩展、多普勒扩展、衰落特征等参数特征，以及对比度和熵等从成像结果中提取出来的图像评价指标。

针对较长时间的接收信号（时间序列）分段问题，针对提出变尺度分段法。

(1) 以多普勒频率为指纹特征，首先根据等宽度窗口内噪声统计量进行大尺度分段。(2) 利用变宽度滑窗法增加噪声个数至噪声数阈值，完成强噪声环境小尺度分段。(3) 利用时态边缘检测 Sobel 算子获取弱噪声环境下斜率变化较大的边缘点作为分段点，实现弱噪声环境下小尺度分段。变尺度分段法将问题 3 中时间序列分为 43 段，拟合误差仅为 0.32，压缩率高达 99.7%。

针对无线信道场景的识别和聚类问题，我们采用了 SVM 作为分类器的基本组成单元。(1) 利用不同场景中提取出来的信道“指纹”特征对 SVM 进行训练，并利用贝叶斯后验概率作为判别准则，可以实现对场景的识别。(2) 为了完成对无标签信道场景的聚类，我们设计了一种基于 SVM 的相邻段聚类算法，实现了对分割后的信道段落进行聚类 and 场景划分。数据包 2、3、4 中数据的处理结果表面，SVM 可以有效根据本文中提取出来的“指纹”特征进行对无线信道的场景进行识别和聚类。利用上述方法判断出问题 2 中两组数据分别属于场景 1（概率接近 100%）和场景 2（概率为 91.9%）。并判断出问题 4 中第一组数据不属于问题 3 中分类，第二组数据位于问题 3 中数据的第 75.4~77m 范围内（概率为 87%）。

经过上述建模分析过程，对无线信道的特征提取和识别聚类问题有了一定程度的认识，对分析无线通信信道和改善无线通信效果具有重要意义。

参考文献

- [1] 数字移动通信[M]. 人民邮电出版社, 1995.
- [2] 吴伟陵, 牛凯. 移动通信原理[M]. 电子工业出版社, 2005.
- [3] 谭立新, 何艳丽. 多径衰落信道的统计特性与仿真研究[J]. 计算机仿真, 2010 (7): 96-98.
- [4] 胡少青, 杨雪松, 袁家劫. 基于时域数据的室内超宽带信道建模[J]. 通信技术, 2013 (3): 10-12.
- [5] 王勤, 万显荣, 杨子杰, 熊俊志. 基于 CLEAN 算法的非连续谱线性调频中断波信号处理[J]. 电波科学学报, 2009, 02: 243-248.
- [6] 酒明远. ISAR成像中微多普勒现象及CLEAN技术研究[D]. 哈尔滨工业大学, 2010.
- [7] Schmitt J M. Restoration of optical coherence images of living tissue using the CLEAN algorithm[J]. journal of Biomedical Optics, 1998, 3(1): 66-75.
- [8] 刘妍妍, 张新, 张建萍. 超分辨重建技术及其研究进展[J]. 中国光学与应用光学, 2009, 02: 102-111.
- [9] 胡晓琴. 超分辨空间谱估计技术应用基础研究[D]. 国防科学技术大学, 2009.
- [10] Frühwirth R. Application of Kalman filtering to track and vertex fitting[J]. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 1987, 262(2): 444-450.
- [11] 丛蓉, 王秀坤, 李进军, 等. 基于层次和密度聚类分析的航迹关联算法[J]. 系统仿真学报, 2005, 17(4): 841-843.
- [12] 袁春兰, 熊宗龙, 周雪花, 等. 基于 Sobel 算子的图像边缘检测研究[J]. 激光与红外, 2009, 39(1): 85-87.
- [13] 何春华, 张雪飞, 胡迎春. 基于改进 Sobel 算子的边缘检测算法的研究[J]. 光学技术, 2012, 38(3): 323-327.
- [14] Kalman R E. A new approach to linear filtering and prediction problems[J]. Journal of Fluids Engineering, 1960, 82(1): 35-45.
- [15] Vapnik V N, Vapnik V. Statistical learning theory[M]. New York: Wiley, 1998.
- [16] 李红莲, 王春花, 袁保宗. 一种改进的支持向量机 NN—SVM[J]. 计算机学报, 2003, 26(8).
- [17] 程克非, 张聪. 基于特征加权的朴素贝叶斯分类器[J]. 计算机仿真, 2006, 23(10): 92-94.
- [18] 李锦善, 王志海, 王中锋. 一种基于假设检验的贝叶斯分类器[J]. 计算机工程与应用, 2009, 44(21): 222-224.
- [19] 张鸿宾, 孙广煜. 近邻法参考样本集的最优选择. 电子学报, 2000, 28(11): 16~ 21.

附录：Matlab 程序代码

```
% ,Ã³ ÌÐòÊµÏÖŒÔÐÂµÀµÃ³¬ ·Ö±æ² îÊÝ¹À¼Æ°ÍÖ÷¼ŒÎáÈ;
close all
clear
%***** Ñ;ÔñÐèÒª ·ÖÎöµÃÊÝ¼Ý
% load('ÊÝ¼Ý°ü1/³¼¹/Test1.mat')
% load('ÊÝ¼Ý°ü3/Sample.mat')
load('ÊÝ¼Ý°ü4/SampleForTest2.mat')
%*****
figure,imagesc(abs(ChannelIR.))

load('ÊÝ¼Ý°ü5/filter.mat')
N=size(ChannelIR,1); %Ãö³ã,öÊÝ
Num=10; %Ã¬ÈÏÃ·¼Œ,öÊÝ²»³¬¹ÝNum
cont=[]; %³öÊ¼»¬Ã·¼Œ,öÊÝ¼ÇÃ¼ÊÝ×é
tao=zeros(10,N); %³öÊ¼»¬Ê±ÑÓÊÝ×é
phase=zeros(10,N); %³öÊ¼»¬ÏàÊÝ×é
ai=zeros(10,N); %³öÊ¼»¬Ê¥ÃäÏµÊÝÊÝ×é
ti=8; %²ãÖµ±ŒÊÝ(·Ö±æÃÊ±ŒÊÝ)
K=ti*size(ChannelIR,2);
% ÊË²¬Æ²²ãÖµ
filter = [filter, zeros(1, 51)];
filter=ifftshift(ifft(fftshift(filter)));
filter=fftshift(fft((filter),K));
[~,mid_filter]=max(abs(filter));
% figure,plot(abs(filter))

for m1=1:N
    x=ChannelIR(m1,:);
    x1=ifftshift(ifft((x)));
    x1=(fft((x1),K));
    x=x1;

    power1 = sum(abs(x).^2);
    var1=var(x);
    for k=1:Num
        [a,Index]=max(x);
        filter2 = circshift(filter', Index-mid_filter);
        filter2 = filter2';
        phase=angle(x(Index));
        a=x(Index);
        x = x(1, :) - a * filter2;
```

```

        power2 = sum(abs(x).^2);
        var2=var((x));
        cont(m1)=k;
        tao(k,m1)=Index;
        ai(k,m1)=abs(a);
        phi(k,m1)=angle(a);
        if (var2/var1>0.3)
            break;
        end
        %         figure,plot(abs(x))
        var1=var2;

    end
end
% figure,plot(cont)
% save Test2_forscene tao ai phi cont
%% »-Ê±ÑÓµÄÊçµäí¼
figure
hold on
for k=1:Num
    plot(800-tao(k,:),'.')
end

%% ÀûÓÃ¼4£¹ØÁª°íÄË²²ÏáÈ;Ö÷¼¶
n=2;
ts=1/100;
Phi=[1,ts,ts^2/2;0,1,ts;0,0,1];
Tao=[ts^2/4;ts^2/2;1];
H=[1,0,0];
Q=10;
R=1;
r_e(:,1)=[tao(n,1),0,0];
r_e(:,2)=[tao(n,2);(tao(n,2)-tao(n,1))/ts;0];
P=[R^2,R^2/ts,0;R^2/ts,ts^2*Q^2/4+2*R^2/ts^2,0;0,0,R];
r_pre(1)=r_e(1,1);
for k=3:length(tao(n,:))
    % ÖÐÄÑ;×í´óÖµ
    %         men=size(s2,1)/8;
    %         temp=s2(index(k-1)-men:index(k-1)+men,k);
    %         [~,M]=max(abs(temp));
    %         index(k)=index(k-1)+M-men;
    %
    zz=tao(n,k);

```

```

%      È¥³ýÆæÒìÖµ
if abs(tao(n,k)-r_e(1,k-1))>K/40
    zz=r_pre(1);
end

[P,P_pre,r_e(:,k),r_pre]=Kalman_Filter(Phi,Tao,Q,H,R,P,r_e(:,k-1),zz);

end

figure
hold on
for k=1:Num
    plot(800-tao(k,:),'.')
end
plot(800-r_e(1,:), 'r')

%

function [x,weiz,R,b1]=line_match(x)
[M,N]=size(x);
MM=fix(M/2);

%   ham=hamming(M);
%   ham=ones(1,M).';
%   Ps=exp(-1j*pi*[0:M-1]');
%   HPS=ham.*Ps;
%   Vec_M=[0:M-1]';
point=8;
LEN = M*point;
MM2 = M*point;
MM_line = (1:MM2)- floor(MM2/2);
Vec_M = MM_line;
Vec_M = Vec_M.';

for n=1:N
    if n==1
        a = fft_inter(x(:,n),point);
        a=abs(a);
%       mopt=[1:LEN]*(a.^2/sum(a.^2));
        [~,mopt] = max(a);
        mopt = mopt-floor(MM2/2);
%       mopt = round(mopt);
%       pin pu
        A_temp = fftshift(fft(x(:,n)),1);

```

```

Sig_new = zeros(LEN,1);
Sig_mid = floor(LEN/2);
sig_mid = floor(M/2);
Sig_new((Sig_mid-sig_mid):(Sig_mid-sig_mid+M-1)) = A_temp;

Sig_new = Sig_new.*exp(1j*2*pi*Vec_M/(MM2)*mopt);
%      a=ifft(x(:,n)).*exp(-1j*2*pi*Vec_M/M*mopt).*HPS,LEN);
a_temp = ifft(fftshift(Sig_new,1));
x(:,n)=a_temp(1:point:LEN,1)*point;
%      a=ifft(exp(1j*2*pi*Vec_M/M*(mopt-MM)).*x(:,n)).*HPS,LEN);
%      x(:,n)=a(1:point:LEN,1)*point;
b1=abs(a_temp);
outR(1)=1;
weiz(n)=(mopt-MM);
else
    b=fft(sqrt(b1));
    a = abs(fft_inter(x(:,n),point));
    a = fft(sqrt(a));
    R=abs(fftshift(ifft(b.*conj(a))));
    %      R=abs(R);
    %      R=real(R);
    [maxR,maxm]=max(R);
    maxm;
    xstar=0;
    % plot(R);
    if (maxm~=1) & (maxm~=M*point)
        f(1)=R(maxm-1);
        f(2)=R(maxm);
        f(3)=R(maxm+1);
        fa=(f(1)+f(3)-2*f(2))/2;
        fb=(f(3)-f(1))/2;
        fc=f(2);
        xstar=-fb/(2*fa);
    end
    outR(n)=maxR;
    mopt=maxm+xstar-floor(MM2/2)-1;%
    maxR;
    mopt;
    mopt=mopt;
    weiz(n)=-mopt;
%      a_temp = fft_inter(x(:,n),point);

A_temp = fftshift(fft(x(:,n)),1);
Sig_new = zeros(LEN,1);

```

```

Sig_mid = floor(LEN/2);
sig_mid = floor(M/2);
Sig_new((Sig_mid-sig_mid):(Sig_mid-sig_mid+M-1)) = A_temp;

Sig_new = Sig_new.*exp(-1j*2*pi*Vec_M/(MM2)*mopt);
%      a=ifft(x(:,n)).*exp(-1j*2*pi*Vec_M/M*mopt).*HPS,LEN);
a_temp = ifft(ifftshift(Sig_new,1));
x(:,n)=a_temp(1:point:LEN,1)*point;
%      b1=abs(a);
%      b1=b1+abs(a);
b1=b1*0.95+abs(a_temp);
%b1=b1*exp(-0.1)+abs(a);
end
end

% for k=1:N
%      x(:,k)=fft(fftshift(x(:,k)));
% end

```